8-11-2020

Neural Coordination of Distinct Motor Learning Strategies: Latent Neurofunctional Mechanisms Elucidated via Computational Modeling

Robert A. Capps
Georgia State University

NEURAL COORDINATION OF DISTINCT MOTOR LEARNING STRATEGIES: LATENT NEUROFUNCTIONAL MECHANISMS ELUCIDATED VIA COMPUTATIONAL MODELING

by

ROBERT A. CAPPS

Under the Direction of Yaroslav I. Molkov

ABSTRACT

In this dissertation, a neurofunctional theory of learning is presented as an extension of functional analysis. This new theory clarifies the distinction— via applied quantitative analysis— between functionally intrinsic (essential) mechanistic structures and irrelevant structural details. This thesis is supported by a review of the relevant literature to provide historical context and sufficient scientific background. Further, the scope of this thesis is elucidated by two questions that are posed from a neurofunctional perspective— (1) how can specialized neuromorphology contribute to the functional dynamics of neural learning processes? (2) Can large-scale neurofunctional pathways emerge via inter-network communication between disparate neural circuits? These questions motivate the specific aims of this dissertation. Each aim is addressed by posing a relevant hypothesis, which is then tested via a neurocomputational experiment. In each experiment, computational techniques are leveraged to elucidate specific mechanisms that underlie neurofunctional learning processes. For instance, the role of specialized neuromorphology is investigated via the development of a computational model that replicates the neurophysiological mechanisms that underlie cholinergic interneurons' regulation of dopamine in the striatum during reinforcement learning. Another research direction focuses on the emergence of large-scale neurofunctional pathways that connect the cerebellum and basal ganglia— this study also involves the construction of a neurocomputational model. The results of each study illustrate the capability of neurocomputational models to replicate functional learning dynamics of human subjects during a variety of motor adaptation tasks. Finally, the significance— and some potential applications— of neurofunctional theory are discussed.

INDEX WORDS: Motor learning, Motor adaptation, Neurofunction, Basal ganglia, Cerebellum, Reinforcement learning, Model-free learning, Model-based learning, Striatal cholinergic interneurons, Tonically active neurons, Biomechanics, Neurophysics, Computational neuroscience

NEURAL COORDINATION OF DISTINCT MOTOR LEARNING STRATEGIES: LATENT NEUROFUNCTIONAL MECHANISMS ELUCIDATED VIA COMPUTATIONAL MODELING

by

ROBERT A. CAPPS

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2020

NEURAL COORDINATION OF DISTINCT MOTOR LEARNING STRATEGIES: LATENT NEUROFUNCTIONAL MECHANISMS ELUCIDATED VIA COMPUTATIONAL MODELING

by

ROBERT A. CAPPS

Committee Chair:	Yaroslav Molkov
Committee:	Igor Belykh
	Vladimir Bondarenko
	Andrey Shilnikov

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2020

DEDICATION

This work is dedicated to my friends and family who have graced me with their love and support through this journey— you all know who you are. I cannot possibly list here everyone who deserves to be mentioned by name.

Foremost, I dedicate this work to my parents. Your persistent altruism never ceases to inspire me. Dad, to quote from one of your favorite books,

> ‘Love’ is that condition in which the happiness of another person is essential to your own.
>
> (Robert Heinlein — *Stranger in a Strange Land*)

To my sister Cici and my brother Kenny— you are among the few who understand the personal significance of this milestone. You have helped me through tough times— I think sometimes without even realizing it.

Auntie J, Suzy, Aunt Cath, Uncle Guy, Uncle Bubby, the Brown-Capps family, the Dickinsons— you guys have consistently been there when I needed a friendly face or a helping hand. I hope you realize just how much you have contributed to my success, especially toward my personal growth as an individual. Thank you.

To my life partner Elizabeth— you rock. I love you.

To Mrs. Kim Buchanan— you taught me to think critically and write thoughtfully. The lessons you shared through your mentorship have proven invaluable. Thank you.

I also dedicate this work to my friends, especially those who have supported me in recent years. To name a few— Bryce Chung, Chris Richardson (Dr. Drank), Xavier Francia (Kitty), Robert Quillen, Erik Orndahl, Ricardo Erazo Toscano, Dmitrii Todorov, Will Barnett, the Two Kevins. Last but not least— a big ”Thank you!” to Dr. Antonio Roque, the instructors, and fellow students who contributed to LASCON 2018.

Thanks, guys! I sincerely couldn’t have done it without your help.

ACKNOWLEDGEMENTS

It would be impossible to acknowledge everyone who has helped me throughout my academic career, but I'll do my best. In no particular order...

I'd like to thank my adviser and the chair of my dissertation committee Dr. Yaroslav "Slava" Molkov for his guidance and mentorship. Slava, I cannot express how grateful I am for your seemingly limitless patience and understanding. Thank you for sharing your knowledge and wisdom.

Special thanks to committee members and mentors Dr. Igor Belykh, Dr. Vladimir Bondarenko, Dr. Andrey Shilnikov.

Dr. Belykh and Dr. Bondarenko, through your teaching, you shared with me foundational knowledge of (bio)mathematics & (bio)physics that has been— and will continue to be— important throughout my research career. Dr. Shilnikov, your *Dynamical Foundations of Neuroscience* course was my first encounter with dynamical systems theory. The hands-on, research-based assignments— and of course your unique lecture style— sparked my interest in computational neuroscience.

I would also like to acknowledge my mentors Dr. Donald Edwards, Dr. Bryce Chung, and Dr. Remus Osan. Thank you sincerely for your guidance and support. Dr. Edwards, thank you for generously inviting me to work with your lab. The experience provided me with practical knowledge and gave me the confidence necessary to pursue a career in neuroscience. Bryce, thank you for your unending patience and for selflessly sharing your time and attention. Remus, thank you for sharing your knowledge and wisdom— thank you also for facilitating educational and memorable experiences in Romania and Hungary.

In addition, I would like to thank Dr. Lisa Armistead, the Dean of Georgia State's Graduate School, for her role in helping me transfer to GSU's Neuroscience Institute— by the way, congratulations on the new role!

Further, I would like to acknowledge the monetary support provided by the Second

Century Initiative (2CI) university fellowship program— without which my graduate education would not have been possible. I would also like to acknowledge the CHDI Foundation grant A-8427, which provided funding for the work described in chapter 4 and chapter 5.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The functional role of striatal cholinergic interneurons in reinforcement learning from computational perspective [Kim et al., 2019] 27

The interplay between cerebellum and basal ganglia in motor adaptation: a modeling study [Todorov et al., 2019] 59

LIST OF ABBREVIATIONS

Kim and Capps et al., 2019

- TANs - Tonically Active Neurons
- RPE - Reward Prediction Error
- TP - TAN pause
- L-DOPA - Levodopa
- DA - Dopamine
- DA Def - Dopamine Deficiency

Todorov and Capps et al., 2019

- BG - Basal Ganglia
- CB - Cerebellum

1

INTRODUCTION

1.1 Neurofunctional theory

In this report, a new theory— a ”complete” neurofunctional theory of learning— is introduced. This new theory originated as an extension of functional analysis, a concept borrowed from analytical psychology (see: section 2.6). In [Piccinini and Craver, 2011] functional analyses are defined as “sketches of mechanisms, in which some structural aspects of a mechanistic explanation are omitted.” The goal of the present work is to optimally distinguish between functionally intrinsic (i.e. essential) mechanistic structures and irrelevant structural details. In this report, a theoretical framework is derived from several independent strands of evidence. This framework generally aims to define *the minimum amount of mechanistic detail that could explain the observed operation of a particular behavior or set of behaviors.*

Neurofunctional theory can be understood in terms of semantic structure— i.e. the biological and physical models from which neurofunctional theory is derived. An important advantage of neurofunctionalism is the potential benefit for empirical analysis of a wide variety of neurobehavioral phenomena. This practical advantage originates with the antecedent theory of Behaviorism (section 2.5). For example, a behavioral study of stress response in expectant mothers could be considered from a neurofunctional perspective, as could a molecular study of serotonin transporter expression in the hypothalamus.

The advantages provided by a neurofunctional perspective can also be explained in semantic terms. Key among these advantages is the concept of *neurofunction*, which is defined as *the operation of (neuro)behavioral processes in terms of minimum description length.* The term “neurofunction” is derived from the meaning of ”function” in “functional analysis,” an approach in behavioral psychology that aims to connect psychological consequences (e.g.

mood, cognition) to specific causative mechanisms (e.g. neurophysiology). Thus, neurofunctionalism can be partially understood from an analytical psychology perspective. Moreover, the definition of *neurofunction* also has roots in a well-known neurocomputational principle— the ”minimum description length” (MDL), which is a formalization of Occam’s razor that originates from computational learning theory [Rissanen, 1978, Peter Grünwald, 1998]. As stated in the 1978 publication that first introduced MDL,

> [MDL can be utilized as] a criterion for estimation of parameters, including the structure parameters, in a model for a random time series has been derived from the single and natural principle: minimize the number of bus it takes to write down the observed sequence.
>
> (Rissanen, 1978)

This description as well as the formal definition of MDL supports the central observation of neurofunctional theory: *Any observed neurofunction (i.e. neurofunctional or neurofunctional unit)— including neurofunctionals with nonlinear temporal dynamics— can be expressed quantitatively*, e.g. as a Bayesian inference problem [Molkov et al., 2009]. Moreover, this realization could be used to derive a neurofunctional conservation law— briefly discussed in subsection 6.4.1 as a future research direction.

Thus, neurofunctional theory is derived from a rich body of existing work in the cognitive sciences that incorporates principles from fields such as computational learning theory, psychology, neurobiology, and linguistics. In general, this theory aims to integrate existing models, theories of neurobiological systems. In this report, these models and theories are presented as evidence, which is integrated to justify the development of a new theory. The majority of the evidence presented in the present work focuses on the neurocomputational underpinnings of learning and behavior, and specifically motor adaptation. In chapter 4 and chapter 5, experimental results are presented and later discussed from a neurofunctional perspective chapter 6. The semantic relationship between computation and neural adaptation is a key takeaway from this dissertation— the interrelated history of these interdependent concepts is summarized in chapter 2. Moreover, relevant experimental results are presented in chapter 4 and chapter 5. In chapter 6, this topic is further expounded upon, and the significance of the experimental results are briefly described.

1.2 Neural differentiation: Striatal cholinergic interneurons

In chapter 4, neurocomputational techniques are utilized to test the hypothesis that population-specific characteristics of striatal cholinergic interneurons— often termed *tonically active neurons* (TANs)— can subtly influence the timing of striatal reinforcement-based encoding of reward. These unique neuronal characteristics are the result of neural differentiation— a process that occurs during early development— that results in maturation of neuronal stem cells into specialized neural species (e.g. ion channels, mechanical/chemical receptors). A complete report of this work has been published and is publicly available as an open-access journal article [Kim et al., 2019].

1.3 Emergence of large-scale networks: Cortico-cerebro-striatal pathways

In chapter 5, neurocomputational techniques are employed to investigate the neurophysiological mechanisms that could be responsible for coordinating between distinct learning strategies during motor adaptation— namely, the existence of a cortico-cerebro-striatal pathway that could be responsible for mediating between error-based learning in the cerebellum and non-error-based reinforcement learning in the basal ganglia. A complete report of this work has been published and is publicly available as an open-access journal article [Todorov et al., 2019].

2

LITERATURE REVIEW

Here, the relevant literature is reviewed to provide sufficient evidence to justify the development of a new neurofunctional theory. Further, this review summarizes the essential scientific concepts to ensure that the present work is accessible to a non-expert audience. The review culminates in a description of the specific aims of this project, which focus on two of the author's publications that serve to illustrate the utility of computational methods to neurofunctionally define neural substrates with respect to specific motor adaptation tasks.

In the following sections, historical examples of relevant experimental protocols and theoretical paradigms are summarized. Each historical example is seated in the context of neurofunctional theory. Further, each section includes an explicit statement explaining how the discussed scientific innovation relates to the author's publications (full-text in chapter 4 & chapter 5).

2.1 Early perspectives: The evolution of learning theory

This section introduces a chronological narrative that summarizes important events across several scientific periods (1900-2000s) that shaped the development of modern learning theory. A plethora of experimental methods and scientific theories that have proven essential to the study of learning were developed during the discussed historical eras. Perhaps surprisingly, many of these important theories and techniques were not formalized— at least in the literature— until the late-1800s/early-1900s.

At the dawn of the 20th century, physiologists increasingly sought to elucidate the mechanisms underlying behavior, introspection, and learning through the development and application of new scientific methods, many of which left a lasting impact on the field. A particularly important consequence of these scientific innovations— including the formalization of experimental protocols— was the standardization of experimental measurements of behavior, which enhanced interdisciplinary collaboration and eased the interpretation of behavioral experiments.

The behavioral experiments during this era (early 1900s) began to address new problems that were previously intractable— e.g. the classification of distinct behavior patterns into categories, which scholars accomplished by inventing new quantitative methods for precise comparison between experimental groups (i.e. population statistics). Spearman's rank correlation coefficient is a particularly notable statistical method developed during this era [Lovie and S, 1996]. These and other innovations were essential to the evolution of cognitive science and the formalization of modern learning theory.

2.2 Behavioral conditioning

Behavioral conditioning (i.e. "conditioning") is a fundamental principle of modern learning theory, which can be summarized as, "any learning procedure that involves behavioral modification through the presentation of stimuli."

As practical terminology, "conditioning" often refers to a set of experimental procedures that attempt to modify (e.g. strengthen/weaken) internal associations between the presented stimuli, environmental cues, and specific behaviors.

The concept of behavioral conditioning plays an important role in the semantic structure of neurofunctional theory. Moreover, this brief explanation of "conditioning" is included to assist the reader in understanding behavioral conditioning more broadly and in specific contexts— e.g. the behavioral psychology movement in the mid-20th century. In section 2.5, these nuances are introduced in the context of specific scientific discoveries. Throughout this dissertation, the historical narrative serves as a natural segue to introduce the author's publications as described in the Introduction chapter 1 and Specific Aims chapter 3.

2.3 Reinforcement & conditioning

"Reinforcement," another important concept in learning theory, is a learning process that occurs when an experimental subject receives a rewarding stimulus as a consequence of performing a specific action, thereby increasing the probability that the subject will repeat that specific behavior [Schultz, 2015]. "Reinforcement learning" (RL) is then the process of repeated reinforcement to induce behavioral modification by strengthening or weakening (i.e. "conditioning") the subject's internal reward-behavior associations.

The concept of "reinforcement" originates with the work of the psychologist Edward Thorndike, who is best known for his contributions to educational psychology and learning theory. Thorndike conducted behavioral experiments to test the ability of dogs, cats and other animals to escape from "puzzle boxes." In these experiments, fasted animals were placed in an enclosure from which they could escape— e.g. by pressing a lever. A food reward was also placed outside the enclosure [Thorndike, 1898]. Moreover, Thorndike speculated that animals must form "...associations of sense-impressions and ideas with impulses to act, muscular innervations," thus predicting the existence of an internal biological mechanism by which animals can associate specific behaviors— e.g. the series of muscle activations necessary to press a lever— with the expectation of receiving an intrinsically rewarding stimulus (e.g. food). This idea paved the way for further investigation into "association" and "reinforcement" as psychological— and eventually neurobehavioral— concepts.

The work of Thorndike and contemporaries such as the Russian physiologist Ivan Pavlov inspired the next generation of psychological endeavors, including Watson's fear-conditioning experiments [Watson and Rayner, 1920] and B.F. Skinner's formalization of "operant conditioning," a type of associative learning procedure in which reinforcement and punishment are used to modify the strength of a behavior. These ideas also play an important role in framing neurofunctionalism, especially in the context of the reversal-learning and learning-perturbation experiments discussed in chapter 4 and chapter 5.

2.4 Classical conditioning

Despite being best known for his contributions to educational psychology, the Russian physiologist Ivan Pavlov was— like Thorndike— originally interested in animal behavior and physiology. Pavlov notably discovered the "conditioned reflex" (i.e. "reflex at a distance") as a result of his experiments in which dogs were trained to salivate in response to the ringing of a bell. These experiments led to the development of "classical conditioning," a learning procedure in which an unconditioned stimulus (e.g. food) is paired with a conditioned stimulus (e.g. ringing of a bell) to elicit an unconditioned reflex/response (e.g. salivation) [Pavlov, 2010].

Despite some conceptual overlap, the distinction between classical and operant conditioning is a key nuance. In classical conditioning, a previously neutral stimulus— one that would not ordinarily trigger a reflex— is paired with an unconditioned stimulus that naturally elicits an unconditioned reflex— an innate, involuntary response. This pairing causes the previously neutral stimulus to become associated with the unconditioned stimulus, and thus is now a conditioned stimulus that does elicit the unconditioned response. In contrast, operant conditioning induces behavior modification by reinforcing voluntary actions (e.g. arm-reaching movements) through the presentation of innately appetitive (e.g. a pleasant sound) or aversive (e.g. an unpleasant sound) stimuli, which respectively serve as rewards or punishments to strengthen or weaken a specific behavior.

2.5 Behaviorism

In the 1920s and '30s, Thorndike and Pavlov's successors sought to establish more concrete theoretical frameworks of learning and behavior. The psychologists John Watson and B.F. Skinner formalized several key hypotheses and experimental observations, from which the psychological theory of behaviorism was born.

In his "Behaviorist Manifesto," Watson describes the fundamental tenets of behaviorism, stating that the study of psychology should be purely objective— and therefore should not attempt to interpret internal cognition or "introspective processes," as an individual's internal cognitive experience cannot be empirically observed. Moreover, the "Manifesto" notably states that psychology should aim to predict and control behavior, rather than to describe mental states [Watson, 1913]. Watson also stated his belief that human behavior is no different than other animal behavior— only differing in complexity. Although much of the core philosophy would later be challenged, Watson and Skinner's behaviorism spawned innovative new methods and psychological concepts.

Notably, Skinner expounded on Pavlov and Thorndike's theories of learning and conditioning. Whereas classical conditioning focused on learned associations between specific involuntary behavioral responses (i.e. "stimulus-response" pairs), Skinner developed a new experimental procedure, termed "operant conditioning," which incorporated voluntary actions, i.e. "operants." Importantly, Skinner defined "operants" as emergent behaviors that can be independent of any external stimulus. Therefore, as units of learning, operants have the key advantage of generalizability— especially in comparison to Pavlov and Thorndike's stimulus-reflex pairs, which require the context of a specific stimulus in order to be studied [Skinner, 1958].

Because of this shift in focus to a more generalizable framing of behavior, operant conditioning continues to be an important part of behavioral science. As an experimental procedure, operant conditioning is adaptable and leads to interpretable results across a range of behavioral and environmental learning contexts— the dependence of learning on environmental contexts became an important topic toward the development of behavior therapy, which is discussed in further detail below [Antony and Roemer, 2003]. Moreover, the formalization of operant conditioning enabled the behaviorists to study nuanced learning characteristics including the temporal dependence of learning rate on the timing of reward presentation— these reward-timing protocols were termed "reinforcement schedules" [Skinner, 1969].

Other notable contributions of behaviorism include the concept of "chaining," a concept formulated by Skinner to explain complex behavior as a combined sequence of simpler behaviors, which could be viewed as an extension of Watson's— originally, Darwin's— assertion that human behavior is simply animal behavior, only more complex [Skinner, 1969].

The development of cognitive-behavioral therapy (CBT)— a common type of psychological intervention that involves challenging and altering unhelpful thoughts, cognitive distortions, and behaviors to improve mental health— is possibly the longest-lasting applied psychological technique that originates with the work of behaviorists, namely "behavior therapy," which was heavily based on results from Watson and Rayner's emotion conditioning experiments in the 1920s [Corsini et al., 2008, Watson and Rayner, 1920]. Behavior therapy was a necessary step toward the development of CBT. However, the development of modern CBT was heavily influenced by cognitive psychology, a field that competed with and eventually replaced behaviorism as the psychological mainstream during the "cognitive revolution" of the 1970s and early '80s.

As mentioned in chapter 1, many of the core tenets of Behaviorism play important roles in the semantic structure of neurofunctional theory. The "reinforcement schedule" is a particularly notable concept. In the following sections, adaptive reinforcement/reward scheduling is described in specific experimental contexts. Moreover, adaptive learning is central to the investigations detailed in the author's included publications.

2.6 Relational frame theory

Although the mainstream focus on behaviorism waned in favor of cognitive approaches, psychological theory continues to be shaped by behaviorist influences. One such theory, relational frame theory, extends some of Skinner's ideas to account for nuances of verbal learning that had previously eluded behaviorists. In relational frame theory, responses can be formulated and interpreted relative to the function of a stimulus— i.e. a word (stimulus) in English can have a different meaning (function) depending on the context [Hayes, 1991].

As a concept, the contextual dependency of language functionality has found applications in functional analytic psychotherapy (i.e. functional analysis), a type of psychotherapeutic approach in which the clinician identifies "clinically relevant behaviors" during therapy sessions that could be important for understanding the client's interpersonal relationships in the real world [Tsai et al., 2009]. Furthermore, the analysis of "function" is an important component of the present study.

The analysis of "function" is a particularly important concept in modern learning theory. (chapter 1). In mathematical terms, a function is defined to be an operation that maps any number of inputs to a single output. Similarly, in neurofunctional theory, a function or neurofunction can be defined in terms of a causal hierarchy— i.e. mechanism(s)— that define the dynamic relationship between a physiological input (e.g. taste sensation) and a neurobehavioral output (e.g. feeding behavior).

2.7 The cognitive revolution

By most accounts, the "cognitive revolution" began in 1959 with Noam Chomsky's scathing review of Skinner's book Verbal Behavior [Chomsky, 1959]. Skinner intended the book to address linguistics and language acquisition from a behavioral perspective, topics that had previously been ignored by behaviorism— primarily because the formation of language is not easily interpretable as a sequence of behaviors. Chomsky and other cognitive psychologists, including the French developmental psychologist Jean Piaget, observed that the human capacity for creating or learning new vocabulary is infinite and independent of any stimulus conditioning— e.g. humans demonstrate the ability to speak and understand sentences that they have never heard before [Piaget et al., 1980]. To account for human language ability, an analysis of internal learning processes is necessary— this focus on introspection is incompatible with the core principles of behaviorism. Overall, Chomsky's criticisms of behaviorism focused on the lack of experimental evidence for many of Skinner's claims. Moreover, Chomsky argued that the cited experiments failed to generalize to human verbal learning— in particular because no animals are known to possess comparable linguistic skills.

The cognitive revolution gave rise to many ideas that continue to influence the cognitive sciences today. One of these ideas is the "modularity of the mind," which states that distinct systems of the mind must cooperate to generate thought or a sequence of actions— i.e. distinct processes have different specific missions that coalesce to accomplish a common goal [Fodor, 1983]. Another key idea in cognitive psychology is "innateness," which surmises that humans must be born with some innate ability to learn language without any formal teaching— a concept that was introduced by Chomsky to explain how humans could know so much with relatively limited sensory input [Kasher, 1998].

Along with these new ideas, the cognitive revolution spawned a renewed interest in scientific questions related to introspection and cognition— e.g. "What is the origin, nature of thought?" These questions were explored by careful documentation of clinical observations, case studies, and new experimental approaches that focused on understanding the internal processes that held the keys to understanding the complexities underlying human cognition. Despite the surplus of new psychological approaches, one of the most influential of these new techniques only came to fruition as a result of advances in mathematics and engineering— the use of computational experiments and the birth of Artificial Intelligence (AI).

2.8 Early AI: Analytical linguistics & cognition

In the mid-20th century, several globewide socio-political events— e.g. the Second World War, Vietnam War, Korean War, Cold War— and monumental scientific discoveries— e.g. Watson and Crick's discovery of DNA's double-helix structure— catalyzed the study of computer automation, resulting in numerous breakthroughs. During this period, applied computational research was profoundly influenced by the cognitive sciences— namely psychology and computational neuroscience [Miller, 2003].

Particularly of note are Noam Chomsky's contributions to the linguistic theories of "universal grammar" and "transformational-generative grammar." The former establishes the genetic component of the human language faculty. In the latter, verbal syntax is analyzed as a structured ruleset (i.e. language) from which word combinations are generated to produce meaningful sentences, and via the application of "operations" (i.e. transformations) new sentences can be produced in the same language. These linguistic concepts are often cited as being key inspirations for the development of some of the earliest computer programming languages— including LISP, LOGO, and TeX [Knuth, 2002]. In a sense, the development of human-readable programming languages was bidirectionally dependent on cognitive— especially linguistic— sciences.

Neuroscience is another of the cognitive sciences that influenced early research efforts in AI. In particular, the formulation of the first computational neuron model by McCulloch and Pitts— and the electrophysiological experiments by Hodgkin and Huxley that later yielded the conductance-based Hodgkin-Huxley model— guaranteed that the earliest computational theories and implementations of AI were correlated with— if not caused by— advances in neuroscience [Hodgkin and Huxley, 1952, McCulloch and Pitts, 1943].

2.9 Early AI: General problem solving

Within this initial ”proving period” for AI, the socio-political influences of the second World War significantly catalyzed the academic focus on AI. As part of the war effort, national governments increased funding for AI research, which centered on gaining a deeper understanding of the internal processes that underlie human cognition and learning. The intended applications of this AI research included the optimization of military personnel training, and the creation of computer-automated general problem-solving systems, specifically for use in codebreaking to decrypt enemy communications, and for statistical optimization of battlefield decision-making, e.g. predicting enemy movement on the battlefield [McCorduck, 2004]. Of note are Alan Turing’s many contributions to AI, including neuromimetic implementations of computer automata and the popularization of reaction-diffusion systems, which he modeled as time-coupled systems of partial differential equations [Turing, 1951].

During the 1940s and early ‘50s, applied outcomes in AI research were mostly limited to the development of “expert systems” that could solve well-defined, highly constrained problems— e.g. the German tank problem, an example of frequentist statistical inference being used to estimate the discrete population distribution [Ruggles and Brodie, 1947]. The system itself could be defined as a computational algorithm that takes as input the observed frequency distribution, then performs a series of computations that produce an output relevant to the stated problem.

Another canonical example of an artificial learning system from this era is the General Problem Solver, a computer program implemented in the late 1950s [Newell et al., 1959]. The General Problem Solver was in part motivated by a desire to characterize human problem solving in terms of algorithmic procedures. The resulting computational experiments concluded that two generic forms of problem-solving exist— planning and means-ends analysis. The former takes advantage of human-like abstract reasoning, which offers interpretability at the cost of computational tractability. In contrast, means-ends analysis is essentially an iterative optimization technique that leverages sensory feedback to constrain the solution— this technique is useful for AI systems but is typically impractical as a human reasoning method.

2.10 Modern AI: Specialized (neuro-)computational models

Up to this point in the history of learning theory, AI research sought to characterize and replicate algorithmic, procedural mechanisms of learning and behavior. Much of this research focused on abstract thought experiments with little-to-no consideration of biophysics or practical implementation. Moreover, some AI research during this period was essentially an engineering effort moreso than scientific endeavor— i.e. was heavily goal-oriented, focused on specific predetermined outcomes (e.g. military, educational applications). But by the 1960s and '70s, the study of AI— and computer science in general— began to shift away from toy models and the military industrial complex. Instead this new generation of computationalists expounded on existing cognitive and behavioral theory, posing specific experimental hypotheses via specialized computer simulations.

2.11 Modern AI: Temporal difference learning

In the early 1980s, specialized computational experiments increasingly focused on the characterization of biological reinforcement learning (RL) experiments, generally adhering to an experimental architecture similar to Skinner's operant conditioning protocols. These computational RL experiments attempt to replicate the ability— that many organisms possess innately— to learn a pattern despite incomplete knowledge of the pattern-generating system— i.e. "model-free learning." As in operant conditioning, the learning procedure can be summarized as, "Generate, test, repeat." This distinguishing characteristic of RL directly contrasts with "supervised" learning— e.g. General Problem Solver— in which learning occurs by direct comparison between the "current pattern" and the "target pattern" [Sutton, 1984].

An important conclusion of these computational RL simulations is that "model-free" predictive learning is made possible by comparison between temporally successive predictions, termed "temporal-difference methods" [Barto et al., 1981, Sutton, 1988, Sutton, 1984]. Moreover, Barto et al. noted some key limitations to temporal-differencing as a method for predictive learning. RL systems that rely on temporal-differencing do not require prior knowledge of the pattern-generation system. However, these simple RL systems can only learn when explicitly provided with a complete map of pattern-reward pairings— i.e. a complete prior understanding of the reward to be received for each pattern— in the literature, this reward-pattern mapping is referred to as a "teacher" signal [Sutton and Barto, 1981].

Despite this limitation, computationalists demonstrated that temporal-differencing methods can provide the advantage of computational efficiency— e.g. reduced memory requirements, improved accuracy, convergence speed— over the existing supervised learning algorithms [Sutton and Barto, 1981]. These discoveries found lasting utility in logistical control processes— enabling global solutions to previously intractable optimal control problems— such as real-time optimization of supply chain management and processes related to industrial manufacturing [Barto et al., 1982, Bowersox and Closs, 1989]. Beyond these engineering applications, the computational learning experiments of the '80s and '90s inspired a wave of multidisciplinary collaboration in the cognitive sciences, yielding a plethora of contributions to learning theory, creative new methodologies, and computational network architectures.

2.12 Modern AI: Ensemble methods

A notable product of this multidisciplinary research effort is the advent of "ensemble learning methods," a type of computational learning algorithm that combines procedural elements of several distinct learning approaches, often characterized by some form of "meta-learning." As an example, a supervised learning algorithm could optimize the amount or timing of reward (i.e. the reward schedule) to maximize accuracy with respect to a statistically-defined distribution of reinforcement learning contexts [Schiff et al., 1996].

Ensemble methods aim to overcome the known disadvantages of a particular learning technique by combining multiple algorithms that possess complementary advantages and disadvantages. This can be accomplished by quantitatively updating model parameters, leveraging heterogeneous statistical information shared between distinct, nested learning processes. For example, consider a hypothetical scenario in which two individuals ("Steven", and "Rachel") cooperate to find a route to the library by relying on communication, thereby accounting for any empirical deficits of each individual learner.

In this scenario, "Steven" is colorblind, and "Rachel" is dyslexic. Steven and Rachel are given a map that contains the necessary information to navigate to the library. Unfortunately, the map is not to-scale and distinguishes between roads by color— i.e. Steven cannot determine the extent of individual roads. Furthermore because the map is not to-scale, navigation is only possible by matching the street names on the map with street signs along the route— i.e. alone, Rachel cannot accurately determine when she should make a turn. Despite these obstacles, the pair can still successfully navigate the route by combining forces. Before setting out, Rachel could identify the road extents, then draw a symbolic representation in black-and-white to assist Steven's interpretation of the map. Steven could then confirm the street names that correspond to turns along the route. Individually, neither Rachel nor Steven could use this unconventional map to reach the library. But through cooperative communication, the pair manage to integrate their heterogeneous observations (e.g. street names, extents) to effectively function as an ensemble.

2.13 Modern AI: Neurocomputational learning models

The specialization of computational learning experiments and the development of ensemble methods are among the most notable outcomes produced by this era of multidisciplinary collaboration in cognitive science, effectively bridging theoretical gaps between cognitive psychology, neuroscience, biophysics, and computer science. These specialized computational experiments of the late-70s, ‘80s and ‘90s were predicated on previous work by the behaviorists in the early-to-mid-20th century, and subsequently the linguistic theory developed during the cognitive revolution. Notably, the multidisciplinary influences in the cognitive sciences became increasingly interdependent, with bidirectional information flow between experimental biologists and computationalists [Barto, 1995, Sutton and Barto, 1981, Williams, 1992].

Thus, computational modeling techniques gained newfound utility— computationalists could design simulations to make statistically sound experimental predictions. Specifically, these virtual learning simulations could be interpreted to predict whether or not a specific neurocomputational mechanism could account for an empirically observed behavior or learning characteristic. These computational predictions garnered the attention of experimental neurobiologists, who could then conduct their own experiments to validate any functional mechanisms that could be inferred by computationally replicating experimental observations. In this way, new avenues in the neurobiological study of learning continue to be directed by previously abstract concepts in computer science and control theory.

2.14 Modern AI: Adaptive learning systems

The collaborative atmosphere among cognitive scientists during the 1980s and ‘90s spawned an assortment of advances in learning theory— namely the exploration of the neurocomputational mechanisms underlying animal intelligence, and the implementation of corresponding neuromimetic controller architectures. Of these neuromimetic control systems, the “associative search network” is particularly relevant to the present work. The “associative search network" is an artificial learning system composed of adaptive components— some of which provide specialized RL capabilities— that extends temporal-differencing to enable robust learning in contexts that were previously computationally intractable [Barto et al., 1981].

Significantly, "associative search networks" can learn without explicit prior knowledge of the pattern-reward mapping (i.e. "teacher") while being robust to statistical noise, i.e. unexplained variability). In contrast to simple temporal-differencing, the associative search network— designed as a closed-loop— can rely on limited knowledge of the reinforcement landscape that is gained during exploration of the environment. This is possible by leveraging specialized network components that can each perform a different learning strategy. Thus, the network can perform an "adaptive search." The system can choose among multiple learning strategies before performing an action, then intelligently encode feedback (i.e. punishments, rewards) relative to a combination of the current environmental context and the chosen learning strategy. Thus, the network can leverage hierarchical, contextual memory of the environment, enabling adaptation to variable pattern-reward schedules— or noise— and informing the selection of future learning strategies. Moreover, this enables the system to intelligently strategize in real-time [Barto and Sutton, 1981, Sutton and Barto, 1981].

2.15 Computational neuroscience: Quantification of neurofunctional substrate

Although the exploration of adaptive learning in silico— e.g. the use of computer simulations— arguably provided limited empirical evidence, these virtual learning experiments notably identified a key gap in understanding— "What latent neural mechanisms or neurofunctional units can explain the unparalleled learning capabilities of neurobiological systems?"

Despite limited experimental evidence, neurobiologists of the era generally accepted that individual cells could perform simple associative learning— e.g. Pavlovian stimulus-response associations in classical conditioning. Pavlov himself speculated that self-organizing neuronal populations in the cerebral cortex could be responsible for these simple associative learning tasks [Pavlov, 2010]. These neurobiological insights inspired further speculation as to the mechanisms that populations of neurons utilize to perform more complex adaptation tasks. Hebbian learning theory is cited as another key inspiration— a theory often summarized as, "Cells that fire together wire together," which implicates synaptic plasticity as a potential mechanism for associative learning in neuronal populations.

Klopf's "Heterostatic Theory" was an essential impetus for reinvestigating the role of neuronal differentiation in adaptation and learning. Klopf observed that individual neurons can have specialized biophysical characteristics— e.g. differentiated receptor types, ion channels, dendritic morphology. Moreover, Klopf speculated that interactions between heterogeneous populations of specialized neurons could explain the sophisticated adaptive learning faculties demonstrated in vivo by neurobiological systems [Klopf, 1979, Klopf, 1972]. These hypotheses heavily influenced Sutton and Barto's design of the "associative search network" with specialized components— i.e. the "adaptive search" and "adaptive critic" elements— that coordinate as a sophisticated computational ensemble to enhance the robustness of learning by adaptation.

Thus, despite limited empirical evidence, a sufficiently complex neural control system such as the one introduced by Barto et al. could play a role in neurobiology, coordinating between distinct learning strategies. Such a mechanism could operate by leveraging large-scale inter-network pathways that connect distinct populations of specialized neurofunctional subsystems (e.g. dopaminergic neurons in the striatum, nuclei of the cerebellum). Such a neural control system could be interpreted as the neurophysiological analogue of the "associative search network" (i.e. "adaptive critic") [Barto, 1995]. Initially these speculations were regarded with caution by mainstream cognitive science, as little empirical evidence existed to pinpoint a specific neural substrate [Houk and Barto, 1992, Klopf, 1982, Wickens, 1980].

2.16 Computational neuroscience: Recent work

This section ties together the experimental and computational results of the early and mid-20th century by describing more recent findings that are particularly important to the present work.

2.16.1 Basal ganglia: The neurofunctional role of dopamine in RL

By the mid '90s, significant experimental evidence demonstrated that the basal ganglia play a key neurofunctional role in operant reinforcement— i.e. the basal ganglia was shown to be involved in RL [Aosaki et al., 1994b, Barto, 1995, Graybiel, 1995, Houk and Wise, 1995, Mirenowicz and Schultz, 1996, Montague et al., 1996, Schultz et al., 1997, Toni and Passingham, 1999]. Dopamine activity in the basal ganglia was correlated with the timing of subjects receiving a reward. From this evidence, it was concluded that (reinforcement) learning (RL) occurs in the basal ganglia within dopaminergic populations. This system continues to be studied as a key site of RL. Moreover, this system can be framed from a neurocomputational perspective as a neurobiological analogue of the temporal-difference learning method proposed by Barto and Sutton, et al. [Barto, 1995].

Electrophysiological evidence further confirmed that dopamine neurons can adapt their firing pattern to optimize the reward expectation, essentially predicting the magnitude and timing of future rewards. The corresponding neurofunctional concept is termed the "reward prediction error" (RPE), now a fundamental principle in modern RL research. In computer models of the basal ganglia, the RPE is calculated as the difference between the expected and actual reward received [Schultz et al., 1997]. Importantly, the RPE is computed as a directional error that indicates whether the current expectation of reward is higher or lower than the actual reward value.

Furthermore, in correspondence with Kopf's "Heterostatic Theory," the basal ganglia's learning neurofunctionality depends on the heterogeneous activity of a variety of different neuronal species— e.g. striatal cholinergic interneurons [Aosaki et al., 1994b, Klopf, 1972].

2.16.2 Basal ganglia: Neurocomputational models

Many computational RL models that focus on specific neural dynamics in the basal ganglia have since been developed. Some have focused on engineering milestones— e.g. efficient optimization of learning parameters for use as a robotic control system [Doya, 1996]— while other computational models quantitatively replicate behavioral phenomena [Suri and Schultz, 2001], or quantitative biophysical relationships inferred from experimental data [Doya, 2002, Frank, 2006, Frank et al., 2007, Franklin and Frank, 2015a, O'Reilly and Frank, 2006].

In relation to the present work, the author co-authored a publication that describes a neurofunctional model of the basal ganglia [Kim et al., 2017a]. The initially published version of this model incorporates essential features of the direct and indirect pathways, which are functionally distinct mechanisms of RL in the basal ganglia. By incorporating both pathways, this neurofunctional model replicates essential adaptive learning properties observed in behavioral experiments— specifically, the indirect pathway functions to suppress previously encoded reward-pattern associations [Kim et al., 2017a].

Further biophysical details of this model are described in a 2017 publication by Teka et al., in which a computational model of neural motor control is defined. This biomechanical model produces goal-directed reaching movements by computationally replicating the motor control pathway from the motor cortex— through spinal cord circuits— to the muscles, which produce arm movements [Teka et al., 2017a].

Importantly, this biomechanical arm-reaching model is a key component of the neurocomputational models developed in the current study. The model defines a neurophysiologically accurate biomechanical arm that incorporates neurofunction-relevant features of motor control, including direction-specific motoneurons in the primary motor cortex (M1), which serve as cortical controllers that produce directionally-tuned motor commands to the muscle fibers, and receive afferent feedback from the muscles via the spinal cord. Moreover, the authors made conclusions regarding the role of specific motor control circuitry— namely that motoneurons and afferent feedback modulate the directional tuning behavior of cortical neurons to produce functional goal-oriented reaching behavior. The incorporation of this biomechanical model can therefore provide additional biophysical detail that could elucidate essential dynamical characteristics of real-world motor adaptation tasks.

This computational neurofunctional framework is further expounded upon in the middle chapters of the current report— the author's included publications.

2.16.3 Cerebellum: A neural substrate for error-based learning

Interest in motor learning was not limited to RL in the basal ganglia during the '90s and early 2000s. A mounting supply of new empirical evidence implicated the cerebellum as not only being involved in low-level motor control but also as a sophisticated neurofunctional learning pathway. Before the discoveries of the '90s, neurobiologists and computationalists had previously speculated the cerebellum was primarily a simple feedforward controller that integrated sensory feedback to produce real-time neural motor control via synapses with descending projections from motor cortex [Ivry et al., 2002, Kheradmand and Zee, 2011, Paulin, 1993].

However, the most recent consensus holds that the cerebellum is involved in more sophisticated motor adaptation and learning tasks [Bastian, 2006, Caligiore et al., 2016a]. Current experimental challenges center on establishing a complete neurofunctional description of the highly interconnected neural pathways in the cerebellum. Experimental evidence shows that the cerebellum can function as a closed-loop control mechanism that adaptively "corrects" descending motor commands by leveraging sensory feedback [Bastian, 2006].

2.16.4 Cerebellum: Computational models

Numerous existing computational models describe some functional supervised learning capabilities of the cerebellum [Doya, 2000a, Houk et al., 1996, Miall et al., 1993]. These models vary in their level of detail and the specific adaptive properties that are described.

Moreover, computational models of cerebellar learning increasingly focus on the cerebellum's ability to adaptively relinquish or acquire control of descending motor commands depending on the degree of certainty that a particular motor adaptation task is tractable using error-based learning [Caligiore et al., 2016a, Todorov et al., 2019].

2.16.5 Neural coordination of distinct learning strategies

Recent evidence supports the notion that a neurofunctional mechanism could explicitly coordinate between RL in the basal ganglia and supervised learning (error-based learning) in the cerebellum to intelligently alternate between these functionally distinct learning strategies [Caligiore et al., 2016a]. A common hypothesis— supported by a growing body of evidence— proposes the existence of a cerebral-cortico-striatal closed-loop neural circuit from which large-scale, inter-network neural communication pathways could emerge [Caligiore et al., 2016a, Doya, 2000a, Houk and Wise, 1995, Todorov et al., 2019]. Such a circuit could play a role in producing generalized motor adaptation, a faculty that humans and other complex neurobiological organisms possess innately. For a visual diagram of the proposed circuit, see 1 below.

Moreover, behavioral experiments have been conducted to elucidate the distinct learning strategies that could be involved during different motor adaptation tasks, some of which have been further validated by neurocomputational models that replicate essential neurofunctional features of the behavioral data [Doyon et al., 2003, Galea et al., 2011, Gao et al., 1996, Hikosaka et al., 2002, Penhune and Steele, 2012]. This line of thought is further expounded upon in the following chapters.

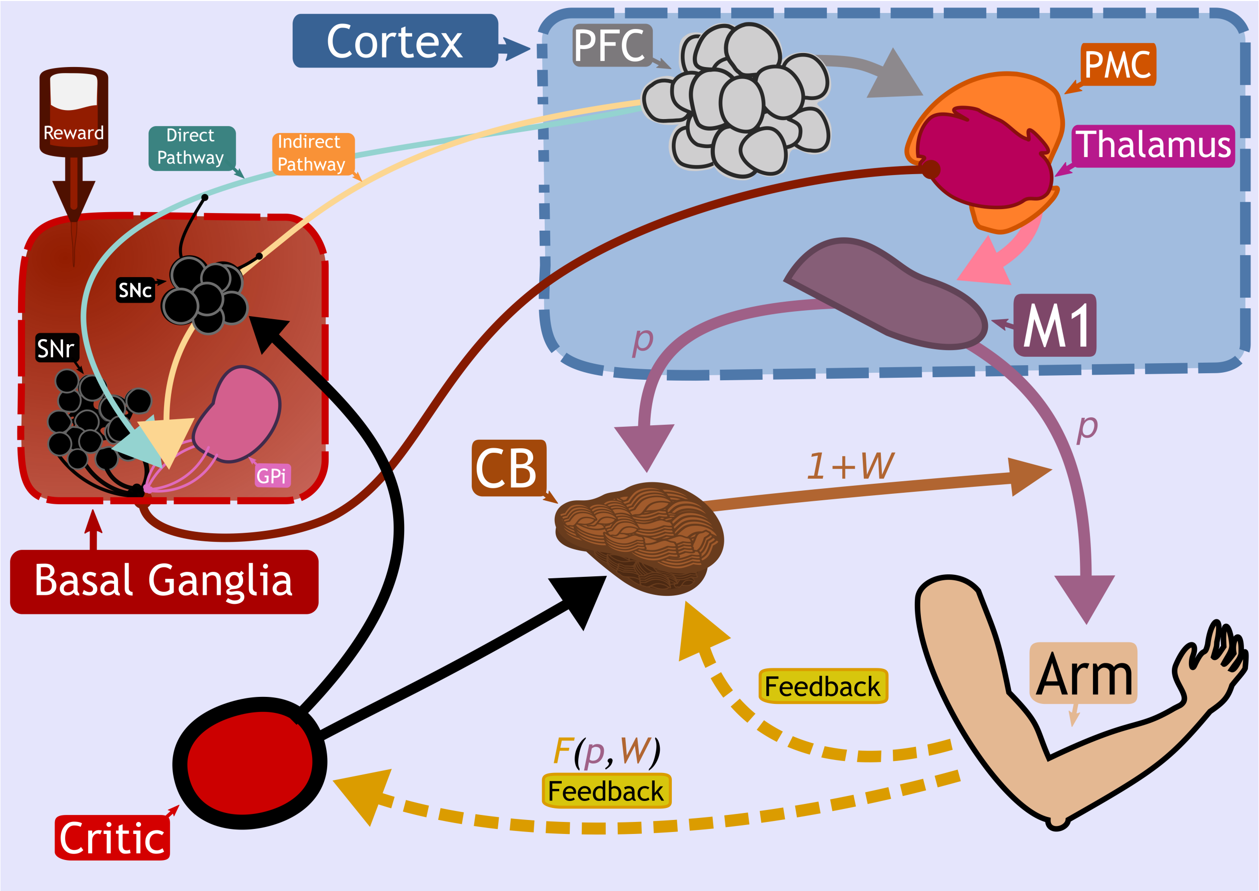

Figure 1 Semi-artistic network diagram of the complete motor adaptation model. Depicted is the proposed neurofunctional substrate that is hypothesized to coordinate between distinct motor adaptation strategies. The relevant substrate includes cortical structures such as the prefrontal cortex (PFC), the pre-motor cortex (PMC), thalamus, and the primary motor cortex (M1). As described in [Kim et al., 2017a], the PFC provides integrated sensory signals to the basal ganglia. The basal ganglia then further integrates this cortical input with dopaminergic reward to enable a reinforcement learning cascade. Also included in the diagram are the cerebellum (CB) and the "critic" (i.e. "adaptive critic"). The CB adaptively corrects the motor commands provided by M1 to the musculature ("Arm"), effectively combining cortical input with sensory feedback from the brainstem. The integrated motor program descends via the spinal cord. In parallel, the critic adaptively switches between distinct learning mechanisms by selectively facilitating or depressing the basal ganglia (reinforcement learning)and CB (model-based/error-based learning).

3

SPECIFIC AIMS

The specific aims of this project are described in this section, and experimental solutions for each aim are summarized. In each experiment, neurocomputational models are constructed to quantitatively define the neurofunctional mechanisms underlying specific motor adaptation processes.

3.1 Investigate the neurofunctional role of striatal cholinergic interneurons in reinforcement learning.

In this specific aim, the goal is to determine the neurofunctional role of striatal cholinergic interneurons (TANs) in the striatum (a nucleus of the basal ganglia). For this investigation, the hypothesis— that TANs can affect the timing and encoding of reward during RL tasks by selectively modulating dopamine release from D2 neurons— is posed by construction of a computational model.

From a neurofunctional perspective, the results of this investigation illustrate an example of neural differentiation, which can subtly influence adaptive learning processes in the basal ganglia. Moreover, the authors validate the underlying neurofunctional hypothesis by constructing a computational model of the relevant neural circuitry— effectively demonstrating the utility of neurocomputational methods in posing complex neurofunctional hypotheses.

3.2 Describe the interplay between cerebellum and basal ganglia during motor adaptation.

For this specific aim, the neurofunctional role of cortico-cerebro-striatal pathways is investigated. The underlying hypothesis for this investigation is that this pathway plays a role in coordinating between distinct learning strategies. This hypothesis is validated by a computational model that is described in further detail in later chapters. These findings could have clinical applications— namely, the development of a robust quantitative method to distinguish between different neurofunctional motor disorders, even those with similar symptomatology.

3.3 Summarize the history, evolution of learning theory.

A specific aim of the literature review is to provide an adequate summary of the history and evolution of learning theory in the 20th century. Specifically, this review is intended to contextualize some fundamental principles of the proposed neurofunctional theory of learning, including the interdependence of behavior and cognition. The review emphasizes the development and utility of important neurocomputational methods— historical context is provided through a description of relevant discoveries, theories, and hypotheses.

Notably, a summary of the cognitive revolution is presented here as an intellectual movement that catalyzed the development of AI and cemented the cognitive sciences as a multidisciplinary field of research that necessarily incorporates psychology, neuroscience, and computation— all of which provide unique perspectives that are required to formulate a complete neurofunctional theory of learning.

Specific computational and experimental discoveries— such as the development of the General Problem Solver and the advent of temporal-difference methods— are summarized to contextualize some distinct neurofunctional learning principles, and to illustrate how these principles can be— and have since been— integrated into modern learning theory. Furthermore, summarization of these innovations also lends itself to contextualizing a variety of important psychological, neurobehavioral, and neurocomputational methods, many of which continue to have lasting impact in the cognitive sciences.

3.4 Describe the significance of a complete neurofunctional theory.

This dissertation aims to introduce a ”complete” neurofunctional theory of learning. Thus, the present work aims to define ”complete” in this context by explicitly referencing historic experimental results that motivate the development of a ”complete” neurofunctional theory— and which illustrate some important advantages of the theory’s semantic structure.

This goal is to some extent accomplished in the literature review, as described in the previous aim (section 3.3). Computational methods— e.g. the adaptive critic— are seated in a neurofunctional context via the literature review. The significance of a complete neurofunctional theory is also described in terms of potential clinical applications— e.g. quantitative prognosis of motor diseases by identification of the neurofunctional characteristics that uniquely define a particular disease (e.g. Parkinson’s, Huntington’s, cerebellar ataxia). The potential applications and general implications of a complete neurofunctional theory are further expounded upon in the Discussion.

4

THE FUNCTIONAL ROLE OF STRIATAL CHOLINERGIC INTERNEURONS IN REINFORCEMENT LEARNING FROM COMPUTATIONAL PERSPECTIVE

4.1 Background & Significance

4.1.1 Background

In this study, we explore the functional role of striatal cholinergic interneurons, hereinafter referred to as tonically active neurons (TANs), via computational modeling; specifically, we investigate the mechanistic relationship between TAN activity and dopamine variations and how changes in this relationship affect reinforcement learning in the striatum. TANs pause their tonic firing activity after excitatory stimuli from thalamic and cortical neurons in response to a sensory event or reward information. During the pause striatal dopamine concentration excursions are observed. However, functional interactions between the TAN pause and striatal dopamine release are poorly understood. Here we propose a TAN activity-dopamine relationship model and demonstrate that the TAN pause is likely a time window to gate phasic dopamine release and dopamine variations reciprocally modulate the TAN pause duration. Furthermore, this model is integrated into our previously published model of reward-based motor adaptation to demonstrate how phasic dopamine release is gated by the TAN pause to deliver reward information for reinforcement learning in a timely manner. We also show how TAN-dopamine interactions are affected by striatal dopamine deficiency to produce poor performance of motor adaptation.

4.1.2 Significance

It is widely accepted that the basal ganglia play an important role in action selection, the process by which contextually appropriate actions are chosen in response to presented stimuli. To determine the appropriateness of an action the basal ganglia perform reinforcement learning to establish action-stimulus associations. This learning process is facilitated by dopaminergic activity in the striatum, where a reward prediction error is encoded by the dopamine concentration excursion from its baseline level. When a subject performs context-appropriate actions, there is a phasic increase in striatal dopamine if the received reward is above the expectation, which means a positive reward prediction error is computed. Over time, the synapses that correspond to appropriate stimulus-action association in the striatal network are strengthened by long-term potentiation, and inappropriate actions are suppressed by long-term depression [Graybiel, 2008, Frank, 2005]. Although this process is well understood from a behavioral perspective, there are still open questions about the underlying neural circuitry.

The neural populations within the striatum consist of GABAergic medium spiny neurons (MSNs), cholinergic interneurons, and GABAergic interneurons [Kita, 1993, Koós and Tepper, 1999, Tepper et al., 2010, Yager et al., 2015, Dautan et al., 2014]. Many previous computational studies have focused on MSNs, which comprise a vast majority of the striatum and are heavily implicated in basal ganglia reinforcement learning [Kreitzer and Malenka, 2008, Wall et al., 2013, Smith et al., 1998]. In contrast, cholinergic interneurons—also known as tonically active neurons (TANs)—comprise a small fraction of the striatal neurons and their functional role is not well understood. In this study, we integrate the results of previous studies into a computational model that includes TANs and highlight their role in propagating reward information during reinforcement learning.

Tonically active neurons (TANs) are so-called because they exhibit tonic firing activity (5 10Hz) [Schulz and Reynolds, 2013, Tan and Bullock, 2008]. TANs receive glutamatergic inputs from the cortex and thalamus [Yager et al., 2015, Ding et al., 2010, Kosillo et al., 2016]. These excitatory inputs convey sensory information during a salient event or the presentation of a reward [Cragg, 2006, Schultz, 2016]. When a salient event occurs, TANs generate a short burst of action potentials, which is followed by a pause in TAN activity for several hundred milliseconds. After this pause, TANs undergo a postinhibitory rebound before returning to normal levels of activity [Aosaki et al., 1994a, Morris et al., 2004, Apicella et al., 2011, Joshua et al., 2008, Schulz and Reynolds, 2013, Doig et al., 2014]. TANs project to various neighboring striatal neurons and affect them by releasing acetylcholine which binds to muscarinic and nicotinic cholinergic receptors present on postsynaptic neurons. Muscarinic receptors are widely expressed in the striatal medium spiny neurons [Galarraga et al., 1999, Franklin and Frank, 2015b]. The nicotinic receptors are present in striatal GABAergic interneurons and axon terminals of the dopaminergic substantia nigra pars compacta (SNc) neurons [Cragg, 2006, Franklin and Frank, 2015b, Shin et al., 2017, Zhang et al., 2018].

The characteristic pause in TAN activity was previously suggested to be important for conveying reward information during reinforcement learning. The TAN pause duration depends on a change in striatal dopamine concentration, which is induced by dopaminergic inputs from SNc [Maurice, 2004, Straub et al., 2014]. This dependence exists because TANs express type 2 dopamine receptors (D2) that have an inhibitory effect on TAN activity when activated [Deng et al., 2007, Ding et al., 2010].

After a stimulus, TANs develop a slow after hyperpolarization (sAHP) that is mainly controlled by apamin-sensitive calcium dependent potassium current (IsAHP). The sAHP lasts several seconds and induces a pause in tonic firing [Bennett et al., 2000, Reynolds, 2004, Wilson, 2005]. Another current, the hyperpolarization-activated cation (h-) current (Ih), is involved in quick recovery from sAHP. Deng et al. showed that partially blocking Ih resulted in a prolonged TAN pause duration, and that Ih was modulated by dopamine primarily via D2 inhibitory receptors [Deng et al., 2007]. Thus, the duration of the TAN pause is modulated by Ih activation, which in turn is dependent on striatal dopamine concentration.

In this study, we revisit previous experimental results to formulate the following interpretations. During baseline tonic firing TANs release acetylcholine, which binds to nicotinic receptors on dopaminergic axon terminals. Thus, during their tonic firing regime, TANs exclusively define the baseline concentration of dopamine in the striatum, independently of the firing frequency of dopaminergic neurons [Rice and Cragg, 2004, Cragg, 2006]. This baseline dopamine concentration corresponds to the expected reward in the determination of the reward prediction error. Furthermore, during the TAN pause, TANs stop releasing acetylcholine, thereby temporarily returning control of striatal dopamine release to dopaminergic neurons. This phasic shift in dopamine concentration corresponds to the received reward; the reward prediction error is represented as the phasic increase/decrease in dopamine concentration from the TAN-defined baseline [Cragg, 2006]. Importantly, this suggests that the TAN pause serves as a time window, during which the phasic release of dopamine encodes the reward prediction error.

In this paper, we introduce a mathematical model of the TAN activity-dopamine relationship that incorporates the sAHP- and h-currents in a rate-based description of the striatal TAN population. In the model, the Ih is modulated by striatal dopamine through D2 receptor activation. Our model provides a mechanistic interpretation of the TAN activity-dopamine concentration relationship; we use our model to elucidate the mechanism by which striatal dopamine modulates the TAN pause duration, and how TAN activity regulates dopamine release. Previously, we implemented a model of reward-based motor adaptation for reaching movements that incorporated reinforcement learning mechanisms in the basal ganglia [Kim et al., 2017b, Teka et al., 2017b]. With that model, we reproduced several behavioral experiments that involved basal ganglia-focused motor adaptation [Kim et al., 2017b]. Presently, we integrate our new model of the TAN-dopamine relationship into our previous reinforcement learning model. We use the integrated model to simulate striatal dopamine deficiency, as occurs in Parkinson's Disease. Even though TANs are known to send cholinergic projections to other striatal neurons, e. g. medium spiny neurons, the model does not account for these projections and focuses exclusively on the implications of interactions between TAN activity and dopamine release in striatum.

4.2 Methods

4.2.1 The model of TAN activity

Our model describes the collective dynamics of a population of striatal tonically active neurons (TANs). The model represents the aggregate firing rate (activity) of the population treated as a smooth function of time t with TAN activity denoted by $V_{TAN}(t)$. The following differential equation governs its dynamics:

$$\tau_{TAN}\frac{dV_{TAN}(t)}{dt} + V_{TAN}(t) = \sigma\left(I_{TAN}(t)\right) \tag{1}$$

where τ_{TAN} is a time constant, $\sigma(x) = \Theta(x) \cdot \tanh(x)$ is a sigmoid function, $\Theta(x)$ is Heaviside's function, and $I_{TAN}(t)$ is a term representing an aggregate input composed of intrinsic current inputs and synaptic inputs to the TAN population:

$$I_{\text{TAN}}(t) = W_{T\text{ hal}} \cdot V_{\text{Thal}}(t) + Drv_{\text{TAN}} + I_{s\mu HP}(t) + I_H(t) \tag{2}$$

Here $V_{\text{Thal}}(t)$ is a thala mic stimulus equal to 1 during stimulation and 0 otherwise, W_{Thal} is a synaptic weight of the thalamic input, Drv_{IAN} is a constant drive that defines the baseline firing rate, $I_{sAHP}(t)$ is a slow after-hyperpolarization cur rent input, and $I_H(t)$ is an h-current input.

The slow after-hyperpolarization current $I_{sAHP}(t)$ is a hyperpolarizing current activated when the TAN activity exceeds certain threshold; the dynamics of this current are defined as

$$\begin{aligned}\tau_{sAHP}\frac{dI_{sAHP}(t)}{dt} + I_{sAHP}(t) = &- g_{sAHP} \cdot \left(V_{IAN}(t) - \theta_{sAHP}\right)\\ &\cdot \Theta\left(V_{TAN}(t) - \theta_{sAHP}\right)\end{aligned} \tag{3}$$

where τ_{sAHP} is a time constant, g_{sAHP} is the activation gain, an θ_{sAHP} is the threshold for activation.

In contrast to I_{sAHP}, the depolarizing h -current $I_H(t)$ activated when the TAN activity is below certain threshold, and its activation is modulated by the dopamine concentration.

Its dynamics are defined by the following equation.

$$\begin{aligned}\tau_H \frac{dI_H(t)}{dt} + I_H(t) = &- g_H \cdot \exp\left(-W_{DA} \cdot [DA](t)\right) \\ &\left(V_{TAN}(t) - \theta_H\right) \cdot \Theta\left(\theta_H - V_{TAN}(t)\right)\end{aligned} \tag{4}$$

where τ_H is a time constant, g_H is the activation gain, W_{DA} is the dopamine weight coefficient, $[DA]$ is the concentration of striata dopamine, and θ_H is the h -current activation threshold.

The temporal dynamics of striatal dopamine are defined by

$$\begin{aligned}\tau_{DA} \frac{d[DA](t)}{dt} + [DA](t) = &[DA]_0 + RPE \cdot \left(1 - \frac{V_{\text{TAN}}(t)}{\theta_{DA}}\right) \\ &\Theta\left(\theta_{DA} - V_{\text{TAN}}(t)\right)\end{aligned} \tag{5}$$

where τ_{DA} is the time constant, RPE is the reward prediction error, θ_{DA} is the nicotinic receptor threshold, $[DA]_0$ is the baseline dopamine concentration.

To calibrate the model, we replicated experimental data published by Ding et al. (2010) who recorded TAN activity from sagittal slices of mice brains while stimulating either thalamic or cortical neurons while blocking D2 receptors with sulpiride or increasing dopamine levels by cocaine (Figure 3). All parameters were tuned to fit the experimental data and their values are listed below:

$$\begin{aligned}\tau_{TAN} &= 20ms, W_{Thal} = 4, Drv_{TAN} = 0.3, \tau_{sAHP} = 700ms \\ g_{sAHP} &= 5, \theta_{sAHP} = 0.3, \tau_H = 700ms \\ g_H &= 20, \theta_H = 0.2, W_{DA} = 1 \\ \tau_{DA} &= 20, \theta_{DA} = 0.01, [DA]_0 = 1\end{aligned}$$

To simulate the effect of sulpiride (Figure 3 B) we set $W_{DA} = 0$ as sulpiride is a selective antagonist of dopamine D2 receptors. To simulate the effect of suppressed dopamine reuptake by cocaine (Figure 3C) we set $[DA]_0$ to three times its control value $[DA]_0 = 3$. We simulated blocking h -current (Figure 3D) by setting $g_H = 0$.

4.2.2 Simulation of behavioral experiments

Integration of TAN-dopamine interactions into the model of reward-based motor adaptation Previously, we published a model able to reproduce key experiments concerned with non-error-based motor adaptation in the context of center-out reaching movements [Kim et al., 2017b]. The model included 3 modules: a 2 pathway (direct and indirect) BG module, a lower level spinal cord circuit module that integrated supra-spinal inputs with feedback from muscles, and a virtual biomechanical arm module executing 2D reaching movements in a horizontal plane [Kim et al., 2017b] for the details). The BG module was responsible for selection and reinforcement of the reaching movement based on reward provided. To study effects of TAN activities on dopaminergic signaling in the striatum, we integrated the model of TAN-dopamine interaction described above into the model of [Kim et al., 2017b]. A schematic of the integrated model is shown in Figure 7.

The model of reinforcement learning in basal ganglia we used in this study was previously published and is described in details in [Kim et al., 2017b]. Here, we only provide short qualitative description. Behavioral experiments studying reinforcement learning mechanisms assume that a choice must be made between several differentially rewarding behavioral options. Unlike decision-making tasks, motor learning does not imply a small or finite number of possible choices. The only constraint is the context of the task, e.g. reaching from a fixed initial position to an unknown destination. Our model has unlimited number of possible actions. As the context, we used center-out reaching movements performed in a horizontal plane. To calculate cortical activity corresponding to different movements, we explicitly solved an inverse problem based on the given arm kinematics. Accordingly, for every possible reaching movement we could calculate the corresponding motor program represented by the activity profiles of cortical inputs responsible for activation of different muscles. To describe different experiments, we define corresponding (arbitrarily large) sets of motor programs that define all possible behavioral choices (actions) in each experimental context.

The classical view of action selection is that different motor actions are gated by thalamocortical relay neurons. In the presented model, we assume that relay neurons can be activated at different firing rates, and their firing rates define contributions of different motor programs to the resulting motor response. More specifically, in our model cortical input to the spinal network is implemented as a linear combination of all possible motor programs in the given context with coefficients defined by the firing rates of corresponding thalamocortical relay neurons. This linear combination can be viewed as an aggregate input to the spinal network from the cortical motoneurons exhibiting activity profiles corresponding to different motor behaviors, e.g. reaching movements in different directions.

The classical concept of BG function is that the BG network performs behavioral choice that maximizes reward. This action selection process results in activation of thalamic relay neurons corresponding to the selected action and suppression of neurons gating other behaviors. Per this concept, each action is dedicated to specific neurons in different BG nuclei. Their focused interconnections form action-related loops which start at the cortex, bifurcate in the striatum into direct and indirect pathways converging on the internal Globus Pallidus (GPi), and feed back to the cortex through the thalamus. Action preference is facilitated by increased excitatory projections from sensory cortical neurons representing the stimulus to direct pathway striatal neurons (D1 MSNs). Suppression of unwanted competing actions is assumed to occur because of lateral inhibition among the loops at some level of the network in a winner-takes-all manner.

In the model, novel cue-action associations are formed based on reinforcement learning in the striatum. Eventually, the preferable behavior is reliably selected due to potentiated projections from the neurons in prefrontal cortex (PFC), activated by the provided stimulus, to D1 MSNs, corresponding to the preferred behavior. In technical terms, the output of basal ganglia model is the activation levels of thalamocortical relay neurons in response to the input from PFC neurons activated by visual cues. Each cure represents one of the possible reaching targets. These levels are used as coefficients of the linear combination of all possible actions which represents the motor program selected for execution. The resulting motor program is used to calculate the endpoint of the movement using neuro-mechanical arm model [Teka et al., 2017b]. Depending on the distance between the movement endpoint and the target position, the reward is calculated as dictated by the experimental context. This reward value is used to calculate the reward prediction error as a temporal difference between the current and previous reward values. The reward prediction error is used as the reinforcement signal (positive or negative deviation of dopamine concentration from its baseline levels) to potentiate or depress synaptic projections from PFC neurons, activated by the visual cue provided, to the striatal neurons, representing the selected actions. See details in [Kim et al., 2017b].

In [Kim et al., 2017b], the reinforcement learning is described as a trial-to-trial change in the synaptic weights of prefrontal cortico-striatal projections as follows:

$$\Delta W_{ji}^1 = \lambda_1 \cdot C_j \cdot D_i^1 \cdot RPE - d_w \cdot W_{ji}^1 \tag{6}$$

$$\Delta W_{ji}^2 = -\lambda_2 \cdot C_j \cdot D_i^2 \cdot RPE - d_w \cdot W_{ji}^2 \tag{7}$$

where: ΔW_{ji}^1 and ΔW_{ji}^2 are the changes in synaptic weights between PFC neuron j and D1$-$ and D2 $-$ MSNsi, respectively, λ_1 and λ_2 are the learning rates, RPE is the reinforcement signal equal to the reward prediction error, C_j is the firing rate of PFC neuron j; D_i^1 and D_i^2 are the firing rate of D1$-$ and D2$-$MSNs i, respectively, and d_w is a degradation rate.

In the integrated model, we assume that learning in the striatum is a continuous process defined by the deviation of dopamine concentration from its baseline value. Therefore, we replace the difference equations above with their differential analogs with reward prediction error replaced with the phasic component of the dopamine level:

$$\frac{d}{dt} W_{ji}^1 = \bar{\lambda}_1 \cdot C_j \cdot D_i^1 \cdot ([DA](t) - [DA]_0) - \bar{d}_w \cdot W_{ji}^1 \tag{8}$$

$$\frac{d}{dt} W_{ji}^2 = -\bar{\lambda}_2 \cdot C_j \cdot D_i^2 \cdot ([DA](t) - [DA]_0) - \bar{d}_w \cdot W_{ji}^2 \tag{9}$$

Considering that dopamine concentration ($[DA]$) excurses from the baseline ($[DA]_0$) during a short pause in TAN activity only, while the degradation process occurs continuously on a lot longer timescale, we can approximately rewrite these equations in a difference form by integrating over the pause duration:

$$\Delta W_{ji}^{1} = \bar{\lambda}_1 \cdot C_j \cdot D_i^1 \cdot \int ([DA](t) - [DA]_0)\, dt - d_w \cdot W_{ji}^1 \tag{10}$$

$$\Delta W_{ji}^{2} = -\bar{\lambda}_2 \cdot C_j \cdot D_i^2 \cdot \int ([DA](t) - [DA]_0)\, dt - d_w \cdot W_{ji}^2 \tag{11}$$

Where $\lambda_{1,2} = \lambda_{1,2} \cdot 0.00125$ if $[DA] \geq [DA]_0$ or $\lambda_{1,2} = \lambda_{1,2} \cdot 0.0025$ if $[DA] < [DA]_o$ All other parameters of BG model remain unchanged and can be found in Kim et al. (2017).

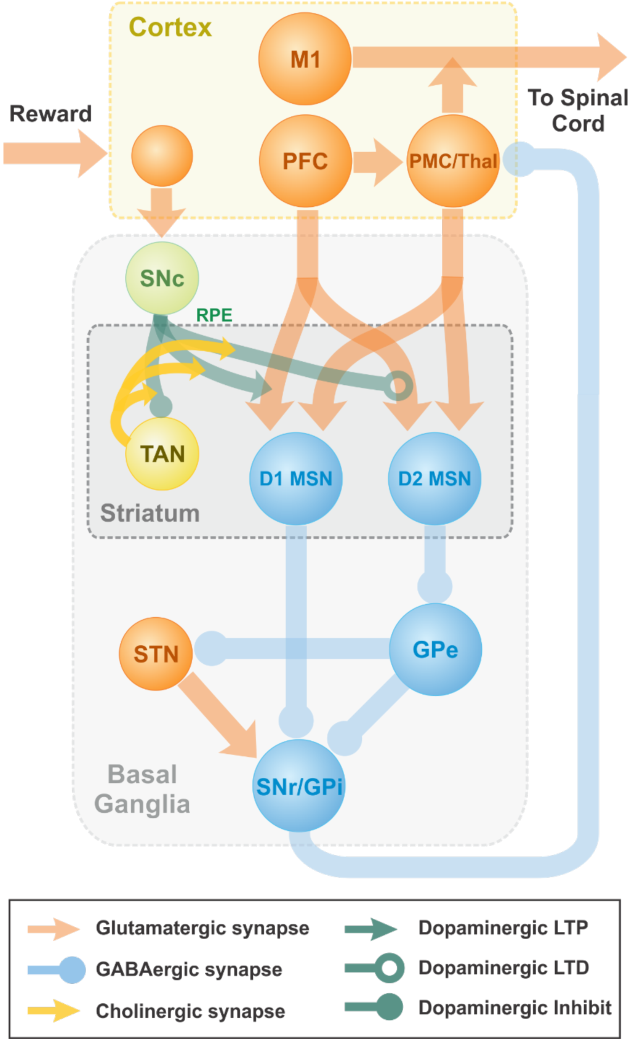

Figure 1 Schematic diagram of two-pathway of basal ganglia integrated with TAN model. Dopaminergic Substantia Nigra pars compacta signal represents the reward prediction error (reward prediction error). PFC (PreFrontal Cortex); M1 (Primary Motor Cortex); PMC (PreMotor Cortex); MSN (Medium Spiny Neuron); SNr (Substantia Nigra pars Reticulata); GPi (Globus Pallidus internal); GPe (Globus Pallidus external); Substantia Nigra pars compacta (Substantia Nigra pars Compacta); STN (SubThalamic Nucleus).

4.2.3 Dopamine deficiency simulation

Striatal dopamine deficiency is caused by degeneration of dopamine producing neurons as observed in Parkinson's Disease patients. Parkinson's Disease is a long-term neurodegenerative disorder of the central nervous system that mainly affects the motor system. Shaking, rigidity, slowness of movements and difficulty with walking are the most obvious Parkinson's Disease symptoms so called parkinsonism or parkinsonian syndrome [Kalia and Lang, 2015]. Motor learning is also impaired [Gutierrez-Garralda et al., 2013]. Aging is also often accompanied by death of midbrain Substantia Nigra pars compacta neurons which causes parkinsonism-like motor disorders [Kalia and Lang, 2015]. Based on the above, we assume that dopamine deficiency results from a reduced number of dopamine neurons which produce proportionally smaller amount of dopamine. To simulate this condition, we multiply the right-hand side of the equation describing dopamine concentration dynamics

$$\begin{aligned}\tau_{DA}\frac{d[DA](t)}{dt} + [DA](t) = \alpha\Bigg(RPE \cdot \left(1 - \frac{V_{TAN}(t)}{\theta_{DA}}\right) \\ \cdot\Theta\left(\theta_{DA} - V_{TAN}(t)\right) + [DA]_0\Bigg)\end{aligned} \tag{12}$$

by a coefficient α between 0 and 1 with $\alpha = 1$ corresponding to 0% dopamine deficiency and $\alpha = 0$ meaning 100% dopamine deficiency, i.e., no dopamine is produced at all. Fifty percent dopamine deficiency used in our simulations assumes that the coefficient used is $\alpha = 0.5, 30\%$ deficiency corresponds to $\alpha = 0.7$, etc.

4.2.4 Levodopa medication simulation

Levodopa is an amino acid made by biosynthesis from the amino acid L-tyrosine [Knowles, 1986]. Levodopa can cross the blood brain barrier whereas dopamine itself cannot and so it is naturally transferred into the brain via blood circulation [Wade and Katzman, 1975]. Then levodopa as a precursor to dopamine is converted to dopamine by the enzyme called DOPA decarboxylase (aromatic L-amino acid decarboxylase) in the central nervous system [Hyland and Clayton, 1992]. Thus, levodopa application increases overall dopamine concentrations in the brain. Levodopa medication is a clinical treatment for Parkinson's Disease patients as dopamine replacement to compensate for the dopamine deficiency. It is unclear whether levodopa improves the function of remaining dopamine neurons or affects baseline levels of dopamine in the brain only.

Our objective was to investigate if increasing the baseline dopamine concentration by levodopa without affecting the phasic dopamine release can improve learning performance in simulated Parkinson's Disease conditions. Thus we mathematically describe the effect of levodopa medication by adding a constant term to the right-hand side of the equation for dopamine concentration

$$\begin{aligned}\tau_{DA}\frac{d[DA](t)}{dt} + [DA]_0 = \alpha \Bigg(RPE \cdot \left(1 - \frac{V_{TAN}(t)}{\theta_{DA}}\right) \\ \cdot \Theta\left(\theta_{DA} - V_{TAN}(t)\right) + [DA]_0 \Bigg) + LDOPA\end{aligned} \tag{13}$$

where LDOPA is an increase in the baseline dopamine concentration due to levodopa administration. Correspondingly, to calculate the phasic component of dopamine dynamics in conditions of dopamine deficiency and/or levodopa medication for the baseline dopamine concentration, we use $\alpha[DA]_0 + LDOPA$ instead of $[DA]_0$.

4.2.5 Simulation environment

Our basic TAN activity-DA release interaction model was developed and simulated in Matlab. Then the model was implemented in C++ to integrate it into our previous model of reward-based motor adaptation described in detail in [Kim et al., 2017b]. All simulations for behavioral experiments were performed using custom software in C++. The simulated data were processed in Matlab to produce figures. For behavioral experiments, we performed 75 simulations (25 before perturbation, 25 with perturbation, 25 after perturbation) per session and results of 8 sessions were averaged (see [Kim et al., 2017b] for more details).

4.3 Experimental Results

4.3.1 Model of the TAN-dopamine relationship

Here we provide a short conceptual description of the model, sufficient for the qualitative understanding of the system dynamics. For equations and details please see Methods.

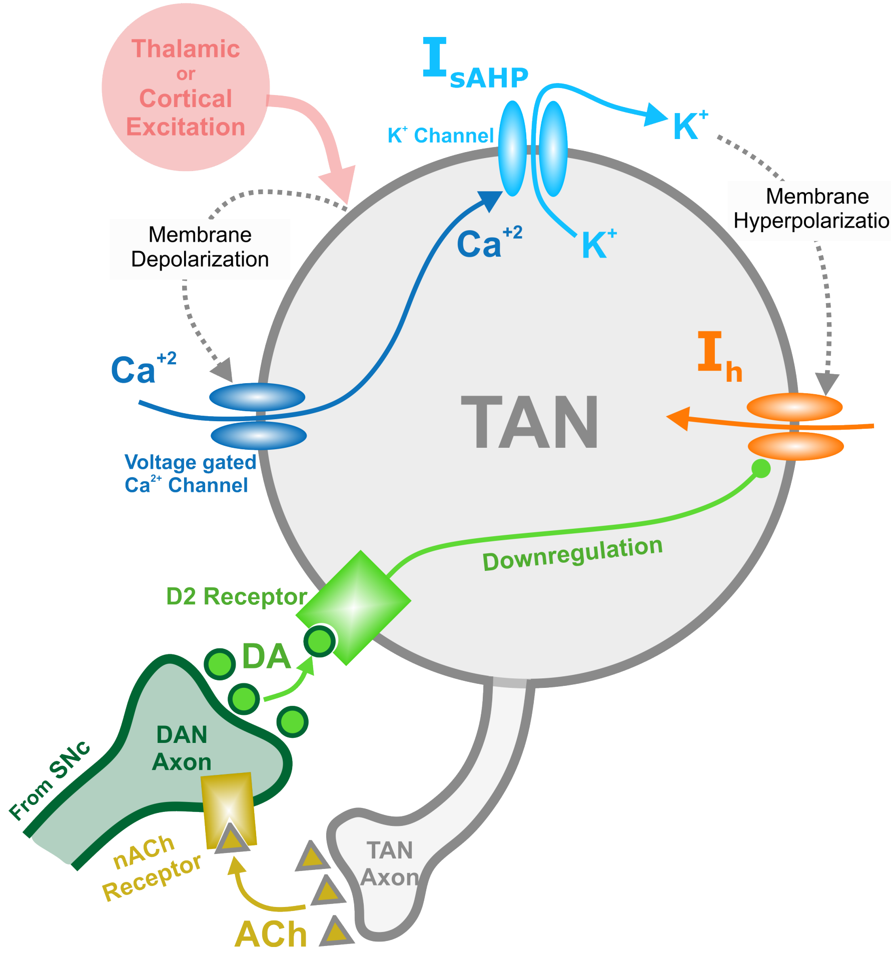

Figure 2 Diagram of the mechanisms involved in interactions between TANs and dopamine-release. Thalamic or cortical excitation leads to membrane depolarization in TANs. In response to depolarization, calcium ions enter through voltage dependent calcium channels, and the slow afterhyperpolarization current (IsAHP) is activated via the efflux of potassium ions through calcium dependent potassium channels. Once the cortical/thalamic excitatory input ends, the efflux of potassium ions causes the membrane to hyperpolarize, which in turn activates the inward dopamine-dependent h-current (Ih) that increases the membrane potential. Furthermore, dopamine (DA) from dopaminergic neurons (DANs) in substantia nigra pars compacta (SNc) binds to D2 receptors on TANs, downregulating the h-current. In concert, TANs produce acetylcholine (ACh), which binds to nicotinic acetylcholine (nACh) receptors on DAN axonal terminals. This cholinergic pathway enables TANs to modulate the release of dopamine into the synaptic cleft. Importantly— since the h-current is downregulated via activation of dopamine D2 receptors— the DA concentration affects the refractory period of TANs.

Rate-based TAN population In the model, we assume that TANs comprise a homogeneous neuronal population, whose activity is described by a single variable representing the normalized firing rate of the population. We also assume that ACh release and the activation of all cholinergic receptors in the model are proportional to TAN activity.

TANs receive excitatory inputs from the cortex and thalamus [Ding et al., 2010, Kosillo et al., 2016, Yager et al., 2015]. These inputs are implemented in the model as a binary input that—when activated—initiates a burst, followed by a pause in TAN activity.

TAN activity is attenuated by the slow afterhyperpolarization (sAHP) current. The sAHP current is activated by TAN depolarization—represented in the model as TAN activity in excess of a specified threshold. The kinetics of this current are defined on a timescale of hundreds of milliseconds. This mechanism—intrinsic to the TAN population—is responsible for generating the pause in TAN activity, following a stimulus from the cortex/thalamus.

TAN activity is also affected by a depolarizing hyperpolarization-activated h-current. This inward current activates when TANs are hyperpolarized, and the timescale of its kinetics is similar to the sAHP current. The h-current thus contributes to the recovery of TANs from the pause in activity. In the model, the h-current deactivates in response to an increase in the concentration of dopamine—an implementation of D2-receptor agonism, which serves as a dopamine-based modulation of TAN activity [Deng et al., 2007]. This mechanism provides the basis for a positive correlation between TAN pause duration and dopamine concentration.

Dynamics of striatal dopamine concentration In the model, the release of dopamine in striatum depends on the firing rate of SNc dopaminergic neurons, which receive cholinergic inputs through TAN-released acetylcholine. In the absence of acetylcholine—which occurs during a TAN pause—dopamine release is proportional to the firing rate of dopaminergic neurons. In contrast—during TAN tonic firing regimes—the release of dopamine is constant and corresponds to the baseline extracellular concentration of striatal dopamine. With increasing values of the cholinergic input to dopaminergic neurons, dopamine release becomes less dependent on the firing rate of dopaminergic neurons, and increasingly dependent on the magnitude of the TAN-provided cholinergic modulation (see Methods for mathematical description).

We also assume that the deviation of the firing rate of dopaminergic neurons from its baseline encodes the difference between the expected and received reward—the reward prediction error [Morris et al., 2004]. Positive reward prediction errors correspond to increases in the firing rate of dopaminergic neurons, and negative reward prediction errors correspond to decreases in the firing rate of the dopaminergic neuron population. To constrain the model, we require that the baseline dopamine concentration is the same, whether it is defined by the baseline firing of the SNc neurons in absence of cholinergic inputs during the pause in TAN activity, or when controlled by those inputs during tonic TAN firing. We refer to deviations from the baseline dopamine concentration as “phasic dopamine release”.

As follows from the above, for striatal dopamine dynamics to encode the reward prediction error—i.e. for reward information to be processed in the striatum [Zhou et al., 2002, Calabresi et al., 2000, Centonze et al., 2003, Pisani, 2003, Cragg, 2006, Joshua et al., 2008]—a pause in TAN activity must occur. In the model (see Fig. 2), a thalamic stimulus produces an initial increase in the TAN firing rate. When the stimulus ends, due to activation of the sAHP current the TAN pause begins. During the pause, TANs stop releasing acetylcholine, resulting in a phasic dopamine release—proportional to the firing rate of dopaminergic neurons. While TAN activity is paused, the sAHP current slowly deactivates, and eventually TAN activity returns to baseline [Cragg, 2006, Aosaki et al., 2010].

Figure 2 depicts the dynamics of TAN activity and dopamine concentration in cases of positive, zero and negative reward prediction error, as generated by the model. If the reward prediction error is positive, the dopamine concentration increases above the baseline during the TAN pause (Figure 2A). Since the h-current in TANs is inactivated via D2 agonism, the increase in dopamine release during the TAN pause prolongs the pause by suppressing the h-current. If the reward prediction error is zero, the dopamine concentration does not change during the TAN pause (Figure 2B), which means the pause is shorter than in the case of a positive reward prediction error. Finally, when the reward prediction error is negative, the dopamine concentration falls below the baseline during the TAN pause (Figure 2C), which upregulates the h-current and thus results in an even shorter pause duration. In summary, the TAN pause duration positively correlates with the reward prediction error in the model.

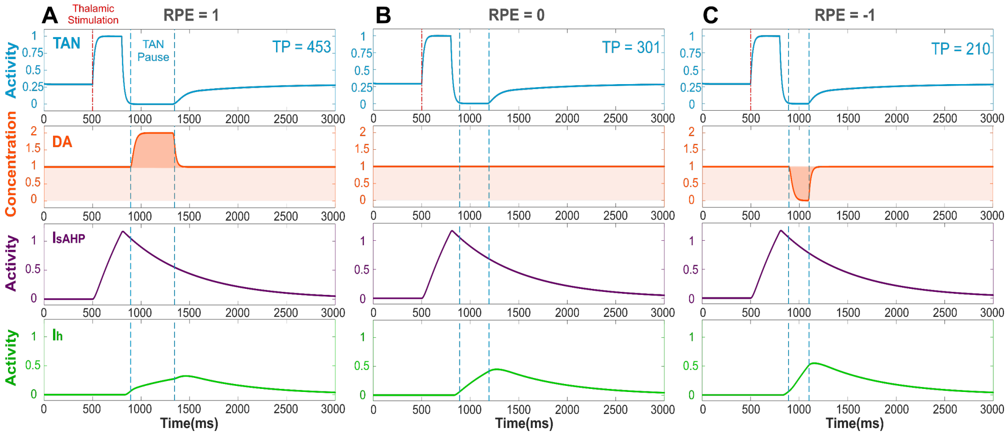

Figure 3 The TAN pause duration positively correlates with the reward prediction error (RPE). Thalamic stimulus induces an initial burst of TAN activity, followed by a TAN pause. The blue curve is TAN activity; the orange curve is dopamine (DA) concentration; the purple curve is the slow afterhyperpolarization current IsAHP and the green curve is the h-current Ih. (A) RPE=1, the dopamine concentration increases during the TAN pause as a result of the positive RPE, which slows down Ih activation and thus prolong the pause. (B) For RPE=0, the TAN pause is shorter, because there is no phasic change in dopamine release, so the concentration of dopamine remains at baseline during the TAN pause. (C) RPE=-1, the TAN pause is even shorter than for RPE=0 because there is a net decrease in dopamine concentration during the pause, which provides the fastest Ih activation and hence, the shortest pause in TAN activity. Thalamic stimulation duration was 300ms. TP stands for TAN pause duration in milliseconds.

4.3.2 Calibration of the model

To calibrate the model, we first simulated the condition without phasic dopamine release and compared the results to those obtained by Ding et al. [Ding et al., 2010]. They experimentally studied changes in TAN activity, which were modulated pharmacologically with drugs affecting dopamine release, reuptake, and binding (Figure 3). We varied the model parameters to reproduce the experimental time course of TAN activity in control conditions as well as after application of sulpiride and cocaine (blue traces in Figure 3). Sulpiride is a selective D2 receptor antagonist; thus, in the model administration of sulpiride corresponds to maximal activation of h-current in TANs (see Methods), which in turn shortens the pause duration. Then—because cocaine is a dopamine transporter antagonist, which results in an increase in extracellular dopamine—we simulated the cocaine condition by increasing the tonic dopamine concentration in the model until the TAN pause duration matched the experimental results.

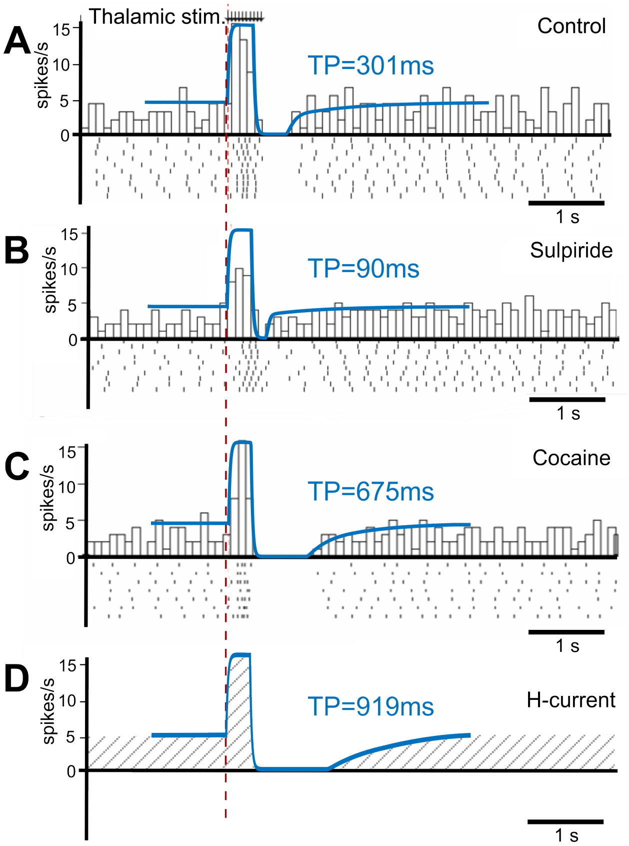

Figure 4 TAN activity as simulated by the model against experimental data. (A-C) Peristimulus time histogram (PSTH) and raster plot from striatal cholinergic interneurons in response to a train (50Hz, ten pulses) of thalamic stimulation. The background figures were reproduced from Ding et al. (2010) with permission. For easier comparison, all simulation results (blue lines) were rescaled down at the same ratio and overlaid on the figures of experiment results. (A) Simulation (blue) and data (gray bars) for control condition. (B) Simulation and data for sulpiride (D2 receptor blockade) condition. (C) Simulation and data for cocaine (dopamine reuptake blockade) condition. (D) Simulation of the hypothetical blockade of h-current. TP stands for TAN pause duration.

Additionally, we performed simulations of complete suppression of h-current (see Figure 3D) by setting the conductance of h-current to zero. This simulation qualitatively corresponds to the experimental results concerned with h-current blockade as described by Deng et al. [Deng et al., 2007].

4.3.3 Striatal dopamine deficiency

Having calibrated the model, we further investigated the implications of the proposed TAN-dopamine interactions. We first simulated the condition of striatal dopamine deficiency, which may be caused, for example, by the degeneration of dopaminergic neurons in the Substantia Nigra pars compacta that occurs in Parkinson's Disease. Because dopaminergic signaling is critical for action selection and learning in the basal ganglia, dopamine deficiency adversely affects those functions. We assumed that the degenerated Substantia Nigra pars compacta neuronal population releases less dopamine during both tonic and phasic modes. Accordingly, dopamine deficiency conditions were simulated by reducing the tonic dopamine concentration by a factor less than 1 and reducing the reward prediction error by the same factor (see Materials and Methods). Thus, both tonic (baseline) and phasic dopamine levels are decreased by the same factor; Figure 4A and 4B show changes in TAN pause and dopamine dynamics in dopamine deficiency conditions. Noteworthy, in the dopamine deficiency conditions, the duration of the TAN pause decreases in response to the reduction in dopamine concentration (Figure 4).

4.3.4 Effects of levodopa medication

Using the model, we investigated the mechanisms of levodopa-based treatments for dopamine deficiency. Levodopa (L-DOPA) is a common medication for Parkinson's Disease patients to increase overall dopamine concentration in the brain [Brooks, 2008, Kalia and Lang, 2015]. Levodopa readily passes across the blood brain barrier and converted to dopamine [Wade and Katzman, 1975, Hyland and Clayton, 1992]. This additional extracellular dopamine propagates nonspecifically throughout the brain. When simulating levodopa treatment conditions, we assume that levodopa administration increases the tonic (baseline) dopamine concentration but does not affect the phasic dopamine release. In the model, the concentration of levodopa is represented as a constant added to the baseline dopamine concentration. Figure 4C shows the corresponding simulation results. Importantly, although phasic dopamine release is unaffected by levodopa, the increase in tonic dopamine prolongs the TAN pause duration.

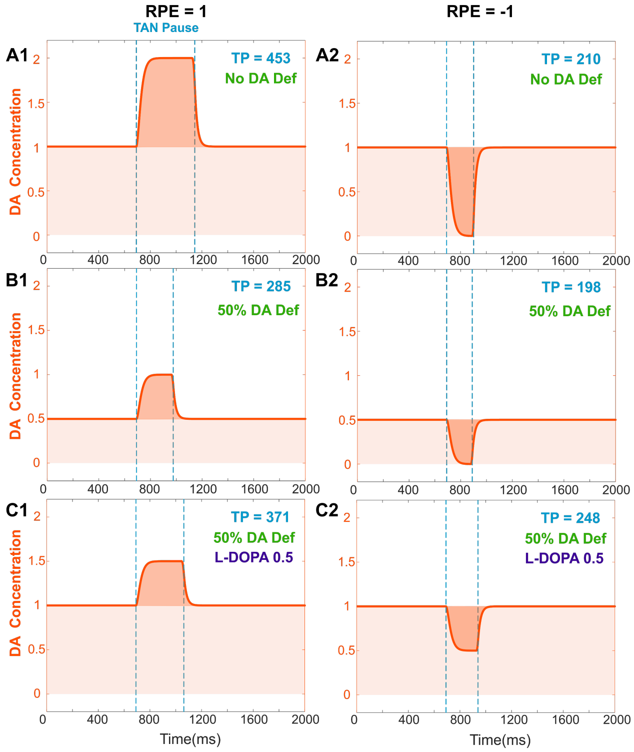

Figure 5 Effects of dopamine deficiency on TAN pause duration (TP, area between two dotted blue lines) and changes in dopamine concentration (orange) with/without levodopa (L-DOPA). In these simulations, a 50% dopamine deficiency (DA Def) causes both the baseline dopamine concentration and the phasic dopamine release to decrease. (A1-2) RPE=1 and -1, no dopamine deficiency for reference. (B1) RPE=1, 50% dopamine deficiency. Normally, the baseline concentration of dopamine would be 1.0. With a deficiency of 50% of dopaminergic inputs, the baseline dopamine concentration is exactly halved; additionally, the phasic release of dopamine decreases in magnitude by 50%, and therefore the duration of the TAN pause also decreases. (B2) RPE=-1. The tonic and phasic release of dopamine are both reduced by the 50% due to dopamine deficiency. During the pause, dopamine concentration converges to zero, so the pause is similar (slightly shorter) to A2. (C1) RPE=1. When levodopa (0.5) is applied, the baseline concentration of dopamine returns to normal (1.0) and the duration of the TAN pause duration increases, but it remains smaller than the one with no DA deficiency (A1). This is because the magnitude of phasic dopamine release is unaffected by levodopa. (C2) RPE=-1. When levodopa (0.5) is applied, the baseline concentration of dopamine returns to normal (1.0) as for RPE=1, but the duration of the TAN pause exceeds the one with no DA deficiency (A2). This is due to the increased (non zero) dopamine concentration during the pause.

4.3.5 Non-error-based motor adaptation during dopamine deficiency

In addition to our analysis of the local effects of dopamine deficiency on the striatal dopamine concentration, we also simulated the effects of dopamine deficiency on motor adaptation by incorporating the current model of TAN-dopamine interactions into our previously published model of reward-based motor adaptation [Kim et al., 2017b] (see Materials and Methods for details). Using this integrated BG model—including the TAN-dopamine interactions—we reproduced the non-error based motor adaptation experiments of Gutierrez-Garralda et al. (2013). In these experiments, healthy subjects, Parkinson's Disease patients, and Huntington's Disease patients threw a ball at a target under dffierent visual perturbation scenarios. In one scenario, each subject's vision was horizontally reversed using a Dove prism so that missing the target to the right was percived as missing to the left, and vice versa—corresponding to a sign change in the percieved error vs the actual error. This perturbation rendered error-based motor adaptation useless. In these experiments, each session was comprized of 75 trials (25 trials before the perturbation, 25 trials with the pertubation, and 25 trials after the perturbation). 8 sessions per subject were performed and averaged. Subjects in the control group gradually overcame the visual perturbation and reduced the distance error, but Parkinson's Disease subjects showed poor learning performance (distance errors fluctuated without any sign of adaptation in 25 trials, Figure 5A). In our simulations, we assumed that dopamine deficiency was the cause of Parkinson's Disease symptoms [Kalia and Lang, 2015]. To see how much dopamine deficiency affects learning performance in the model, we performed multiple simulations with changing dopamine deficiency conditions from 0 to 90% (see Methods for details). The simulation of 0% dopamine deficiency (Figure 5B, control) shows a trend of decreasing errors, which accurately reproduces the experimental results of control subjects in [Gutierrez-Garralda et al., 2013] (Figure 5A, control). As we can see in Figure 5B (Dopamine Deficiency), at 50% dopamine deficiency, learning performance is poor and is similar to the experimental results in Parkinson's Disease patients (Figure 5A, PD). For over 50% dopamine deficiency, average distance error remains at the initial level for all 25 trials, while error fluctuation and standard distance error decrease (result not shown). In summary, almost no learning occurs in the model when dopamine deficiency exceeds 50%.

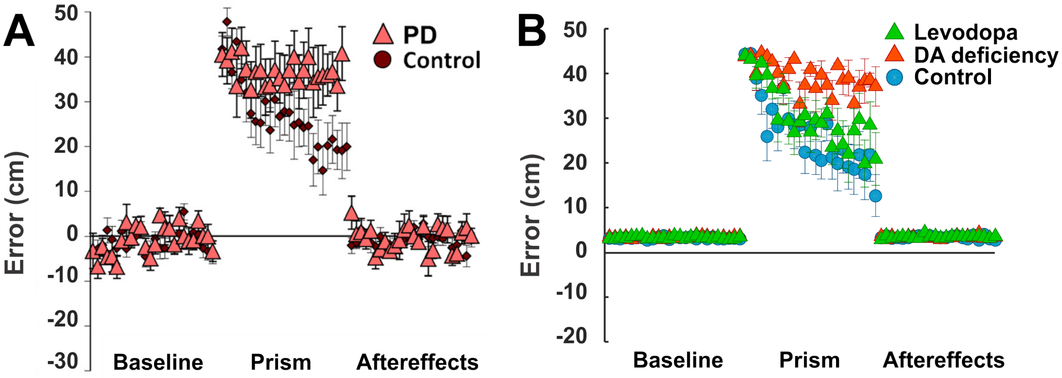

Figure 6 Non-Error based motor adaptation in 50% of dopamine (DA) deficiency condition with/without levodopa medication. (A) Results of ball throwing tasks performed by healthy people and Parkinson's Disease (PD) patients. During experiment, a dove prism was used to horizontally flip subjects' vision as perturbation. This figure was adapted from Gutierrez-Garralda et al. (2013) with permission. (B) Simulation results with levodopa medication. Levodopa means the condition of 50% dopamine deficiency with levodopa medication ([LDOPA]=1.0). Colored center markers (triangle or circle) are average error values of 8 sessions and error bars represent standard errors. 1 session = 75 trials (Baseline = 25 trials, Prism (visual perturbation) = 25 trials and Aftereffects = 25 trials).

4.3.6 Recovery of non-error-based motor adaptation with levodopa

To investigate the effects of levodopa medication on reinforcement learning in the striatum, again we simulated the same experimental settings. In the model, dopamine deficiency was set at 50% to simulate Parkinson's Disease conditions and simulations were performed with varying levodopa values representing additional striatal dopamine converted from levodopa medication. Figure 5B (Levodopa) shows the simulation results.

At levodopa values corresponding to 100 recovery of the baseline dopamine concentration, the average error decreases siginificantly at the end of the perturbation trials (Figure 5B, Levodopa). Thus, the overall learning perfomance of the model significantly improves as a result of levodopa administration.

However—although the learning performance improves—the performance of levodopa-medicated patients is still noticably worse than in control subject simulations. This performance difference can be easily understood in the context of our model of TAN-dopamine interactions. In the model, when levodopa is introduced, the tonic concentration of dopamine returns to healthy baseline levels, but the amplitude of phasic dopamine release is not recovered (compare Figure 4A1 with 4C1). Therefore, our integrated model simulations suggest that Parkinson's patients can partially regain learning performance following levodopa administration—due to the increase in tonic dopamine concentration—but a full recovery is impossible without a corresponding increase in phasic dopamine release.

4.4 Observations & Conclusions

In this study we investigated the relationship between striatal dopamine and TAN activity; specifically, we elucidated the mechanism by which this interaction affects reinforcement learning in the striatum. Striatal TANs temporarily pause their tonic firing activity during sensory or reward events. During tonic firing regimes, TAN activity defines the baseline striatal dopamine concentration via nicotinic ACh receptors (nAChR) activation on dopaminergic axon terminals [Rice and Cragg, 2004]; thus, the TAN pause enables a temporary variation of dopamine release. The duration of the TAN pause is important as it creates a window of opportunity for the dopaminergic neurons to transmit information about the reward prediction error by phasically modulating the dopamine concentration in the striatum. In turn, the concentration of dopamine determines the duration of the TAN pause by modulating the h-current via D2 receptors in TANs [Deng et al., 2007]. Accordingly, in our model, the TAN pause enables the phasic release of dopamine, and the duration of the TAN pause varies with dopamine concentration.

One of the objectives of this study was to extend our previous model by adding details of the striatal circuit concerned with cholinergic modulation of dopamine release. By doing so, we were able to investigate how TAN activity contributes to reinforcement learning mechanisms in simulated behavioral experiments.

In the model, phasic dopamine levels are defined by the activity of dopaminergic neurons, which codes the reward prediction error. Deviations of striatal dopamine concentration from its baseline underlie the plasticity of cortico-striatal projections to medium spiny neurons, representing a basis for reinforcement learning in the striatum. These deviations last for the duration of the pause in TAN activity. Therefore, the magnitude of long-term potentiation or depression of cortico-striatal projections depends on the pause duration, which may affect learning performance.

TANs express D2 dopamine receptors, which are inhibitory. Through this mechanism, the duration of the pause in TAN activity positively correlates with striatal dopamine concentration. In conditions of dopamine deficiency, the baseline dopamine concentration is reduced, which also shortens the duration of the TAN pause.

Based on our model predictions, we speculate that levodopa medication improves learning performance in Parkinson's patients by increasing the baseline dopamine concentration and thus prolonging the pause in TAN activity—even though the magnitude of phasic dopamine excursions may be not affected by this medication.

4.4.1 Dopamine release and cholinergic regulation

Within the Substantia Nigra pars compacta—a structure in the midbrain—are dopaminergic neurons that project to the striatum. These dopaminergic neurons are known to encode reward-related information by deviating from tonic baseline activity [Schultz, 1986, Hyland et al., 2002]. Striatal dopamine release occurs via vesicles at local dopaminergic axon terminals [Sulzer et al., 2016].

However, the amount of dopamine released is likely to be not always defined by the firing rate of the presynaptic neuron. Cholinergic activity plays a major role in modulation of dopamine release in the striatum. For example, synchronized activity of striatal TANs directly evokes dopamine release at the terminals—regardless of the activity of dopaminergic neurons [Cachope et al., 2012, Threlfell et al., 2012]. TANs release acetylcholine (ACh), which binds to nicotinic receptors on the axons of dopaminergic neurons—and when these cholinergic inputs are activated, dopamine release is independent of electrical stimulation frequency [Rice and Cragg, 2004]. However, when these nicotinic receptors (nAChRs) are blocked, the magnitude of dopamine release becomes proportional to the stimulation frequency [Rice and Cragg, 2004]. Therefore, it is necessary for the cholinergic inputs to dopaminergic neurons to cease so that dopamine release reflects the firing activity of the presynaptic neurons.

Our model assimilates the above observations via the following assumptions. Baseline striatal dopamine concentration is determined by the presynaptic action of ACh on dopaminergic terminals [Threlfell et al., 2012] through nAChR desensitization. With no cholinergic inputs, e.g. when TAN activity ceases or nAChRs are blocked, the firing rates of dopaminergic neurons define the dopamine release. In other words, the phasic component of dopamine release is determined by Substantia Nigra pars compacta activity, which codes the reward prediction error. Therefore, the functional role of the pause in TAN activity is to allow the striatal dopamine concentration to vary, thus creating a window of opportunity for dopaminergic neurons to deliver the reward information to and enable reinforcement learning in the striatum.

Variations in the phasic release of dopamine reflect the reward prediction error [Hollerman and Schultz, 1998, Schultz, 1999]; thus, in the case that the reward received is exactly the same as the expected reward—reward prediction error is zero—the dopamine concentration should not change during the TAN pause. In the model, as explained above, the baseline dopamine concentration is constrained by cholinergic inputs from TANs, and during the pause, dopamine release is controlled by the firing rate of dopaminergic neurons in the Substantia Nigra pars compacta. Therefore, we constrained the model by requiring that Substantia Nigra pars compacta firing corresponding to a reward prediction error value of zero (RPE=0)—in absence of cholinergic input during the pause—leads to exactly the same dopamine release as during normal TAN activity. The exact homeostatic mechanisms responsible for such tuning remain open for speculation.

In our model, we did not differentiate between different parts of striatum in terms of cholinergic regulation of dopamine release. However, it was reported that the nucleus accumbens shell, the most ventral part of striatum, has a distinctive modulation mechanism of dopamine release with much higher activity of acetylcholinesterase minimizing nAChR desensitization, which is different from nucleus accumbens core and dorsal striatum [Shin et al., 2017]. There is also evidence that DA release in nucleus accumbens is modulated by ACh not only through nicotinic but also via muscarinic receptors of several types activation of which has different effects on DA concentration [Shin et al., 2015]. Our model does not account for this.

In our model, we focused on the functional role of TAN activity-dopamine interactions in reinforcement learning. Thus, we did not consider the effect of TANs on other striatal neuron types. For example, MSNs are known to receive cholinergic inputs via muscarinic M1 and M2 receptors. Functional role of these projections was discussed elsewhere. In particular, other computational models proposed that TANs might have a timing control function to hold and release MSNs [Ashby and Crossley, 2011, Franklin and Frank, 2015b]. Besides TANs and MSNs, many other types of interneurons have been identified in striatum, such as parvalbumin fast spiking interneurons, neuropeptide Y interneuron, calretinin interneurons, Tyrosine Hydroxylase interneurons [Tepper et al., 2010, Xenias et al., 2015, Tepper et al., 2018]. Functional roles of these interneurons and their relationships with cholinergic interneurons are not clearly understood. However, this does not rule out the possibility, that some of these neuron types interact with TANs and thus may play a role in TAN activity regulation.

4.4.2 TAN pause duration

In the model, the pause in TAN activity is initiated by transient excitatory corticothalamic inputs. Furthermore, the duration of the pause is dependent on the extracellular dopamine concentration [Deng et al., 2007, Ding et al., 2010, Oswald et al., 2009]. To replicate this dependence, we calibrated the duration of TAN pause in the model to in vitro experimental data from Ding et al.

It is important to note that longer thalamic stimulation means stronger activation of the slow after-hyperpolarization (sAHP) current, and hence more time is required for its subsequent deactivation. This prediction is consistent with the in vitro studies by Oswald et al. In their experiments, a higher number of stimulation pulses did generate stronger afterhyperpolarization in TANs below their resting potential—and accordingly evoked a longer pause in TAN activity. In addition, several in vitro and in vivo experiments agree that the magnitude of thalamic input positively correlates with the TAN pause duration [Oswald et al., 2009, Doig et al., 2014, Schulz et al., 2011]. Although we cannot directly compare our simulation results with their data, our TAN model exhibits a qualitatively similar relationship between input duration and pause duration.

To illustrate this relationship, we performed simulations, varying the duration of thalamic stimulation (from 100 to 400ms) as shown in Figure 6A. The duration of the TAN pause increases non-linearly in response to increasing thalamic stimulation duration. Interestingly, this increase in the pause duration is stronger for higher reward prediction error values, which is because of the larger phasic dopamine concentration when the reward prediction error increases. The reward prediction error is independent of the thalamic stimulus duration, and the pause duration is sensitive to both variables. Thus, we manipulated each variable independently to show the dependence of the pause duration on both.

Furthermore, the TAN pause duration is dependent on any change in the extracellular dopamine concentration—not just the RPE-determined phasic dopamine release. Therefore, we also produced simulations demonstrating the effects of dopamine deficiency as well as the effect of levodopa administration on the TAN pause duration. Importantly, dopamine deficiency has almost no effect on the TAN pause duration when the reward prediction error is at a minimum (see the orange line in Figure 6B). This model behavior follows from the observation that the reward prediction error correlates with the magnitude of phasic dopamine release. If the reward prediction error is at its minimum possible value (in our model, RPE=-1), then neither the amount of phasic dopamine nor the duration of the TAN pause can be decreased by dopamine deficiency conditions. In contrast, the administration of levodopa affects the TAN pause duration without any dependence on the reward prediction error. This follows from the fact that levodopa alters the baseline concentration of dopamine—not the phasic dopamine release—which is not dependent on the reward prediction error.

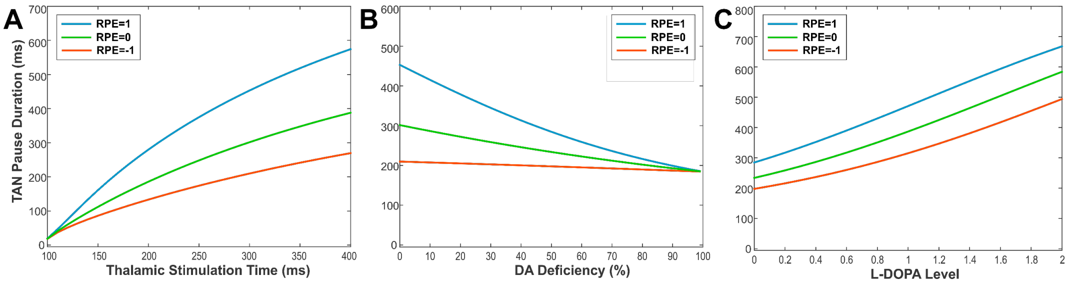

Figure 7 (A-C) The changes in TAN pause (TP) duration by three different factors: the duration of thalamic stimulation, the percentage of dopamine (DA) deficiency, the L-DOPA level in 50 DA deficiency condition when RPE (Reward Prediction Error) = 1 (phasic, reward), 0 (tonic baseline), and -1 (phasic, aversive) respectively. (A) The changes in TP duration by the duration times of thalamic stimulation. The increment of thalamic stimulation duration increases TP duration for all RPE values. The difference of TP duration between RPE=1 and RPE=-1 keeps increasing nonlinearly as increases in thalamic stimulation duration. (B) The changes in TP duration by the percentages of DA deficiency. The increased percentage of DA deficiency decreases TP duration when RPE=1 and 0. For RPE=-1, the TP duration is nearly independent of the amount of DA deficiency, which is the result of RPE=-1 corresponding to the minimum possible DA concentration during the TP. Therefore, the TP duration for RPE=-1 is unaffected by the degradation of dopaminergic inputs. The deviation difference of TP duration from RPE=0 between RPE=1 and RPE=-1 keeps decreasing nonlinearly as increases in percentage of DA deficiency, which means minimizing the time difference between reward and aversive conditions for reinforcement learning and in turn deteriorating the learning performance. (C) The changes in TP duration by the levels of L-DOPA in 50 DA deficiency condition. In response to the administration of L-DOPA, the TP duration increases similarly for all RPE values. This follows from the fact that L-DOPA alters the baseline concentration of dopamine, but does not affect the phasic dopamine release.

4.4.3 Comparisons with other models

The model presented here is not the first computational model of TAN activity. For example, Tan and Bullock previously developed a computational model incorporated h-current as an intrinsic property of TANs [Tan and Bullock, 2008]. Their model was also a non-spiking model that focused on the generation mechanism of TAN-specific activity patterns, which the authors attributed to intrinsic TAN properties. Even though their model accounted for modulation of TAN activity by dopamine level, it did not include a mechanism that affects the dopamine release, which our model did.

Ashby and Crossley also developed a BG model that included Hodgkin-Huxley style spiking TANs with h-current [Ashby and Crossley, 2011]. Their model emphasized the inhibitory effect of TAN activity on striatal medium spiny neurons (MSNs) through muscarinic receptors. They proposed that tonic TAN activity normally suppresses MSN firing, which is released during the TAN pause. Similar idea was exploited in the computational model of BG circuits by [Franklin and Frank, 2015b] who proposed that the pause in TAN activity is formed by local striatal inhibition to code the uncertainty and regulate learning rates through cholinergic projections to MSNs. The model we propose significantly differs from these two models with respect to the gating function of the pause in TAN activity. Our model focuses on cholinergic dopamine regulation and does not incorporate direct cholinergic projections to—or GABAergic projections from—MSNs.

To the best of our knowledge, the model proposed here is the first that incorporates bidirectional effects of cholinergic and dopaminergic signaling in the striatum and explores the implications of these interactions by simulating real and hypothetical behavioral experiments in realistic settings. This was made possible by embedding our implementation of TAN-dopamine interactions into the model of reward-based motor adaptation we previously published [Kim et al., 2017b].

4.4.4 Impaired learning in Parkinsonians and the effect of levodopa medication

Striatal dopamine deficiency in Parkinson's Disease is concerned with degeneration of dopaminergic neurons which results in smaller amounts of dopamine released. This affects both the baseline striatal dopamine concentration and phasic excursions of dopamine concentration that encode the reward prediction error. Our model predicts that lower dopamine concentration also leads to shortening of the pause in TAN activity, during which the phasic dopamine component drives reinforcement learning in the striatum. Using the model, we find that dopamine deficiency influences learning performance in the BG not only due to smaller magnitude of the learning signal, but also by decreasing the duration of the pause in TAN activity. From our simulation results, we found that 50% of dopamine deficiency in the model is sufficient to induce as poor learning performance as observed in Parkinsonians. This finding is consistent with the experimental data on striatal dopamine deficiency in Parkinson's Disease patients [Scherman et al., 1989] where it was reported that Parkinsonian symptoms appear when striatal dopamine deficiency exceeds 50%.

Levodopa is one of common treatments for early stage Parkinson's Disease patients [Brooks, 2008, Kalia and Lang, 2015]. Levodopa administration increases Parkinson's Disease patient's UPDRS (Unified Parkinson's Disease Rating Scale) score by two or three times [Brooks, 2008, Chen et al., 2016, Beigi et al., 2016]. In Gutierrez-Garralda et al.'s experiments [Gutierrez-Garralda et al., 2013], Parkinson's Disease patients were tested in the morning before taking their levodopa medicine to avoid levodopa effects on the results. According to a report, a standard dose of intravenous levodopa infusion increased the striatal dopamine level by 5-6 times [Zsigmond et al., 2014]. Due to the lack of data, it is hard to know by how much the oral intake of levodopa increases dopamine concentration in the striatum. However, from the conventional dosage for Parkinson's Disease patients [Brooks, 2008], we can infer that oral levodopa may take more time to increase striatal dopamine levels and have less efficacy on striatal dopamine levels than intravenous levodopa infusion. In our simulations, levodopa 1.0 (2 times higher than baseline dopamine in 50% dopamine deficiency) caused the learning performance to recover close to the control levels (see Figure 5B). This effect is solely provided by the prolonged pause in TAN activity due to the levodopa-induced increase in baseline dopamine concentration. Interestingly, the extended pause duration at levodopa 1.0 is close to the one in control (no dopamine deficiency) conditions (see Figure 6C). The required increase of the baseline dopamine concentration by levodopa administration and the one predicted by the model is within a ballpark range.

4.4.5 Alternative TAN pause mechanisms

In our model, the pause in TAN activity is induced by a cortico-thalamic excitatory input which causes after-hyperpolarization. However, other mechanisms for TAN pause generation have been proposed. For example, there exist inhibitory projections from GABAergic neurons in ventral tegmental area (VTA) to the cholinergic interneurons in nucleus accumbens [Brown et al., 2012]. Brown et al. (2012) were able to generate a pause of TANs in nucleus accumbens by optogenetically activating VTA GABAergic projection neurons and link this to potentiation of associative learning.

Interestingly, regardless of how the pause is generated, our model would exhibit the same qualitative features of interactions between TAN activity and DA release. Indeed, TAN recovery from the pause would still depend on activation of depolarizing h-current negatively modulated by DA through D2 receptors. Therefore, TAN pause duration would positively correlate with DA concentration thus providing the same basis for our conclusions.

On a side note, GABAergic inhibition of TANs has not been found in dorsal striatum [Zhang and Cragg, 2017], which means that external inhibition cannot represents the primary mechanism of the pause in dorsal striatal TAN activity. The same lab has recently provided further evidence that the pause in TAN activity is associated with intrinsic properties of striatal cholinergic interneurons, induced by an excitatory input, mediated by potassium currents, and modulated by dopamine [Zhang et al., 2018].

5

THE INTERPLAY BETWEEN CEREBELLUM AND BASAL GANGLIA IN MOTOR ADAPTATION: A MODELING STUDY

5.1 Background & Significance

5.1.1 Background

Motor adaptation to perturbations is provided by learning mechanisms operating in the cerebellum and basal ganglia. The cerebellum normally performs motor adaptation through supervised learning using information about movement error provided by visual feedback. However, if visual feedback is critically distorted, the system may disengage cerebellar error-based learning and switch to reinforcement learning mechanisms mediated by basal ganglia. Yet, the exact conditions and mechanisms of cerebellum and basal ganglia involvement in motor adaptation remain unknown. We use mathematical modeling to simulate control of planar reaching movements that relies on both error-based and non-error-based learning mechanisms. We show that for learning to be efficient only one of these mechanisms should be active at a time. We suggest that switching between the mechanisms is provided by a special circuit that effectively suppresses the learning process in one structure and enables it in the other. To do so, this circuit modulates learning rate in the cerebellum and dopamine release in basal ganglia depending on error-based learning efficiency. We use the model to explain and interpret experimental data on error- and non-error-based motor adaptation under different conditions.

5.1.2 Significance

Motor learning is a process of acquiring skills to perform an appropriate motor task in response to a sensory cue, e.g. precise reaching with a mouse pointer to a target spot shown on the screen. Motor adaptation is a form of motor learning to overcome movement perturbations caused by novel environment or altered sensory feedback. During motor adaptation, future movements are corrected using error information acquired on previous trials [Izawa et al., 2008]. Representation of the movement error depends on available sensory components. For example, during reaching movements under unexpected perturbation, visual feedback can provide a vector displacement of the movement endpoint relative to the target position. This vector error may be used by the central nervous system to adjust motoneuron activity and eliminate the effects of the perturbed environment. It has been suggested that this process involves the cerebellum that adjusts the internal model of the body based on information about movement error [Izawa et al., 2012]. This type of motor adaptation is often referred to as supervised or error-based learning [Doya, 2000b].

Error-based learning fully relies on availability and correct representation of the movement error. However, in certain conditions the necessary information about error is limited or unavailable. Several related experimental conditions were proposed. For example, in reaching experiments, the subject's arm can be covered with a non-transparent screen and visual feedback is transformed by rotating the image of the movement on a computer screen [Galea et al., 2010], inverting the image with a Dove prism [Gutierrez-Garralda et al., 2013], etc. When adequate visual feedback is not available, error-based learning becomes impossible. However, motor adaptation can still be observed if some other measure of the movement inaccuracy is provided. This may include a score representing the distance from the movement endpoint to the target, or a reward for a reach performed within a target spot [Izawa and Shadmehr, 2011, Shmuelof et al., 2012]. This type of motor adaptation is referred to as non-error-based learning and is mediated by the basal ganglia (BG) via reinforcement of more successful movement attempts [Izawa et al., 2012, Doya, 2000b].

The emergent view is that motor adaptation can be simultaneously provided by two distinct learning mechanisms: (1) an error-based learning mechanism performed by the cerebellum and (2) a non-error-based learning mechanism that operates in BG. Importantly, these two learning mechanisms differ in their use of sensory information (vision, proprioception, etc.). In the error-based (supervised) learning mechanism, sensory input is explicitly used to compute a vector displacement between the final position and a given target position, and the magnitude of this displacement is reduced with every trial. In contrast, the non-error-based (reinforcement) learning mechanism does not explicitly calculate the distance to a target and relies on whether the “reward” value for the performed movement is better than for the previous ones.

Various experimental setups were implemented to experimentally distinguish between error based and non-error-based learning. The most intriguing results were obtained in experiments where visual feedback was so distorted that efficient error-based learning was no longer possible. For example, Gutierrez-Garralda et al. [Gutierrez-Garralda et al., 2013] used a Dove prism as a visual perturbation while the participants attempted to throw a ball at a target, so that their perceived error was inverted. In these conditions, faulty cerebellar error-based learning would cause participants to perform progressively worse at the task. The control group of participants was still able to adapt to the perturbation, while the groups with impaired basal ganglia function failed to adapt. Interestingly, in experiments where instead of a Dove prism, a shifting wedge prism (that did not change perception of the displacement) was used, participants adapted to the perturbation regardless of the basal ganglia integrity. Because of this, Gutierrez-Garralda et al. [Gutierrez-Garralda et al., 2013] suggested that participants suppressed malfunctioning error-based learning mechanisms and engaged the BG-based reinforcement learning to successfully adapt to the visual distortion used. However, it remains unclear what triggered one of the two mechanisms to operate or disengage depending on conditions.

Previously, we suggested a mathematical model of reinforcement-based motor adaptation [Kim et al., 2017b] for planar reaching movements [Teka et al., 2017b]. The objective of the present study was to explain how supervised or reinforcement learning could be engaged or suppressed during motor adaptation depending on the perturbation. To do so we augmented the model by including a cerebellar compartment performing trial-to-trial movement correction based on supervised learning. Using the extended model, we suggest a simple intrinsic mechanism that may orchestrate the involvement of error- or non-error-based learning via regulating the cerebellar learning rate and dopaminergic signaling in basal ganglia.

5.2 Results

In this study, we modeled the neuromechanical control of reaching movements in humans and used this model to reproduce and explain the results of previous experimental studies demonstrating motor adaptation during reaching [Shmuelof et al., 2012, Schlerf et al., 2013]. In these experiments the participants did not have direct visual feedback from the arm. Instead, the arm endpoint was represented as a cursor on a display. During each experiment, different targets appeared on the screen and the task was to move the arm so that the cursor would reach the target. Such a setup allowed an experimenter to easily alter the visual feedback if needed. The applied perturbations included image rotations around the movement starting point and reflection across to the vertical axis. In presence of the perturbation, the participants had to learn to reach the target relying on the distorted visual feedback or other available information. To prevent the subjects from making correction during the movement, they were either forced to perform movements as quickly as possible or the cursor was only shown on the screen when the movement was complete.

Previously we developed a model for neural control of goal-directed reaching movements that simulated the entire pathway from the motor cortex through spinal cord circuits to the muscles controlling arm movements [Teka et al., 2017b]. In that model, the arm consisted of two joints (shoulder and elbow), whose movements were actuated by six muscles (4 single-joint and 2 two-joint flexors and extensors). Cortical inputs were calculated by a cortical "controller" based on (a) an internal model of the arm, (b) a proposed straight-line trajectory to a target position, and (c) a predefined bell-shaped velocity profile. The neural controller generated a motor program (six time-varying signals) producing a task-specific activation of low-level spinal circuits that in turn induced the muscle activation pattern resulted in the intended reaching movement. In our present study, the internal model of the body used to calculate the motor program, or the perception of the target position could be perturbed, resulting in an imprecise or completely faulty reaching movement. Therefore, we augment the model by sequentially adding structures responsible for different forms of motor learning allowing us to reproduce the experimental data on motor adaptation to the applied perturbations.

The overall organization of the Results section is as follows. First, we incorporate the error-based learning mechanism performed by the cerebellum, in the model. The mechanism modifies the motor program based on supervised learning. Then, we show that if the perception of movement error is strongly distorted, the error-based adaptation quickly worsens the performance. Experimentally, it was found that the subjects switch to a different non-error-based mechanism of motor adaptation mediated by BG in the latter situation [Gutierrez-Garralda et al., 2013]. Following our previous publication [Kim et al., 2017b], we add the BG network to the model which participates in forming the cortical motor program based on reinforcement of previously successful reaching attempts or on suppression of unsuccessful ones. Via simulations we show that if not orchestrated, the supervised and reinforcement learning processes do not get along. To avoid their interference, we incorporate a simple mechanism which regulates the learning rate in cerebellum and dopaminergic signaling in striatum and show its efficiency in endogenous switching between error-based and non-error-based learning mechanisms. Finally, we compare our model performance with existing literature data on the subject.

5.2.1 Supervised learning in cerebellum during reaching

In the present model we used the previous model [Teka et al., 2017b] as a basis, but added a model of cerebellum as an artificial neural network with an input from the cortical controller and the output to the spinal cord network (Fig 1), whose function was to alter the motor program in order to correct the movement. The synaptic weights of the cerebellar network were updated based on a classical error back propagation procedure [Dreyfus, 1990] using a squared distance between the movement endpoint and the target (the squared error) as a cost function (see Methods for details). Simply said, if the produced movement failed to precisely reach the target, the model adjusted the weights in the cerebellar network proportionally to the derivatives of the squared error with respect to the weights. The coefficient of proportionality is called the learning rate which defines how quickly the system adapts to the perturbation. The model also accounted for a trial-to-trial degradation of the cerebellar network weights which reflected "forgetting" process in the system and defined how quickly the accumulated correction would wash out if visual feedback was disrupted.

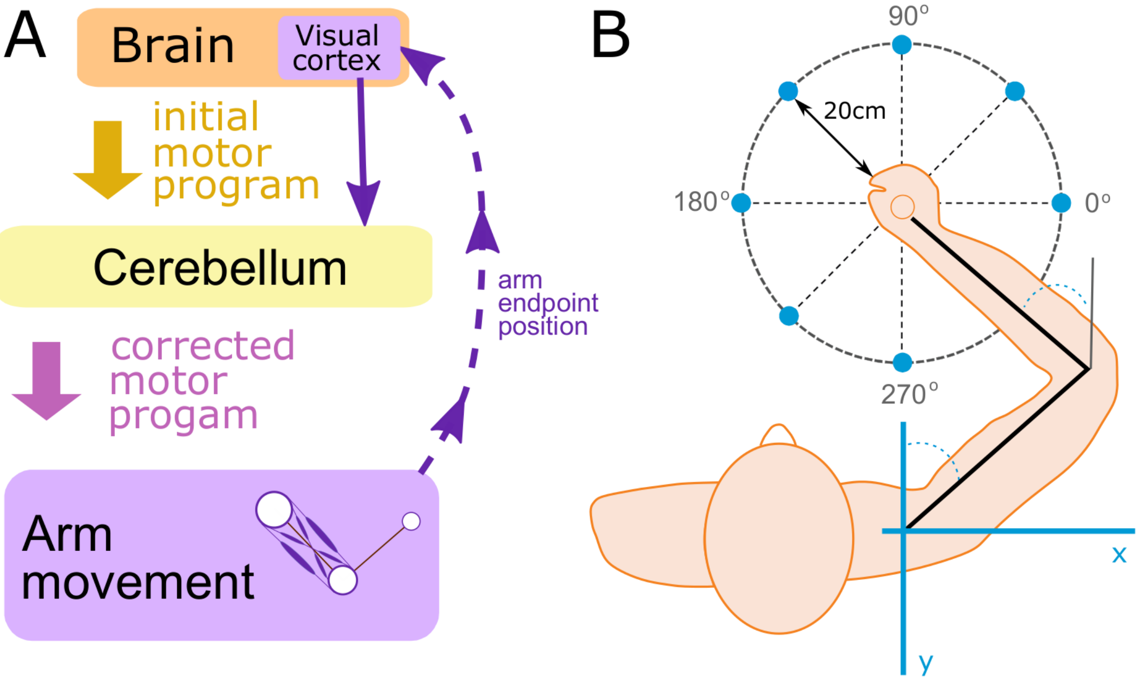

Figure 1 General model structure. (A) To perform a reaching movement the brain creates a motor program and sends it as an input to the cerebellum implemented as an artificial neural network. The cerebellum modifies the motor program and sends it as an input to the neuromechanical arm model. The perceived displacement between the movement endpoint and the target position (the vector error) is fed back to the cerebellum (visual feedback) and used to calculate the adjustment of weights in the cerebellar network using error back propagation. (B) We simulated reaching movements which started from a fixed initial position, aiming to reach a target located on a circle centered at the starting point with a radius of 20 cm. The direction of the movement relative to the body position was defined by the angle as shown.

To illustrate the process of error-based motor adaptation provided by the cerebellum, we simulated reaching movements in conditions of two different visual perturbations. Each simulation consisted of three consecutive phases. During the first phase, we simulated reaching to a target located in 90 degrees direction (see Fig 1B for definition of directions) with veridical feedback (i.e. the perceived endpoint position and expectation of where it should appear as a result of the movement coincided). During the second phase, a visual perturbation was introduced. Here we used two different perturbations: a shift and a rotation. The rotation perturbation mimicked the rotation of the image around the initial movement point by 30 degrees counterclockwise so that the perceived target position was at 120 degrees (see Fig 1B). Note, that during such a perturbation the perceived vector error (the perceived displacement of the movement endpoint relative to the target) is rotated by the same angle relative to the actual vector error. The shift perturbation consisted in parallel translation of the image in such a way that the target was perceived at the same location as during the rotation perturbation. The difference was in perception of the vector error which was not perturbed during the shift perturbation. During the third phase, the veridical feedback was restored.

Fig 2 depicts motor adaptation results to these two perturbations as generated by the model. During the first “BASELINE” phase the model generated movements that reached closely to the target within natural movement variability (see Methods for details). After perturbation (see ADAPT in Fig 2), the model missed the target by the angle which initially coincided with the magnitude of the perturbation of 30 degrees, but then the angular error gradually reduced to below 10 degrees on subsequent trials as the cerebellum learned to correct the movement. When the perturbation was removed (see POST in Fig 2), the perceived and actual target positions coincided again. However, at the beginning of the POST phase, the cerebellum network was primed to counteract the previous perturbation, so it required certain number of trials for the system to come back to baseline. This phenomenon is often referred to as an aftereffect of a perturbation.

Incomplete elimination of errors, observed in the end of ADAPT epoch of Fig 2 is well-known effect [Vaswani et al., 2015], widely observed in experimental studies of motor learning. To our knowledge so far there is no theory for the neural origin of it. Phenomenologically it is usually interpreted as a result of the balance between learning and “forgetting”. In a linear Kalman-filter like state space model of the CB it is implemented by adding a damping term to the equation. Our model also has this damping term (which we call “degradation”) that leads the error to saturate at a non-zero level defined by the balance between learning and degradation processes. See Methods section for implementation details.

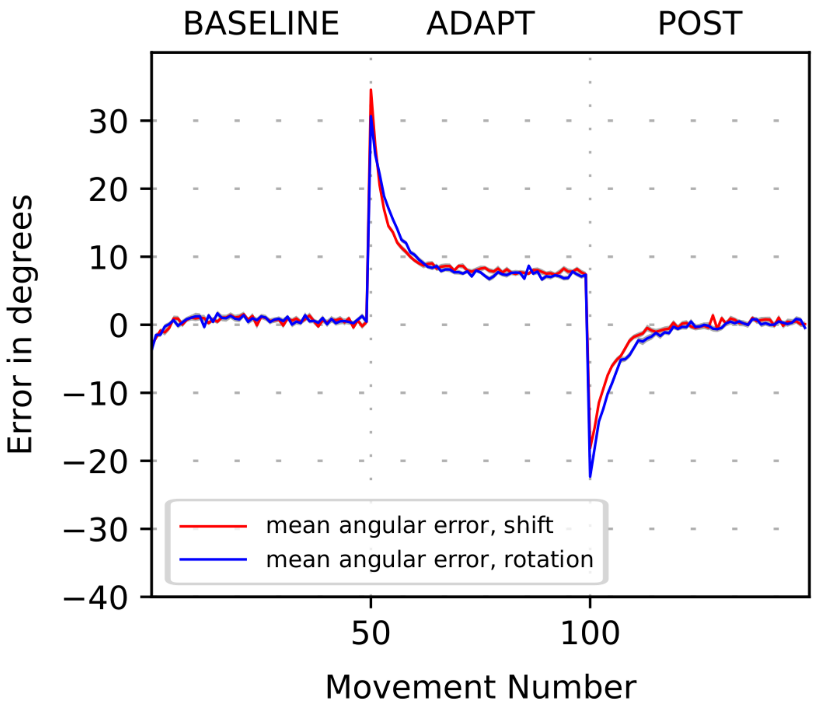

Figure 2 Simulation of cerebellar-based adaptation to mild perturbations. The plot shows the movement error in degrees versus trial number during simulation of the 3-phase experiment when a visuomotor perturbation is introduced in the beginning of ADAPT phase and removed in the beginning of POST phase. The perturbations are the shift (red line) and 30-deg visuomotor rotation (blue line) (see text for details). Solid lines show the 32-run per trial average of the angle difference between the perceived target position and the perceived hand/cursor position versus the trial number. After the perturbation is introduced, the initial error is approximately equal to the shift/rotation angle. Then it converges to the asymptote below 10 degrees. The asymptote is nonzero due to "forgetting" effect presented in error-based learning, corresponding to the damping term in the cerebellum model. After the perturbation is removed, the error changes its sign and abruptly increases in magnitude again (perturbation aftereffect). Then it converges back to zero.

As mentioned above, the rotation perturbs the perception of movement error and thus may adversely affect the supervised learning process relative to the shift. Interestingly, for a perturbation of the magnitude used, the courses of adaptation to the shift and the rotation were very similar. Thus, 30 degrees rotation did not distort the vector error to noticeably impair the error-based motor adaptation.

5.2.2 Model calibration

We used the experimental protocol and data from the study of Schlerf et al. [Schlerf et al., 2013] to calibrate the model parameters (see Schlerf et al. context in Methods), and reproduce some of their results. In short, in their experiments, the participants performed slicing movements to reach the targets appearing on the screen at a 10 cm distance from the starting location. The directions to the targets were uniformly randomly distributed within the sector from 75-15 to 75+15 degrees (see Fig 1B for directions). The participants were divided in two groups: healthy controls and cerebellar degeneration patients.

First, the participants had to adapt to visuomotor rotation introduced in 5-degree steps. There were 4 incremental steps with successive rotations by 5, 10, 15, and 20 deg. The 20-deg rotation was then repeated and followed by 5-deg decrements involving rotations by 15, 10, and 5 deg. Each step involved 16 trials. Next, there was an alternation between blocks of trials with no rotation and blocks of trials with an imposed 20-deg rotation. The rotation was introduced twice, with each block sandwiched between no-rotation blocks.

We used these data to calibrate the model by tuning the learning rate, the degradation rate and the movement endpoint variability magnitude in both control and cerebellar patient groups to fit the adaptation times and asymptotic errors in all cases. Fig 3 represents our simulations which are remarkably similar to the experimental data in Fig 2 from [Schlerf et al., 2013]. Fig 3A shows simulations of 650 trials (reaching movements) as deviations of movement directions from the unperturbed target. The solid black line shows the perturbation angle which changes from 0 to 20 deg and back again in stepwise manner between trials 200 and 400, and abruptly between trials 450 and 600. In the model, after calculating the movement endpoint, we add to it a random number with standard deviation characterizing the magnitude of uncontrolled movement variability. This results in a significant noise in movement direction as seen in Fig 3A. In Figs 3B and 3C to reduce the noise we show the averages over 16 runs for the perturbation epochs only. In this representation, one can clearly see the adaptation process with subsequent saturation of the error at a certain level (the asymptotic error, see Fig 3C).

The saturation of the error happens when balance is achieved between the learning and degradation processes, i.e. when the change in weights due to learning in the cerebellar network is precisely compensated by their decay. Therefore, the asymptotic error is a decreasing function of the learning rate and an increasing function of the degradation rate. In contrast, the adaptation time (time to saturation) is decreasing with both rates which makes it possible to infer the learning rate and the degradation rate from Schlerf et al. data with high accuracy.

In study of Schlerf et al. [Schlerf et al., 2013] there was a striking difference between the asymptotic errors in the control group and cerebellar ataxia patients (see Fig 3C for our simulations reproducing the same). As explained above, greater asymptotic errors can result from the lower learning rate or from faster degradation. Schlerf et al. used a simple Kalman filter model to analyze the experimental data which showed that patients with cerebellar degeneration had slightly higher learning rate and significantly shorter memory corresponding to the higher degradation rate. Besides, the variability of movement endpoint in cerebellar patients was significantly larger than in healthy controls. Our results are largely consistent with Schlerf et al. conclusions. Specifically, the movement endpoint noise magnitude (motor noise hereinafter) was increased to replicate the larger movement variability observed in ataxic patients. Motor noise magnitude per se hardly influenced the adaptation efficiency (not shown). To reproduce data from ataxic patients, we have increased the cerebellar degradation rate, which allowed the model to reproduce the asymptotic errors in this group (Fig 3).

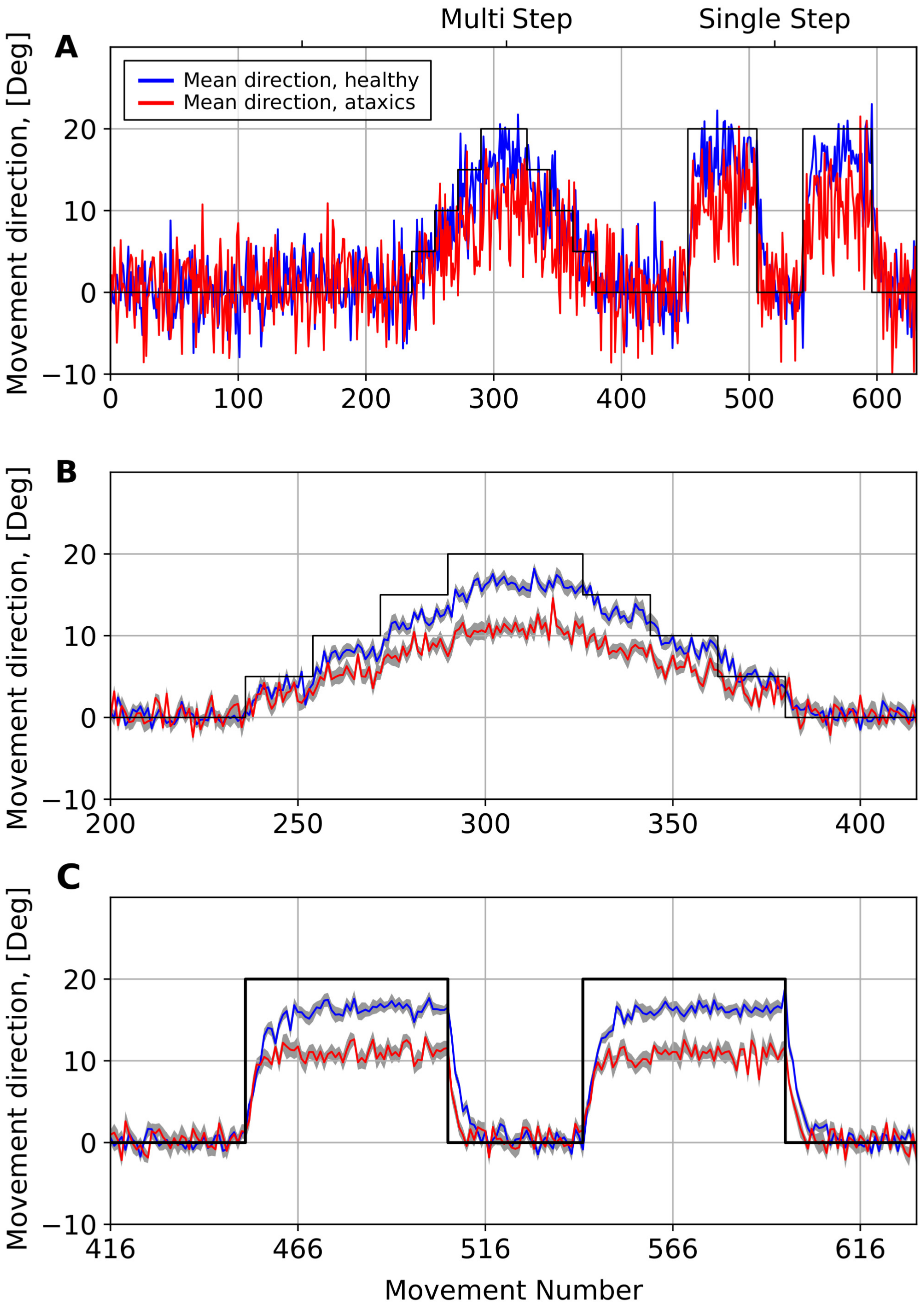

Figure 3 Simulations of adaptation to the visuomotor rotation perturbations. Each plot shows reaching angle dynamics which is defined here as a difference between the perceived target position and the movement direction. In this simulation (similar to Schlerf et al. [Schlerf et al., 2013] experiments) a visuomotor rotation first was introduced and removed in a stepwise manner (Multi Step phase), then it was introduced and removed abruptly (Single Step phase). The solid black line shows the actual target position relative to the perceived target position (rotation angle); the blue line shows simulation corresponding to the control group in Schlerf et al.; the red line shows the simulation with increased degradation rate and increased noise in the movement endpoint to mimic the cerebellar ataxia group data from Schlerf et al. (see text for details); the light grey areas show the standard error of the mean (SEM) over 16 runs. (A) one trial simulation of the control (blue) and ataxic (red) conditions; (B) 16-runs average multistep perturbation adaptation close-up (C) 16-runs average single step perturbation adaptation close-up.

5.2.3 Cerebellum model fails to adapt to strong perturbations.

After calibration, the model presented above provides motor adaptation closely reproducing experimental data in the context of mild visual perturbations. However, as we describe below, it failed to adapt to strong visual perception perturbations. As examples of such perturbations we simulated the rotation by 90 degrees and reflection of the visual field in horizontal direction (see Fig 4). Both perturbations not only required the model to overcome a significant distance between the expected and perceived cursor location but should also deal with strong distortions of the vector error.

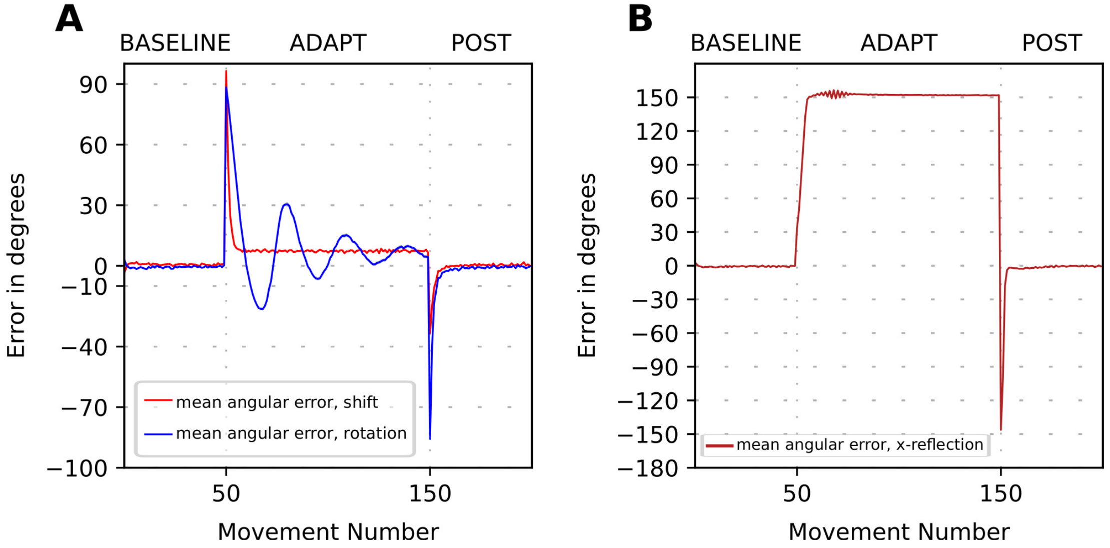

Figure 4 (Lack of) Error-based adaptation to strong visuomotor perturbation. The plot shows movement error in degrees versus trial number during simulation of the 3-phase experiment when a visuomotor perturbation is introduced in the beginning of ADAPT phase and removed in the beginning of POST phase. Solid lines show the 32-run average angle difference between the perceived target position and the perceived hand/cursor position. (A) Simulation of adaptation to 90 degrees visuomotor rotation (blue line) and 90 degrees shift (red line). (B) Simulation of adaptation to the x-reflection perturbation. Note that the perturbation size is the same for the shift and rotation perturbation in (A). However, the trial-to-trial dynamics of the error are very different, as well as the magnitudes of the aftereffects.

Reflection perturbations were previously used in throwing experiments [Gutierrez-Garralda et al., 2013] where a dove prism was used to horizontally reverse the visual field and shift the target's position. Under such a visual perturbation, the sign of x-coordinate of the vector error is reversed, and therefore when a participant misses the actual target to the right, she observes the displacement of the movement to the left of the perceived target. In case of error-based motor adaptation the correction should consist in throwing further to the right, thereby increasing the magnitude of the error distance trial-to-trial. As this did not happen in their experiments, the authors of [Gutierrez-Garralda et al., 2013] concluded that error-based learning was not possible.

We simulated the model of cerebellar correction in strong rotation and reflection contexts (see Methods for details). The simulation results in Fig 4 clearly show that the model performance was much worse when the vector error was significantly perturbed compared to mild perturbations. For the 90 deg rotation, the model produced oscillations instead of gradually converging to the target (compare the shift and rotation cases in Fig 4A). This behavior occurred because the x and y coordinates of the vector error received by CB were transformed by the rotation in such a way, that a flawed correction produced by the cerebellum model did not make the next reaching endpoint to become closer to the target.

During the reflection perturbation, the cerebellar corrections became so faulty that the model diverged from the desired target (Fig 4B). This happened because the sign of the x coordinate of the vector error was inversed. Thus, each adjustment of the cerebellar network weights shifted the reaching endpoint along the x axis in the direction opposite to the desired one. Additionally, since the error magnitude increased for each trial, these faulty corrections became progressively stronger on consecutive trials. In the simulation shown in Fig 4B, the direction of the movement diverges from the target.

Fig 5 illustrates that the performance of the cerebellum model deteriorates continuously with the increase of the perturbation magnitude. In these simulations, we varied the angle of visual rotation and observed the resulting changes in the model's performance. In Fig 5, we show the adaptation phase dynamics. The model was able to successfully adapt to rotations as large as 60 degrees. However, for stronger rotations, the trajectory of the movement endpoint during adaptation started to curl as the vector error components became increasingly distorted by the rotation, as the rotation angle increased. For moderate rotation angles (see 75 deg rotation in Fig 5) the model movement still converged to the desired target in a manner known in ODE theory as "stable focus". However, for rotation angles greater than 80 deg, the convergence to the target no longer happened, and the endpoint trajectory started to spiral around the target (Fig 5C) with progressively larger amplitude as the rotation angle increased.

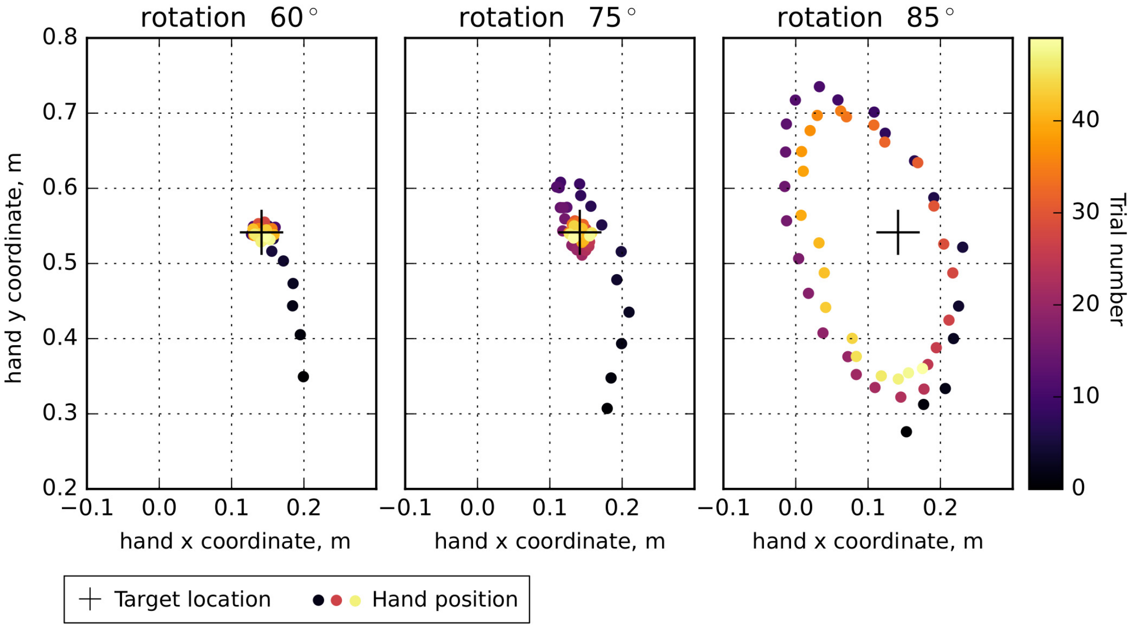

Figure 5 Adaptation to visuomotor rotation with increasing angle. The panels show successive movement endpoints obtained from simulation during adaptation to rotation by 60, 75 and 85 degrees. The color of each point represents the trial number. The '+' shows the target position. Only the adaptation phases of individual simulations are plotted. To emphasize the effect, the degradation rate in the cerebellum was set to zero. As the rotation angle increases, the trajectory first starts spiraling while converging to the target (the middle panel) and then exhibits oscillations around the target with non-decaying amplitude (right panel).

5.2.4 Combining error-based and non-error-based mechanisms

In the previous section, the supervised/error-based learning performed by the model of cerebellum was the only motor adaptation mechanism. As we noted before, it was previously suggested that when error-based learning becomes impossible (e.g. in case of strong visual rotations/reflections, see above), reinforcement/non-error-based learning mechanisms presumably operating in BG are engaged. Thus, we added a model of BG that was responsible for action selection and reinforcement (Fig 6). In this model architecture, BG are involved in action selection hence contribute to formation of the motor program sent downstream in response to a visual cue representing the target position for reaching movements (Fig 6, see [Kim et al., 2017b] for details).

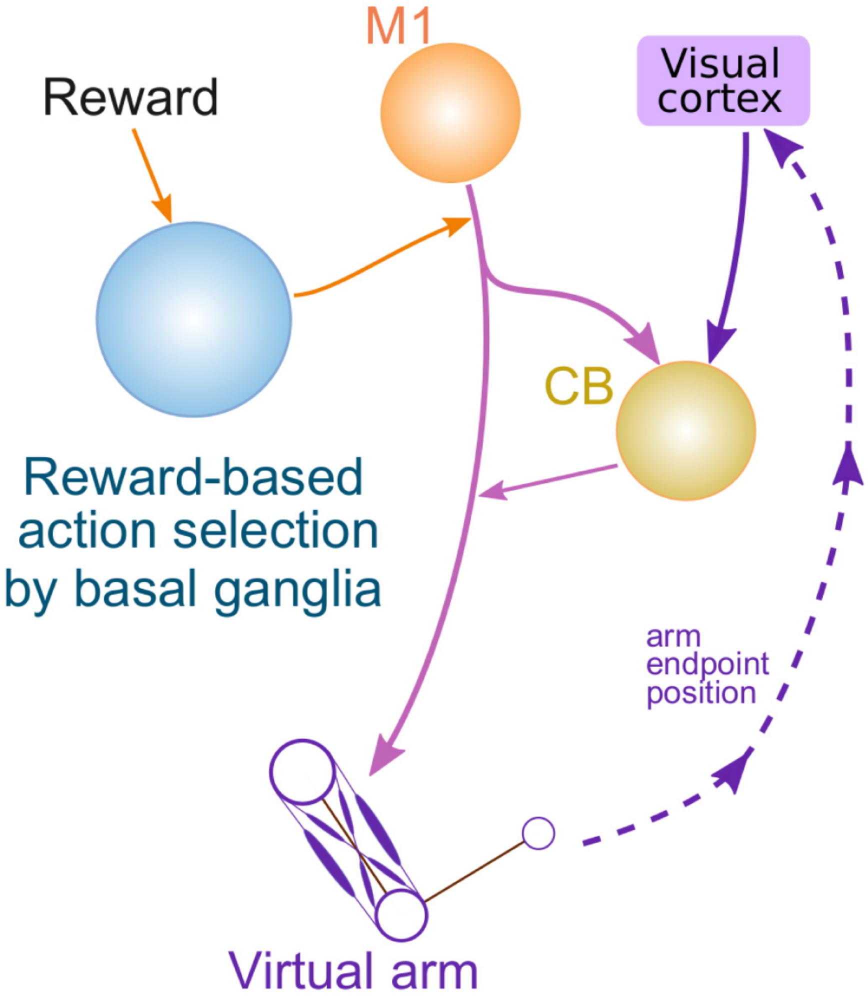

Figure 6 General architecture of the model with cerebellum and basal ganglia. In this version of the model basal ganglia select a motor program for execution among possible behaviors generated in the motor cortex (M1). The cerebellum (CB) receives a copy of the motor program and modifies it before the result is fed to the movement system.

After the movement was complete, the trial performance was evaluated based on how close the movement endpoint was to the target which defined the reward amount. Depending on experimental settings, the reward can be explicitly defined (e.g. fixed reward amount was provided in the case when reaching ended within the target spot) or represented as a measure of satisfaction by the results of movement performed. In the latter case, to calculate the reward amount we used a score monotonically decreasing with the distance from the movement endpoint to the target position (see Methods for details).

The action selection process in BG was modulated by the reinforcement learning procedure. Reinforcement or suppression of the selected actions depended on the sign (and magnitude) of the reward prediction error (RPE) which was implemented as a temporal difference between the rewards obtained on the current and previous trials (see [Kim et al., 2017b] for details). Positive RPE indicated that performance improved on the current trial, and, therefore, the selected action was reinforced. Negative sign of the RPE indicated that the performance worsened, so future selection of this action was suppressed which triggered exploration process for more rewarding actions [Kim et al., 2017b].

Exploration due to negative reinforcement disables error-based motor adaptation in case of mild perturbations. Using the augmented model, we simulated the mild perturbation contexts with the reward provided if the reaching endpoint is within a fixed distance from the target. We found (see Fig 7) that after addition of BG, the model could adapt to a visual shift and mild rotation (not shown), but the time course of this adaptation was different (compare to Fig 4). When the perturbation was introduced at trial 50 (Fig 7), the model missed the target, and the reward was not provided. This resulted in a negative reward prediction error (RPE) which led to a negative reinforcement of the previously performed movement. The BG network suppressed the selection of that movement and initiated the exploration process (see increased variability of the errors between trials 40 and 50 in Fig 7). Eventually, the model generated a reaching movement with an endpoint within the rewarding spot. The provided reward resulted in a positive RPE which reinforced selection of the same movement on the subsequent trials.

Note that during exploration process, the CB network continued to generate corrections of the motor program to reduce the movement error on subsequent trials expecting that the same cortical input would be provided. However, these corrections were not efficient as a

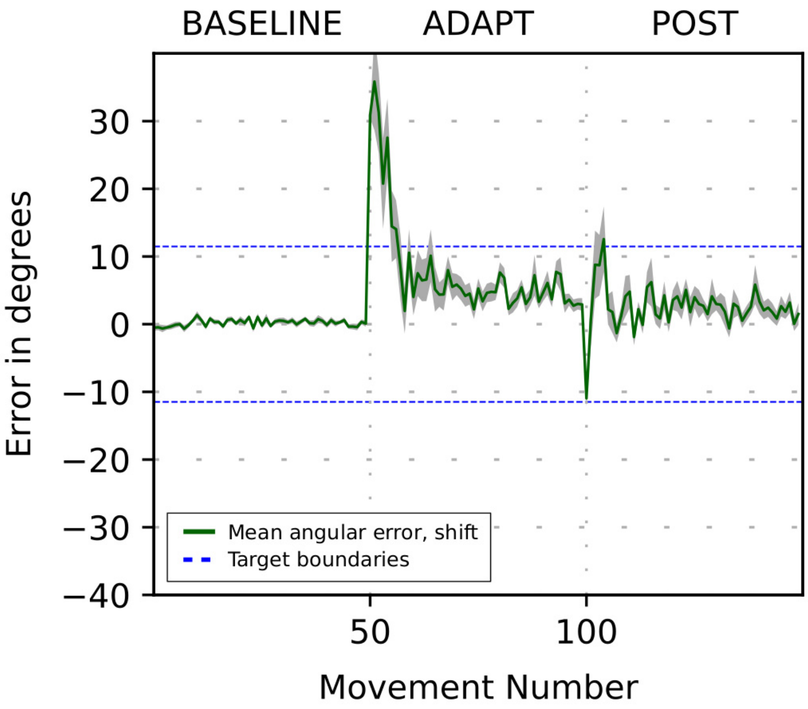

Figure 7 Adaptation to a mild perturbation in the model with cerebellum and basal ganglia. The plot shows error magnitude versus trial number during simulation of the 3-phase experiment when a visuomotor perturbation is introduced in the beginning of ADAPT phase and removed in the beginning of POST phase. Solid line shows 32-runs average angle difference between perceived target position and perceived hand/cursor position. Grey area shows SEMs. Blue dashed lines show the rewarding range.

different action was selected on every trial during exploration. During the adaptation, the accumulated change in the CB network synaptic weights appeared to be insignificant which was evident from the very weak aftereffect as the perturbation was removed after trial 100 in Fig 7.

Faulty CB corrections destroy reinforcement-based adaptation in case of strong perturbations. We have shown above that when the vector movement error is distorted by a strong visual rotation or a reflection of the visual field, the error-based mechanisms lead to divergence of the movement endpoint from the target instead of adaptation to the target. Here we show that the same effect is observed in presence of reinforcement-based mechanisms, too. The simulations of the model with reinforcement and error-based mechanisms combined are shown in Fig 8.

This version of the model also produced oscillations after the rotation by 90 deg (Fig 8A). Similarly, in case of the reflection perturbation (Fig 8B), the time course of the error was virtually identical to the one produced by the model with error-based adaptation mechanism only (compare to Fig 4B). Fig 8B illustrates that the inverted x component of the vector error produced growing cerebellar correction that ultimately resulted in the movement in the direction opposite to the target regardless of the motor program selected by BG.

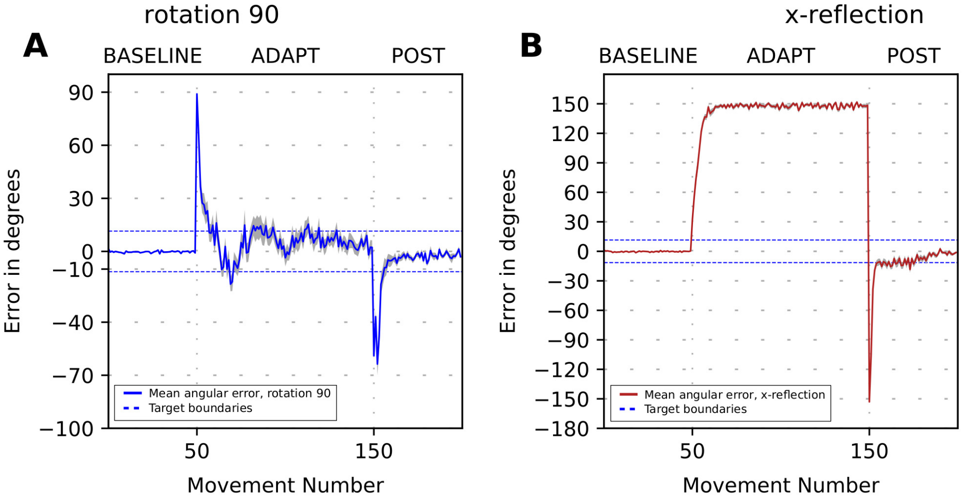

Figure 8 Adaptation to strong perturbations in the model with cerebellum and basal ganglia. The plot shows error magnitude versus trial number during simulation of the 3-phase experiment when a visuomotor perturbation is introduced in the beginning of the ADAPT phase and removed in the beginning of the POST phase. The solid black line shows the 32 runs average angle difference between the perceived target position and the perceived hand/cursor position during adaptation to the 90-degree rotation (A) and the x-reflection (B) perturbations. Dashed lines show the rewarding range. Grey area shows SEMs. BG addition neither improve adaptation to the strong rotation, nor does it make the adaptation to the reflection perturbation possible.

Context-dependent control of error-based learning Results presented in the paragraphs above suggest that visual perturbations, which strongly distort vector error feedback (e.g. visual reflections), make motor adaptation impossible. However, this prediction contradicts the experimental evidence. Indeed, previous studies suggest that the motor control system stops relying on error-based learning mechanisms in situations when perception is strongly perturbed [Gutierrez-Garralda et al., 2013, Telgen et al., 2014]. This can be achieved if there is a mechanism that prevents the cerebellum from further adjusting the motor command given that previous adjustments did not lead to improvement in performance.

As discussed in previous sections, cerebellar corrections become faulty when the vector error provided to CB is unreliable (e.g. x-reflection perturbations). In these situations, the CB's error based (supervised) learning mechanism cannot assist in reaching closer to the target. When error feedback is adequate, it is possible to predict the reduction of the movement error resulting from the adjustment of the CB synaptic weights (see Methods). Therefore, a simple comparison between the predicted and actual changes of the movement error can serve as a robust indicator of the error feedback reliability. Therefore, we propose and implement the performance assessment component (the critic) that compares the predicted and actual vector error in the model (see Fig 9.A). If the two are consistent (i.e. the difference between the predicted and observed errors is small), the critic increases the rate of the supervised learning in CB on subsequent trials (see Methods). If the predicted and actual error changes are inconsistent, the critic reduces the learning rate. Predicted and actual errors are considered consistent (inconsistent) if their difference is within (larger than) a certain threshold. The threshold is proportional to the amplitude of the movement endpoint noise, which is constant (in particular, it is not under the model control, see Methods for details).

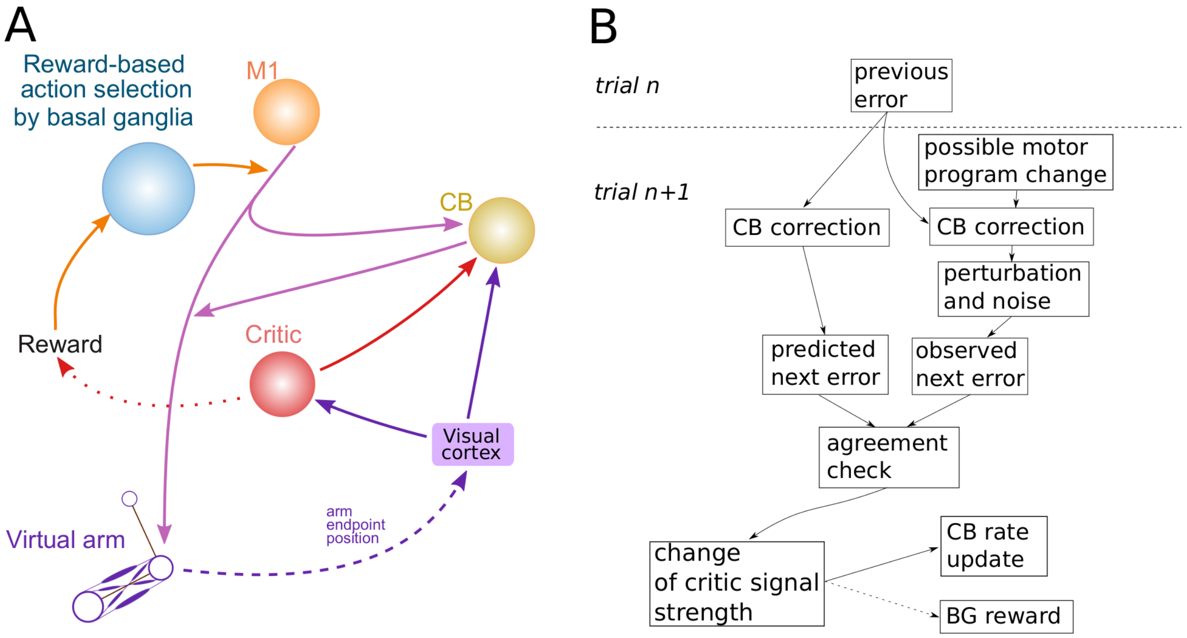

Figure 9 Basal ganglia and cerebellum interaction through the critic and critic mechanism. Panel A: In this version of the model the critic node controls the learning rate of the cerebellum (CB) and offsets the reinforcement signal in basal ganglia. It increases or decreases its output depending on whether the predicted error after correction corresponds to the observed one or not. The critic output is used to set the CB learning rate and add to the dopamine release in striatum. In case the error-based/cerebellar correction is successful, the CB learning rate is increased on the next trial, and the striatal dopamine concentration is increased to prevent negative reinforcement of the previously selected action. In case the error-based correction failed, the CB learning rate is reduced, and the striatal dopamine concentration is not modified thus allowing for negative reinforcement. Panel B: The diagram clarifies how the critic output is computed based on the different pieces of information from the previous (n) and current (n+1) trials. After the previous trial cerebellum modifies the synaptic weights and forms a prediction about the next error value based on that modification. At the current (n+1) trial the actually observed error can differ from the one predicted by the cerebellum due to influence of the factors that cerebellum cannot control and detect directly: possible change of the motor program (e.g. because of basal ganglia activity), perceptual perturbations and noise. If the observed error agrees with the predicted one, the critic signal is increased. If it strongly disagrees, the signal is decreased. Otherwise it is left unchanged. The critic output is used to set the future CB learning rate and add to the reward prediction error when calculating the dopamine level in striatum.

The performance of the model for strong visual rotation and reflection of the visual field is shown in Fig 10. At the baseline, the CB learning rate fluctuated around 3 (Figs 10A2 and 10B2). During the adaptation phase, the critic immediately diagnosed that CB corrections did not produce the predicted effect and reduced the learning rate to near zero. After the perturbation was removed, the CB learning rate returned towards the baseline value.

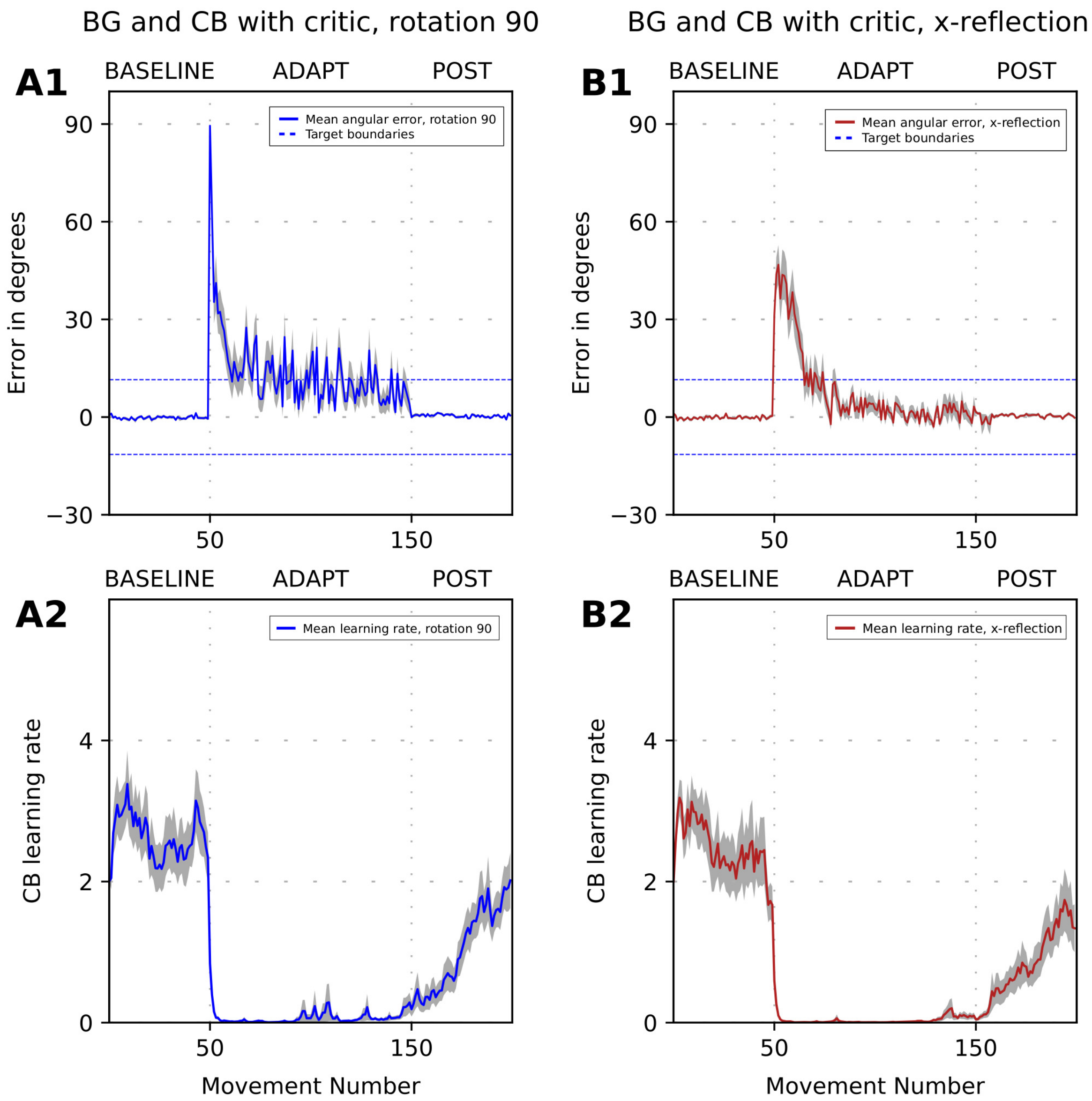

Figure 10 Simulation of adaptation to strong perturbation with the critic-controlled learning rate. The plot shows error size versus trial number during simulation of the 3-phase experiment when a visuomotor perturbation is introduced in the beginning of ADAPT phase and removed in the beginning of POST phase. Panels (A1), (B1) show the 32-runs average reaching error dynamics for the adaptation to 90-deg rotation (A1, blue line) and x-reflection (B1, red line). Panels (A2), (B2) show 32-runs average cerebellum (CB) learning rate dynamics for 90-deg rotation (A2, blue line) and x-reflection (B2, red line). Dashed lines show basal ganglia (BG) target size. Grey area shows SEMs. Critic control of the learning rate allows to adapt to both strong perturbations. Note absence of aftereffects.

As mentioned above, the critic detects faulty cerebellar corrections to prevent further adjustment of synaptic weights in the cerebellar network. Once the learning process is suppressed, the synaptic weights start decaying, which, in turn, gradually diminishes the accumulated faulty correction by means of the cerebellar state degradation. In Fig 10, by the end of the adaptation phase, the cerebellar correction almost fully vanished which is evident from the lack of aftereffects after the perturbation is removed.

Context-dependent control of reinforcement learning Adding the critic component controlling the learning rate in the cerebellum based on the performance assessment helped to suppress faulty error-based adaptation in case of strong visual perturbations. However, in case of mild perturbations (e.g. a visual shift or rotation by a small angle), when the error-based learning was supposed to be the operating adaptation mechanism, our model failed (see Fig 11 and previous sections). Instead, it exhibited reinforcement-based adaptation (triggered by negative reinforcement) and subsequent exploration when the perturbation was introduced. Therefore, in case when the cerebellar correction is efficient, there should be a mechanism that offsets the effect of negative reinforcement and thus prevents exploration.

Reducing BG-induced errors artificially corresponds to turning off BG learning completely – these are simulations shown on Fig 2. The automatic switch happens with the help of the critic mechanism and is described in subsection "Error-based vs. non-error-based learning and effects of neurodegeneration in BG" below. There the BG involvement is suppressed during adaptation to the perturbation where cerebellum could perform without problems.

We propose that the control signal generated by the critic regulates both supervised learning in the cerebellum and reinforcement learning in BG. In the cerebellum, it defines the learning rate as described above. In BG, it adds to the signal representing the reward prediction error (see Fig 9 and Methods). In this manner, negative RPE resulting from the perturbation combines with a positive input from the critic to form a positive reinforcement signal. This provides an efficient mechanism to disable negative reinforcement and exploration in case when cerebellar correction is adequate. In case when cerebellar correction becomes faulty, the signal from the critic changes to zero, which removes positive offset of the reinforcement signal and thus reenables exploration and reinforcement-based motor adaptation.

We explore the regime where BG and CB work together being regulated by the critic in subsection "Error-based vs. non-error-based learning and effects of neurodegeneration in BG". We put this part of the manuscript in "Model validation" section because there we reproduce the study [Gutierrez-Garralda et al., 2013] in the context similar to the ones presented above.

5.2.5 Model validation

In this section we test our model performance against literature data that include experimental paradigms combining both error- and non-error-based motor adaptation.

Error-based vs. non-error-based learning and effects of neurodegeneration in BG To study both the error-based and non-error-based motor adaptation and to test how the BG related neurodegenerative diseases, such as Huntington's (HD) and Parkinson's (PD) ones, affect the non-error based learning, Gutierrez-Garralda et al. [Gutierrez-Garralda et al., 2013] used an experiment during which participants were required to throw balls at a target. During the experiment, either a dove or a wedge prism was used to perturb visual perception of the participants. The dove prism reflected the image about the vertical axis, and wedge prism shifted the visual field horizontally. As mentioned above, the reflection perturbation inverted the sign of the throw mismatch along the horizontal axis, while the shift did not perturb the vector error. Correspondingly, Gutierrez-Garralda et al. assumed that the mechanism of motor adaptation in case of the shift perturbation was an error-based type, while in case of the reflection the participants used a non-error-based motor learning.

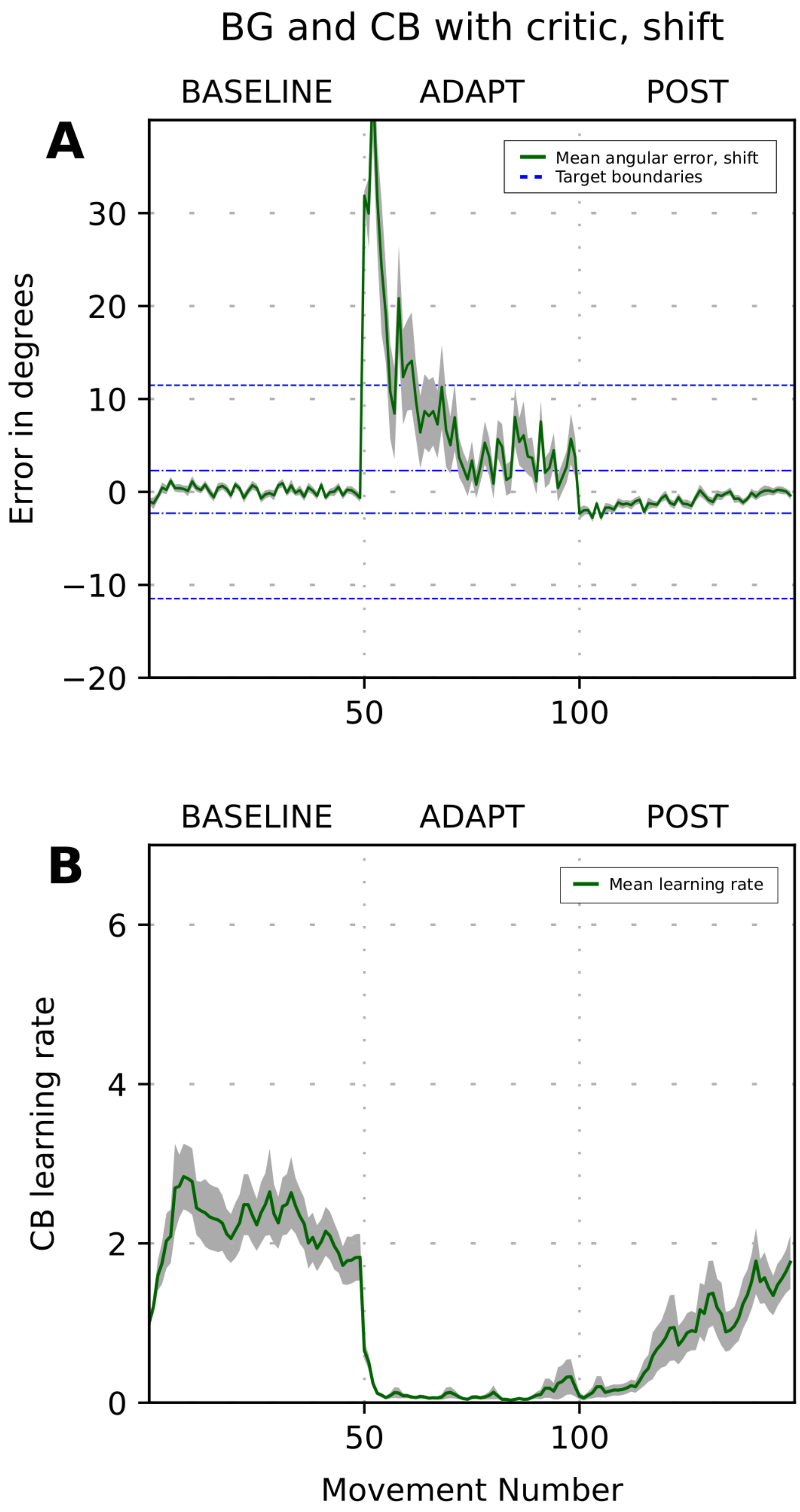

Figure 11 Basal ganglia take over from cerebellum during adaptation to a mild perturbation with the critic-controlled learning rate in cerebellum. The plot shows error magnitude versus trial number during simulation of the 3-phase experiment when a visuomotor perturbation is introduced in the beginning of ADAPT phase and removed in the beginning of POST phase. (A) Solid green line shows 32-run average angle difference between the direction to the target and to the endpoint of the reaching movement; (B) 32-run average CB learning rate dynamics. Dashed lines show the rewarding range, dashed-dotted lines show endpoint noise amplitude. Grey area shows SEMs.

Three subject groups participated in this experiment: the healthy control group, the group of patients with Parkinson's Disease (PD group), and the group with Huntington's Disease (HD group). All three groups were equally successful in adapting to the shift perturbation (see Fig 4A in [Gutierrez-Garralda et al., 2013]) suggesting that PD and HD do not affect error-based learning. However, PD and HD groups were not able to adapt to the reflection as opposed to the control group (Fig 4B in [Gutierrez-Garralda et al., 2013]). Therefore, it was proposed that non-error-based motor adaptation is mediated by reinforcement learning in BG, while error-based learning occurs in some other brain structures, presumably in cerebellum [Izawa et al., 2012, Doya, 2000b, Gutierrez-Garralda et al., 2013, Donchin et al., 2012].

With our model, we simulated an analogous experiment involving reaching movements [Kim et al., 2017b]. In Gutierrez-Garralda et al. experiments, the subjects were not provided with an explicit reward. In our simulations, we assumed that the reward obtained at each trial could be described by a score monotonically decreasing with the error (distance to the target) (see Methods). Therefore, the closer the movement endpoint was to the target, the higher was the reward. PD and HD conditions were simulated in the same way as in [Kim et al., 2017b] (also described in Methods). In short, PD condition involved a reduced rate of reinforcement learning in BG to mimic degeneration of dopaminergic neurons in substantia nigra pars compacta (SNc). Huntington's disease condition was simulated as a reduction in output of the indirect pathway striatal medium spiny neuron population.

Our simulations shown in Fig 12 are in a good agreement with Gutierrez-Garralda et al. experimental data (Fig 4 in [Gutierrez-Garralda et al., 2013]): (1) Simulated adaptation time-course and aftereffects for control and both PD and HD conditions are very similar in case of the visual shift/wedge prism perturbation (Fig 12A); (2) In all conditions there are virtually no aftereffects following the reflection /dove prism perturbation (Fig 12B); (3) In PD and HD conditions, unlike controls, there is no adaptation to the reflection perturbation (Fig 12B).

The mechanistic interpretation provided by the model supports the idea that since visual shift does not perturb the vector of the movement error, all group of participants in [Gutierrez-Garralda et al., 2013] successfully used a cerebellar error correction mechanism (error-based learning [Gutierrez-Garralda et al., 2013]). After the perturbation was removed, the correction, accumulated in the cerebellum, led to a significant aftereffect (Fig 12A). However, in case of reflection perturbation, the cerebellar learning mechanism became inefficient, and, therefore, suppressed by the critic. The participants switch to reinforcement mechanisms mediated by BG (non-error-based learning [Gutierrez-Garralda et al., 2013]). This explains both the lack of the aftereffect and why the adaptation is not occurring in BG-impaired subjects (Fig 12B).

Reinforcement-dependent memory retention In another recent publication by Shmuelof et al. [Shmuelof et al., 2012], action reinforcement was implicated in motor memory retention. The experimental setup used in that study provides an excellent test for our model. In their experiments, human subjects performed fast reaching movements. They received an explicit reward (via a tone) if they succeeded to reach a target within a predetermined distance (see details in Methods, and in [Shmuelof et al., 2012]). Most of the time, they also received complete visual feedback provided on a screen.

There were four slightly different versions of the protocol; all of them included 6 phases: (1) during the baseline phase, participants reached to a target for 20 trials; (2) during the adaptation phase, a 30-degree visual rotation was applied to the image on the screen for 60 trials; (3) during the phase they called “asymptote”, for 80 trials the rotation angle remained the same for the participants to better learn this condition (protocols 1 and 2); however, in some experiments the visual feedback was turned off (protocols 3 and 4); (4) with both visual and reward feedback on, participants had to adapt to a larger 45-degree rotation perturbation for 30 trials; (5) in the error clamp phase, during which false visual feedback showed perfect performance, participants were provided with reward and visual feedback; both indicated that the target was successfully reached regardless of the actual arm movement direction either for 60 trials (protocols 1 and 3) or 100 trials (protocols 2 and 4); (6) finally, in the washout phase, all perturbations were removed (as in the baseline phase) for 40 trials.

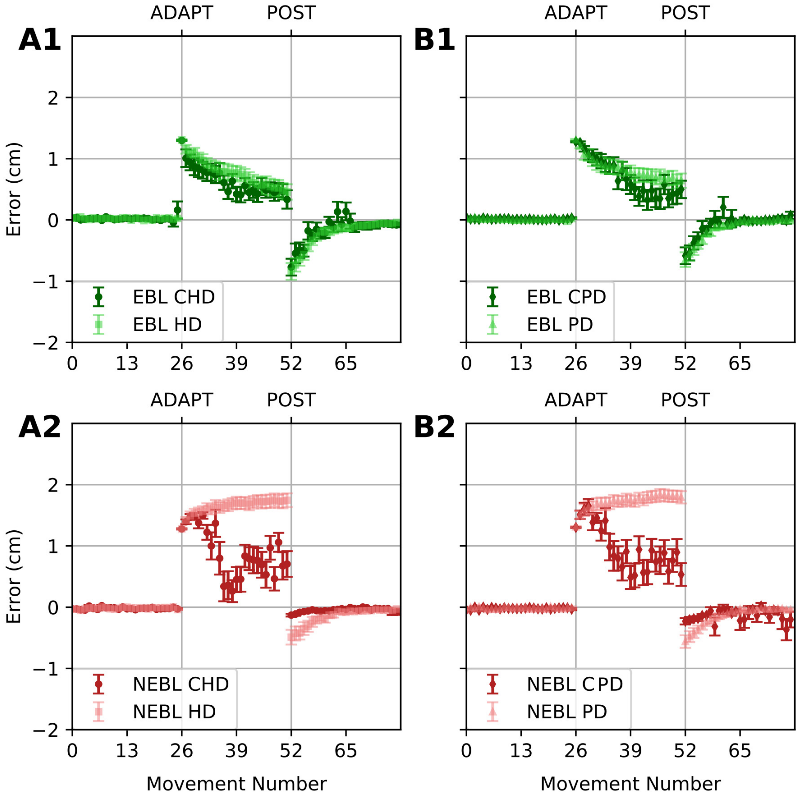

Figure 12 Simulation of error-based and non-error-based adaptation with impaired basal ganglia function. These simulations reproduce Gutierrez-Garralda et al [Gutierrez-Garralda et al., 2013] data. (A1, B1) 16-run averaged reaching error for the adaptation to a 30-degree shift perturbation. (A2, B2) 16-run averaged error for the adaptation to an x-reflection perturbation. Shape and darkness of the makers code different conditions: light squares show the simulation of Huntington's Disease (HD) conditions, and light triangles are used for the simulation of Parkinson's Disease (PD) conditions. Dark circles and diamonds are used to show control conditions for HD (CHD), and for PD (CPD), respectively. Green color is used for the shift perturbation which engages the error-based learning (EBL). X-reflection perturbation simulation (where non-error-based learning (NEBL) is involved) is shown in red. Error bars show SEMs. Controls reduce the error in both conditions but show almost no aftereffects after x-reflection perturbation. In the shift perturbation simulations (A1, B1), the error reduces equally for both control and PD/HD conditions. After the x-reflection (A2, B2), adaptation occurs in control conditions, but not in PD and HD conditions.

The major difference between the described experiments was whether the visual feedback was provided in phase (3) or not. Even though there was no significant difference in subject's performance during this phase (see Fig 2 in [Shmuelof et al., 2012]), there was a striking difference in behavior during the error-clamp phase (5). Specifically, the subjects that were constantly provided with visual feedback during the asymptote phase (5), were gradually converging their movements to the unperturbed target. In contrast, subjects who were not provided with visual feedback during the asymptote phase, adapted their movements to the target position rotated by 30 degrees (same angle used in the asymptote phase). Therefore, the subjects that did not have visual feedback during 30 degrees rotation, were able to "recall" their motor response to this perturbation during the error-clamp phase. Based on this result, Shmuelof et al. [Shmuelof et al., 2012] speculated that during the phase when visual feedback was not provided, the subjects followed a different kind of motor learning process (reinforcement learning), suggesting that reinforcement learning was necessary for motor memory retention. Important to note that Shmuelof et al. reported that only 20% of the participants could consciously detect the presence of the rotation perturbation and even less of them noticed the presence of the error clamp.

Our model reproduces Shmuelof et al. data (compare Fig 13 and 2 in [Shmuelof et al., 2012]) and provides an interesting mechanistic interpretation of their results. In Fig 13 blue curves show the reaching direction vs. trial number for the group who was continuously provided with both the reward (binary error (BE)) and visual feedback (vector error (VE)). Accordingly, this group is labeled "BE+VE". Red curves depict the reaching directions for the group who did not have visual feedback during the asymptote phase (between trials 80 and 160). This group data is labeled "BE". During the simulation, the BE+VE group relied on the cerebellar error-based mechanisms for adaptation exclusively. In contrast, the BE group switched to reinforcement learning during the asymptote phase when no visual feedback was given. So, during the asymptote phase, the BG explored and selected a new rewarding action/motor program. Then, during the 45-degree rotation phase, since visual feedback was provided again, the system switched back to the error-based learning, thus reinforcing the motor program found during the asymptote phase. During the error clamp phase, the degradation of the CB network synaptic weights (the "forgetting" process) gradually diminished cerebellar correction and revealed the unadjusted motor program which appeared different for the two groups. In the BE+VE group, where BG were never involved, this action coincides the default response observed during the baseline phase. In the BE group, however, the motor program was replaced through exploration and reinforcement to the movement in 30-degree direction during the asymptote phase.

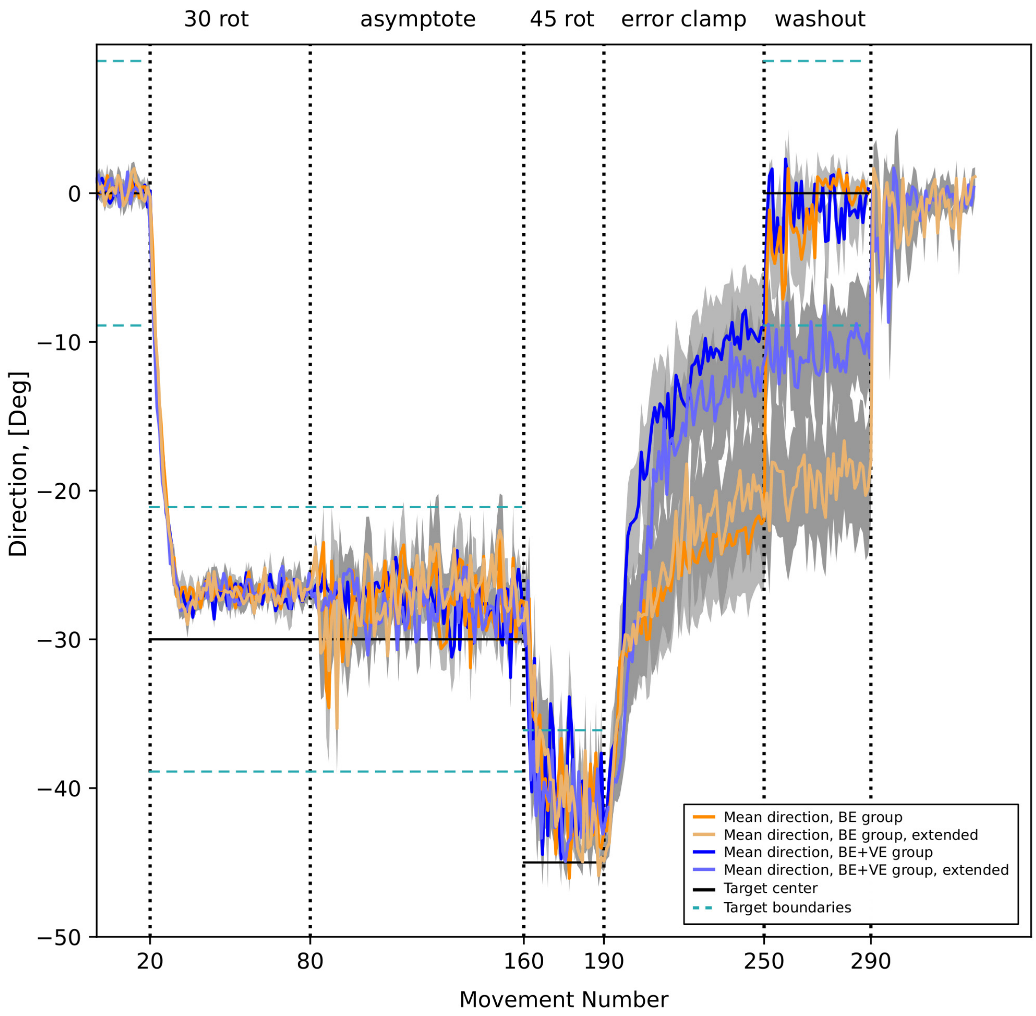

Figure 13 Simulation of the reinforcement-dependent memory retention. These simulations reproduce Shmuelof et al. [Shmuelof et al., 2012] data. Blue lines show the 16-run average of the reaching angle dynamics for the model that constantly received both binary and vector error feedback (BE+VE). Orange lines show 16-run average of the reaching angle dynamics for the model that received only binary feedback between trials 80 and 160 (BE). Lighter blue and lighter orange lines correspond to the simulations where the error clamp was extended by 40 trials. Thick black horizontal lines show the target center, thin grey horizontal lines show target boundaries. Grey areas show SEMs. During error clamp the orange BE curves (unlike the blue BE+VE curves) approximately converge to the level where the binary feedback only was provided between trials 20 and 160.

Learning deficits in cerebellar ataxia and parameters of the critic Previously we have shown that it is possible to reproduce Schlerf et al [Schlerf et al., 2013] data using constant learning rate and increasing the degradation rate for the ataxic condition simulations. Now we show that there is another way of qualitatively reproducing the same data if the learning rate is controlled by the critic. Here we assume that BG are not active, so the critic only controls the learning rate of the CB model.

As explained above, critic decisions are based on whether the difference of the observed and predicted errors falls within a certain threshold, or not. Therefore, by modifying this threshold size, the critic decisions can be modified. We assume that the threshold is proportional to the participant's estimate of its movement variability, i.e. in control conditions it is equal to the standard deviation of the baseline endpoint noise. Our simulations (not shown) confirm that when the model uses critic thresholds corresponding to underestimated movement variability, it leads to reduced adaptation levels. These simulations support the hypothesis about the misjudged movement variability in cerebellar condition formulated in [Schlerf et al., 2013].

For consistency, we also verified that, as in case of the constant learning rate (see Model Calibration section), the increased endpoint noise with movement variability estimate used by the critic adjusted correspondingly, does not produce learning deficits. However, if one additionally increases the degradation rate in the cerebellum, the adaptation deteriorates as in Fig 3.

5.3 Discussion

As previously stated, both basal ganglia and cerebellum may be involved in motor adaptation. Although the functional role of each compartment has been investigated, the interplay between the two structures remains poorly understood [Caligiore et al., 2016b]. We used mathematical modeling to characterize the functional interactions between these structures during motor adaptation.

We introduced a novel model of the network implementing error-based (supervised) learning in cerebellum. Conceptually, this model can be considered a state-space model, like the one discussed in [Thoroughman and Shadmehr, 2000]. However, in contrast with more abstract error-based motor adaptation models, our implementation of the cerebellar network is incorporated into the movement system that produces biophysically accurate movements [Teka et al., 2017b]. There are more detailed models of the cerebellum (e.g. [Luque et al., 2014]), but that was not our goal to model cerebellar networks per se. We were rather interested in mechanisms that engage different learning systems depending on context.

In our study, we assume that the role of basal ganglia is to select an action in response to a visual cue. In the context of reaching movements, an action is a temporal pattern of inputs (a motor program) to the motoneurons which activate arm muscles and, thus, generate a movement to a desired target position. We also assume that a copy of the selected motor program is sent to the cerebellum whose role is to calculate and apply a correction to the motor program that reduces the magnitude of the movement error observed on a previous trial. Under these assumptions, we show that if unregulated, action selection and error correction mechanisms come into conflict. We conclude that only one of them should be active at a time and propose a critic mechanism that controls learning rates in both structures. Specifically, the critic predicts the effect of cerebellar correction calculated on the previous trial and compares it with actual movement error perceived by the subject. If the results are consistent, the critic increases its output which defines the learning rate in the cerebellum and increases dopamine release in basal ganglia. If the observed error differs too much from the prediction, the critic decreases its output, which suppresses learning in the cerebellum and reduces dopamine release in basal ganglia. Note, that comparing the magnitudes of the predicted and observed errors is equivalent to comparing predicted change (reduction) of the error between consecutive trials and the observed change of the error. In the context of adaptation to force field perturbations, the idea of variable learning rate, regulated by consistency of the environment has been previously proposed [Castro et al., 2014]. Our implementation of the critic component follows a somewhat similar idea, though the results are difficult to compare directly because of different experimental protocols being considered.

We show that the suggested critic mechanism provides switching between error-based and non-error-based learning depending on the efficiency of the former. In a nutshell, once the perturbation is introduced, the target is missed resulting in a negative reinforcement (via a reduction of dopamine release) of the movement previously selected by basal ganglia. However, if the correction applied by the cerebellum is successful, the critic facilitates learning in the cerebellum while simultaneously returning dopamine release in basal ganglia to a normal level thus preventing the negative reinforcement of the action from happening. In case the cerebellar correction did not work as planned, the critic stops the learning process in the cerebellum, and allows basal ganglia to block the unsuccessful movement and to initiate exploration of more rewarding actions through the reinforcement learning process.

It is common to classify learning as explicit or implicit [Huberdeau et al., 2015] and in the context of this study we work under the assumption that all the learning happening in the experiments we reproduce with our model is implicit. I.e. we assume that the feedback manipulations and other experiment conditions (such as high movement speed requirements), used in [Gutierrez-Garralda et al., 2013, Shmuelof et al., 2012, Schlerf et al., 2013], have prevented cognitive strategies to develop in participants. We stress that PFC in our model PFC node only provides results of direct sensory processing (like CB error and perception of cue activation). It does not play a direct role in the implicit reinforcement learning.

The neural mechanisms described in this chapter are likely accompanied by further latent mechanisms that could enable more nuanced coordination between supervised and reinforcement learning strategies. However, in the context of “all models are wrong, but some are useful”, we made a simplest possible model that reproduces effects observed in various behavioral experiments. “Simplest” here means within a class of models that follow existing well-established knowledge from electrophysiological and anatomical studies. We have checked several most natural ways of the system organization and ruled out all besides the one with the critic affecting both BG reward and the CB learning rate. Roughly speaking, we show that supervised and reinforcement learning cannot coexist, some switching should happen, provided that one restricts oneself to the classical models of BG and CB like the ones we have used.

We don’t claim that our model is the only one possible, but since to our knowledge it is the first one of this kind, it should be a good start to verify (or falsify) it with more detailed behavioral/electrophysiological studies and create a more accurate model later. We think that currently there is not enough experimental data available to create a plausible model describing more sophisticated interactions between the learning systems. We hope that our model can help to design such experiments.

As noted above, to build our model we had to make a general assumption about the existence of the functional dichotomy between BG and CB. There are, however, recent articles [Wagner et al., 2017, Heffley et al., 2018] challenging this view, so we stress that this dichotomy is a hypothesis and not a well-established fact. Calcium transients in granule cells have been shown to correlate to reward prediction [Wagner et al., 2017]. Parallel fibers could also be a source of contextual input to the cerebellar nuclei.[Heffley et al., 2018] show that calcium transients in climbing fibers contextually correlate to motor outcomes rather than to motor error. The authors indicate that these results support the hypothesis that climbing fibers provide sensorimotor predictions.

In our model the error is encoded somewhere outside of the circuit. We generally presume that it is communicated to the model directly from the cortex, but for the normal functioning of the model it is not really important which structure supplies the error information. Experimental research has identified that the climbing fibers and the dopamine neurons are strong candidates for encoding errors. We see such data contributing to more accurate description of BG (e.g. [Kim et al., 2019]) and CB models separately, whereas we tried to use BG and CB models that are as simple as possible and study interactions between them instead.

5.3.1 Choice of level of detail

Our cerebellum model is essentially a nonlinear version of the standard Kalman filter-based model [Schlerf et al., 2013]. And locally it still works in a linear “gradient-descent” way. We have used a nonlinear model instead of a linear one because we wanted to reproduce experiments involving naturalistic complex movements in 2D space. One of the benefits of such approach is the ability to explore effects appearing for strong visuomotor perturbations like rotations by large angles and visual field reflections.

5.3.2 Adaptation efficiency comparison

Since the critic essentially turns off involvement of either BG or CB in the learning process, the adaptation efficiency of the full model is the same as when the corresponding structure is the only one active from the beginning.

In case when both types of system can lead to successful adaptation reinforcement learning would generally have slower convergence but can potentially stop very close to the target (provided that the rewarded spot it small enough), whereas supervised learning would give fast convergence, but it will stop at the point where degradation (damping/forgetting) in the CB model compensate the learning rate. So in total the answer to the question "Which system adapts better when both potentially can do it?" would depend both on the notion of "better" (e.g. whether one wants fast adaptation or precise adaptation or both) and on a relationship between BG learning rate, reward size, reward spot size, CB learning rate and CB degradation rate.

5.3.3 Physiological basis for control of motor learning

Although some aspects of cerebellar learning have been characterized, the neurophysiology that underlies cerebellar learning is not fully understood. For instance, while it is known that the CB relies on vector error for learning and that a critic mechanism is necessary for certain motor adaptation tasks, it is not yet clear exactly where or how these mechanisms occur in vivo.

Previous studies suggest that a combination of simple (vector error) and informed (predicted and actual error comparison) feedback is necessary for cerebellar motor adaptation. One potential source of informed feedback is the inferior olive (IO), a nucleus in the medulla, which receives sensory input and provides feedback to the cerebellum via climbing fiber inputs. Gellman et al. [Gellman et al., 1983] demonstrated that subpopulations of IO cells respond to specific sensory modalities. Furthermore in [Gellman et al., 1985], they showed that IO cells were excited in response to unexpected stimuli – e.g. if a participant unexpectedly touched an object during exploration – with subpopulations of olivary neurons responding as sensation-specific “event detectors”. When unexpected stimuli occur, excitation in IO neurons elicits “complex” spikes in cerebellar Purkinje cells, thereby strongly inhibiting neuron activity in the cerebellum. Thus, it is thought that the IO inhibits cerebellar output when reliable sensory information is not available, and therefore the IO likely corresponds to the cerebellum-controlling critic in the model. Anatomical evidence suggests that in order to determine the reliability of sensory input the IO may compare error signals from different sources, though the exact nature of this comparison mechanism is not yet fully understood [Zeeuw, 1998].

Although excitation of IO cells produces complex spikes in Purkinje cells, complex spiking can be caused by other inputs to the cerebellum. In a follow-up study to [Gellman et al., 1985], it was demonstrated that complex spikes could occur in the cerebellar Purkinje cells without the introduction of unexpected stimuli [Wang et al., 1987]. Thus, there are other sensory inputs, aside from the IO, that cause complex spikes during cerebellum-directed movements; these non-IO inputs likely provide vector error to the cerebellum during normal learning [Wang et al., 1987].

Another possibility how CB learning rate could be controlled is via some sort of direct control of plasticity in the cerebellum. On the reinforcement learning side the reward modification by critic (which can be a part of IO as noted above) can be done either via modification the signal received by SNc or by direct modulation of dopamine amount in the striatum.

5.3.4 Choice of model parameter values to simulate different experiments

Our model can be easily modified to produce functional motor adaptation consistent with different experimental protocols. Several related experiments have been selected from the literature allowing for investigation of BG-CB interactions across a range of different motor behaviors. By including a variety of experimental protocols in our simulations, we demonstrated that our model allows generalization across different modalities (e.g. types of movement, implicit/explicit reward, etc.). At the same time, model parameters were kept consistent whenever possible so that simulation results related to different protocols could be compared. Yet, certain parameters were adjusted, such as reward components' magnitudes, pools of actions used for different tasks, or movement variability as the experiments we model did not have identical experimental settings. We also had to adjust some of the critic parameters. We believe that differences in experimental conditions justify these parameter modifications. In Schlerf et al. experiments [Schlerf et al., 2013], the participants made slicing movements wearing a cotton glove (to reduce friction between the hand and table surface) to random targets of 1 cm width within 10 cm from the start location. In Shmuelof et al. study [Shmuelof et al., 2012], the protocol required subjects to perform fast reaching movements, having arm supported on a lightweight sled that hovered on air cushions created by compressed-air jets (allowing frictionless planar motion of the arm) to 1 cm wide targets 8cm away from the start location. In Gutierrez-Garralda et al. study [Gutierrez-Garralda et al., 2013], participants made throwing movements. All these experiments address similar types of motor adaptation and are likely to have similar mechanisms involved, but movement-related and reward-related parameters can vary. A detailed experimental study investigating the validity of our assumptions about the parameters in each particular case is desired, though.

In the current work habit formation (increase of the strength of projections from PFC cue-encoding neurons to the thalamus) does not play a major role, since it is a relatively slow process. As we explored in [Kim et al., 2017b], habituation can play an important role in experiments with larger numbers of trials (and also in the pre-learning phase if it is present) compared to the motor learning experiments we were interested in. Here the numbers of trials were relatively small and the results would be the same if we were to turn off the habit formation in the model completely. We did not do explicit pre-learning of the model in this study and just set "habitual" associations artificially in the beginning. It would not be correct to say, however, that habits themselves don't play any role at all in the current study, since the number of the trials when BG starts to explore depends on the strength of the initial habitual associations – roughly speaking, the stronger the initial association, the larger the number of punishments from unsuccessful trials that the BG has to receive before it starts to explore.

The model of CB alone has two main internal parameters: (constant) learning rate and degradation rate. There is also an external parameter – noise amplitude. Changing the learning rate controls the speed of adaptation. Also a nonlinear relation between the degradation rate and the learning rate controls the final error level, where learning is balanced by forgetting (changing the degradation rate has a much stronger influence on it). The performance does not drop much with the increase of the noise level.

We studied parameter sensitivity of CB with critic active in the Schlerf et al [Schlerf et al., 2013] context (healthy subjects and cerebellar patients adapting to abruptly and gradually introduced rotations). We found that in this case the performance (average size of error) worsens for higher noise levels, but only if it is increased without increasing the model's estimate of the environmental variability (i.e. if the noise becomes stronger but the model is aware of that, it performs equally well). There is no explicit learning rate, because it is controlled by the critic. There is however the coefficient regulating the maximal learning rate. Increasing degradation reduces the performance in a way similar to the model with a fixed learning rate. There are also parameters regulating the rate at which the critic increases the learning rate if it thinks that everything works fine, and the rate at which it decreases the learning rate when it thinks the opposite. The effects of changing these parameters are the most prominent under the increased noise condition: both reducing the former and increasing the latter (separately) lowers the performance. See Methods subsection "Simulation with critic active" for exact values and more details.

5.3.5 Error-based adaptation and cerebellar deficits

Via simulation of the experimental results published by Schlerf et al. [Schlerf et al., 2013], we demonstrated that our cerebellar correction model replicates motor adaptation data to single step and multi-step visual rotations. We also reproduced adaptation deficiencies observed in ataxia patients by increasing the degradation rate of synaptic weights in the cerebellar network to represent a reduction in memory related coefficient inferred by Schlerf et al. [Schlerf et al., 2013] using a Kalman filter model.

Interestingly, Schlerf et al. have not found any significant alterations in the learning rate related parameter of their model in ataxia patients, even though their motor variability was increased. This is consistent with Butcher et al. results [Butcher et al., 2017], where it was shown that increased movement noise alone does not lead to deficits in adaptation. Our model also shows no change in the cerebellar learning rate if an increase in movement variability is accompanied by the proportional increase of the thresholds used by the critic. In fact, instead of increasing degradation, one could reduce the learning rate (though the amount of required change is higher) to produce similar changes in the adaptation levels. However in this case these adaptation levels (asymptotic error values) would be reached significantly slower which was not observed in the behavioral data by Schlerf et al. [Schlerf et al., 2013].

Integration of the critic into the model provides additional possibilities to interpret Schlerf et al. [Schlerf et al., 2013] results. Our simulations show that similar learning deficits appear if one changes some of the critic parameters. In particular, the increase in the asymptotic error in multistep condition can be obtained by reducing the movement variability estimate used by the critic to evaluate the efficiency of the cerebellar correction (or increasing the endpoint noise level without adjusting the parameters of the critic). It goes in agreement with Schlerf et al. hypothesis [Schlerf et al., 2013] that ataxics have a wrong estimate of their own movement variability. Similar effects were observed when the reaction of the critic to the threshold crossing was modified. However, the changes of these critic parameters led to a much more dramatic reduction of the learning rate, than the increase of degradation.

In Schlerf et al. analysis [Schlerf et al., 2013] the learning rate of control participants was greater after the single step perturbation, than after the multi-step more gradual perturbation. In our simulations we did not observe this effect. In our simulations the difference of the learning rate in controls in multistep and single step conditions observed for some parameter values happens because in multistep condition unexpected effects (noise and perturbations) are smaller on average, than in a single step condition, thus the critic is less likely to reduce the learning rate. In a broader context, in our model both underestimating and overestimating the uncontrollable movement variability by the critic can have negative effects on learning. Underestimated variability can lead to overly strong reduction in the learning rate for mild perturbations, which results in reduced ability to adapt using error-based learning and may trigger BG-driven exploratory behavior. Overestimated variability tends to drive cerebellar learning rates close to maximum possible and prevent the system from switching to non-error-based learning mechanisms, if it becomes necessary.

5.3.6 Predictions

The model has implications that can be tested experimentally. It predicts, for example, that for very precise movements, where the expected movement variability is low, the participants are less likely to adapt even to mild perturbations, in comparison to the case when their movement variability is artificially increased. The good test for the model would be a Gutierrez-Garralda-like experiment with reaching movements under the rotation perturbation instead of the Dove prism and with explicit rewards. It would be important to check whether the observed effects qualitatively change for different movement variability magnitudes. Another way to test the model is to perform a series of experiments with different visuomotor rotation angles and check at what angles participants switch from error-based to non-error-based learning.

Another important prediction of the model is variations of the striatal dopamine concentrations. Specifically, our model predicts that during error-based adaptation the dopamine levels are significantly higher compared to the non-error-based scenario. To test this prediction, one could measure the striatal dopamine concentration during adaptation to visuomotor perturbations implying different (error-based vs. non-error-based) learning mechanisms.

5.4 Conclusions

The proposed model of motor adaptation to perceptual perturbations includes two distinct learning structures, cerebellum and basal ganglia, responsible for error-based and non-error-based motor learning, respectively. We demonstrate that based on existing experimental evidence, it is necessary to have a mechanism regulating involvement of cerebellum and basal ganglia depending on the perturbation. We suggest that the involvement of a particular learning mechanism is regulated by the same signal depending on the consistency of the internal model used by the brain to predict the movement results. The resulting model reproduces data from several experiments, involving interaction between error-based and non-error-based learning mechanisms.

5.5 Methods

We distinguish between the model of the motor adaptation system, which is supposed to reproduce activity of some parts of the participant's brain, and the model of the experimental environment, which is supposed to describe factors and events that the participant does not control. We call the latter "a context".

The model aims at representing a human subject performing reaching movements with indirect visual feedback of her/his arm during a sequence of trials. We assume that the hand position is not directly visible but is displayed on a screen as a cursor. At the beginning of each trial the participant places his hand to a fixed starting point and the attempts to move the arm to a target position appearing on the screen 20 cm away from the starting position. The visual feedback (hand position-cursor position correspondence) can be artificially perturbed during the experiment. Depending on the context, the participant receives a reward if the movement endpoint occurs within the target spot.

We used different models of the motor adaptation system of increasing complexity and different contexts. Each context corresponded to one of the experimental protocols, which included experiment-specific alterations of the visual feedback and reward paradigms.

5.5.1 Neuro-mechanical model of the arm

To simulate center-out reaching tasks, we previously designed an arm model simulating 2D center-out reaching movements described in details in [Teka et al., 2017b]. In short, the model consists of two rigid links connected by hinge joints (shoulder and elbow) actuated by six Hill-type muscles [Harischandra and Örjan Ekeberg, 2008]. These include four single-joint muscles: the shoulder flexor (SF) and extensor (SE), the elbow flexor (EF) and extensor (EE), which control rotation of either the upper arm or forearm around the corresponding joint. The other two muscles, namely: bi-articular flexor (BF) and extensor (BE) are two-joint muscles that attached to both joints and simultaneously control movement around them. The arm movement depends on the combination of multiple muscle activations and is restricted to the horizontal plane. The dynamics of the arm motion is derived from the Lagrange equations, which take into account the Coriolis and centrifugal forces, joint viscoelastic forces, and muscle forces [Teka et al., 2017b]. The proprioceptor afferent feedback from each muscle (Ia and Ib) projecting to the spinal cord was derived and modified from [Prochazka, 1999].

The model of spinal cord comprises complex interconnections among motorneurons, interneurons, including Renshaw cells, Ia- and Ib- inhibitory interneurons and correspondent afferent feedbacks (see [Teka et al., 2017b] for details). The spinal circuitry receives descending inputs from the motor cortex that activates corresponding motoneurons, which in turn drive arm movements via activation of muscles. Interneurons and Renshaw cells mediate interactions within spinal circuits and modulate cortical signals to the arm muscles. In addition, spinal reflexes, namely: stretch reflex, autogenic inhibition reflex and recurrent inhibition of motoneurons, which play an important role in arm kinematics [Franklin and Wolpert, 2008], are incorporated into the model of spinal cord. The detailed description of the spinal circuitry can be found in previous publications [Teka et al., 2017b, Franklin and Wolpert, 2008, Markin et al., 2015].

The position and movement of the arm is controlled by the time varying input descending from cortex which we refer to as a motor program. So, the motor program is a time-dependent 6-dimensional vector of signals innervating the spinal cord circuitry, $\boldsymbol{p}(t)$. To actuate reaching movements, we implemented cortical neurons as a cortical "controller" that solves an inverse problem

$$F(\boldsymbol{p}, 0) = \vec{x}_0$$

where $\vec{x}_0$ is the target position. The inverse problem solution is based on a proposed straight-line trajectory to a target position and a predefined bell-shaped arm endpoint velocity profile. Thus, the controller generates a motor program that produces a task-specific activation of lowlevel spinal circuits that in turn induce the muscle activation pattern realizing the intended reaching movement. Therefore, calculation of the correct motor program relies on the information about target position which in turn depends on how reliable visual feedback is. During visual perturbations, the target position is perceived incorrectly, which thus results in a displacement of the movement endpoint relative to the actual target position, i.e. in a movement error.

5.5.2 Model of cerebellum

The role of cerebellum in the model is to calculate a correction to the cortical motor program that brings the movement endpoint closer to the target position. This correction is assumed to be additive and is calculated based on the information about the movement error acquired on the previous trial. The corrected motor program $\boldsymbol{p}_{cor}(t)$ is calculated by linearly transforming the original motor program

$$\boldsymbol{p}_{cor}(t) = (1 + W)\boldsymbol{p}(t)$$

where W is a 6×6 correction matrix. This correction can be viewed as a result of processing the initial motor program by a linear artificial neural network mapping 6 inputs to 6 outputs with synaptic weights W. Weight matrix update follows classical error back propagation algorithm which is a one-step iteration of the gradient descent method to minimize the magnitude of the vector movement error $\vec{e}$:

$$W_{new} = W - \frac{1}{2}\lambda \frac{\partial}{\partial W} \vec{e}^T \vec{e} - \gamma W,$$

where parameter λ defines the convergence speed, and γ defines exponential decay of the synaptic weights in absence of learning signal (forgetting). Hereinafter we refer to λ and γ as learning and degradation rates, respectively. Taking into account that the movement error is the vector difference between the movement endpoint position and the target position $\vec{x}_0$, we have

$$\vec{e} = F(\boldsymbol{p}, W) - \vec{x}_0,$$

where $F(\boldsymbol{p}, W) := F\left(\boldsymbol{p}_{cor}, 0\right)$ is the movement endpoint as a functional of the initial motor program $p(t)$ and the correction matrix W as calculated by the neuro-mechanical arm model with the corrected motor program. Then the equation for the weights update takes the form

$$W_{new} = W - \lambda \frac{\partial F}{\partial W} \vec{e} - \gamma W,$$

where $\partial F / \partial W$ is a $6 \times 6 \times 2$ tensor whose components are the derivatives of the end movement positions with respect to correction matrix entries.

5.5.3 Basal ganglia model

The model of reinforcement learning in basal ganglia we used in this study was previously published and is described in details in [Kim et al., 2017b]. Briefly, the model is an extension of the classical two-pathway BG model from [Frank, 2005] to the case of many possible actions. Here, we only provide short qualitative description of our model. Behavioral experiments studying reinforcement learning mechanisms assume that a choice must be made between several differentially rewarding behavioral options. Unlike decision-making tasks, motor learning does not imply a small or finite number of possible choices. The only constraint is the context of the task, e.g. reaching from a fixed initial position to an unknown destination. Our model has unlimited number of possible actions. As the context, we used center-out reaching movements performed in a horizontal plane. To calculate cortical activity corresponding to different movements, we explicitly solved an inverse problem based on the given arm dynamics as described above. Accordingly, for every possible reaching movement we could calculate the corresponding motor program represented by the activity profiles of cortical inputs responsible for activation of different muscles. To describe different experiments, we define corresponding (arbitrarily large) sets of motor programs that define all possible behavioral choices (actions) in each experimental context.

The classical view of action selection is that different motor actions are gated by thalamocortical relay neurons. In the presented model, we assume that relay neurons can be activated at different firing rates, and their firing rates define contributions of different motor programs to the resulting motor response. More specifically, in our model cortical input to the spinal network is implemented as a linear combination of all possible motor programs in the given context with coefficients defined by the firing rates of corresponding thalamocortical relay neurons. This linear combination can be viewed as an aggregate input to the spinal network from the cortical motoneurons exhibiting activity profiles corresponding to different motor behaviors, e.g. reaching movements in different directions.

The classical concept of BG function is that the BG network performs behavioral choice that maximizes reward. This action selection process results in activation of thalamic relay neurons corresponding to the selected action and suppression of neurons gating other behaviors. Per this concept, each action is dedicated to specific neurons in different BG nuclei. Their focused interconnections form action-related loops which start at the cortex, bifurcate in the striatum into direct and indirect pathways converging on the internal Globus Pallidus (GPi), and feed back to the cortex through the thalamus. Action preference is facilitated by increased excitatory projections from sensory cortical neurons representing the stimulus to direct pathway striatal neurons (D1 MSNs). Suppression of unwanted competing actions is assumed to occur because of lateral inhibition among the loops at some level of the network in a winner-takes-all manner.

The classical model predicts that novel cue-action associations acquired based on reinforcement learning rely on BG network integrity. However, multiple experimental studies have shown that pharmacological blockade of GPi, the BG output structure, does not lead to significant impairments in performing well-learned tasks. Consistent with these experiments, it was suggested that acquired associations may become "direct" projections within the cortex bypassing the BG network. In our implementation of the model, competing actions are suppressed by lateral inhibition in the population representing thalamocortical neurons, independent of BG network integrity.

In the model, novel cue-action associations are formed based on reinforcement learning in the striatum. Eventually, the preferable behavior is reliably selected due to potentiated projections from the neurons in prefrontal cortex (PFC), activated by the provided stimulus, to D1 MSNs, corresponding to the preferred behavior. On a longer timescale, repetitive execution of the same action in response to the same stimulus leads to habituation of the response via long-term potentiation of the direct projections between the corresponding PFC and thalamocortical relay neurons based on Hebbian learning. At the same time, due to degradation of synaptic connections between the PFC and striatum in absence of reinforcement, BG involvement in the action selection process gradually decreases. Eventually, the behavior becomes a habit, which is automatically selected solely based on direct cortico-cortical projections.

In technical terms, the output of basal ganglia model is the activation levels of thalamocortical relay neurons in response to the input from PFC neurons activated by visual cues. Each cure represents one of the possible reaching targets. These levels are used as coefficients of the linear combination of all possible actions which represents the motor program selected for execution. The resulting motor program is used to calculate the endpoint of the movement using Neuro-Mechanical Arm Model (see above). Depending on the distance between the movement endpoint and the target position, the reward is calculated as dictated by the experimental context (see below). This reward value is used to calculate the reward prediction error (RPE) as a temporal difference between the current and previous reward values. The RPE is used as the reinforcement signal to potentiate or depress synaptic projections from PFC neurons, activated by the visual cue provided, to the striatal neurons, representing the selected actions.

Mathematically, the model is implemented as a 700-dimensional hybrid dynamical system made of distinct 7 populations of rate neurons (D1 MSN, D2 MSN, GPe, STN, GPi, thalamus), each containing 100 rate neurons (each representing a neuronal subpopulation); every neuron corresponds to a one of the parallel loops within basal ganglia-thalamocortical circuit. The parallel structure is only violated in the thalamus node where each loop inhibits all other loops, creating a “winner takes all effect” which prevents strong simultaneous activation of different loops. The simulation of a single trial goes as follows. First, the rate differential equations are solved until a point attractor is reached (which empirically always happens with this system). This point attractor describes activities of all the 700 rate neurons, including the ones in the thalamus. We then take the activities of the thalamic neurons and use them as input to the neuro-mechanical arm model, which in turn produced a movement. After that the reward is supplied (or not) to the model, depending on whether the movement endpoint was in the proximity of the target center (target size is a model parameter). Using the reward value the model computes the RPE and updates the reward prediction value. Finally, the RPE value is used to update to strengths of connections between the PFC and D1 MSN, D2 MSN. PFC neurons (whose number is equal to the number of cues in the experimental setup) are not simulated and are modeled simply by real numbers meaning that they provide constant signal while BG reaches a decision.

See details in [Kim et al., 2017b].

5.5.4 Model of the Critic

In the presence of visual perturbation, cerebellum may produce corrections that do not lead to improvement in movement accuracy. To cope with that, the critic in the model predicts the result of the implemented correction and compares it with the actual improvement in performance on the next trial. To calculate the expected error after correction, the critic uses the internal model

$$\vec{e}_{exp} = F\left(\boldsymbol{p}, W_{new}\right) - \vec{x}_0$$

where W_{new} is an adjusted correction matrix as described above, and $\vec{x}_0$ is the actual target position.

The critic compares the expected error with the error $\vec{e}_{real}$ observed on the next trial which is subject to possible visual perturbation. If the observed error agrees well with critic's prediction, then the critic concludes that CB is functioning correctly and increases the learning rate λ in the cerebellum for future adjustments. If there is no agreement, the critic assumes that either the cerebellar correction was not the main reason for the error change (e.g. the altered motor program), or CB was relying on the distorted error when adjusting the correction matrix. In both cases, the critic decreases the learning rate to suppress faulty cerebellar learning. Specifically, the critic calculates the magnitude of the difference between the errors

$$\text{mismatch } = |\vec{e}_{\text{real}} - \vec{e}_{\text{exp}}|$$

and then forms its output

$$\text{critic} = \begin{cases} \text{speed}_{up} \cdot \lambda, \ \text{mismatch } < t_{\text{low}} \cdot \kappa \\ \lambda, \quad t_{\text{low}} \cdot \kappa \leq \text{mismatch} \leq t_{\text{high}} \cdot \kappa \\ \frac{\lambda}{\text{slow}_{\text{down}}}, \text{mismatch} \geq t_{\text{high}} \cdot \kappa \end{cases}$$

Here the critic compares the mismatch between predicted and perceived errors with two threshold values. These thresholds are multiples of the estimate of the movement variability κ which is assumed to be equal to the standard deviation of the endpoint noise. The critic treats any mismatch less than t_{low} times the endpoint variability estimate κ as consistent with prediction, any mismatch greater than $t_{\text{high}} \cdot \kappa$ as inconsistent with prediction, and any mismatch in between these threshold is treated as inconclusive. The critic output is used to set as a new value for the learning rate λ for the next trial. Therefore, in case of the consistent prediction it is increased by a factor of speed $_{up}$, in case of the inconsistent prediction it is reduced by a factor of slow $_{\text{down}}$ or remains unchanged if the test is inconclusive. The values of the parameters used are specified in each simulation context.

We bound the critic output by 0.001 from below and by $a_{opt} \cdot \lambda_{opt}$ from above, where $\cdot \lambda_{opt}$ is defined as the value of λ that leads to a complete elimination of the vector error in one step ($\vec{e}_{exp} = 0$) in absence of any visual perturbation. Also, if the current error magnitude is extremely small (less than 1mm) the critic signal is accepted equal to current value of λ, i.e. the learning rate does not change. The critic signal is also used to offset the reinforcement signal DA in the basal ganglia mode:

$$DA = RPE + \alpha \cdot \text{ critic}$$

where RPE is the reward prediction error, and coefficient α depends on the context (see below). In case cerebellar corrections are efficient in reducing the movement error, the critic output is close to its maximum, which ensures positive reinforcement of the currently selected action regardless of the reward provided. When cerebellar corrections are faulty, the critic output is close to zero, which suppresses learning in cerebellum and creates prerequisites for the reward-based learning in basal ganglia.

5.5.5 Simulation algorithm

The general flow of information in the model is as follows. First, basal ganglia select a motor program to be executed. Then cerebellum modifies the program and sends it as an input to the neuro-mechanical model of the arm which executes the program and calculates the movement endpoint which is used to calculate the vector error and the reward. Because of visual perturbation, the perceived vector error may be distorted depending on the experimental context. The critic evaluates whether cerebellum correction worked as expected and sets its output accordingly. Then basal ganglia module updates synaptic weights of the projections from PFC to striatal neurons based on the reward received and the critic input, and cerebellum updates the correction matrix based on the vector error received and the learning rate set by the critic.

More formally, the simulation of each trial can be divided into following stages, some of which can be skipped depending on the particular model type (see text and Fig 9.B):

1. Selection of the motor program by the BG model, based on presented cue.

2. Application of the CB correction using current CB correction matrix.

3. Use of the resulting motor program as an input to the arm model to simulate the movement and calculate reaching endpoint.

4. Addition of Gaussian (with mean zero and standard deviation 0.005m = 5mm noise to the endpoint and application of the visual perturbation that transforms the vector error.

5. Critic evaluation of the predicted and perceived error agreement.

6. Calculation of the reinforcement signal as a sum of the reward and the critic signal.

7. BG learning: use of the reinforcement signal to update the synaptic weights of PFC projections to the striatum.

8. CB learning: Update of the correction matrix based on the vector error, and the critic signal as a learning rate.

5.5.6 Simulation contexts (models of experimental protocols)

Description of each of context includes the list of targets and (initial) cue-action associations used, sequence of cue activations, reward description as well as some additional minor details.

Test 3-phase contexts First, we describe simple contexts that we use to demonstrate model performance using progressively augmented architecture (Figs 2, 4, 7, 8, 11, 12). In total we use three sensory cues (C1 and C2, C2') and three actions (A1 and A2, A2') which corresponded to reaching toward North-East (T1, 45 degrees), North-North-East (T2, 75 degrees) and North-West (T2',135 degrees) targets at 20 cm distance from the starting point. We set habitual associations $C1 \rightarrow A1$, $C2 \rightarrow A2$ and $C2 \rightarrow A2$ represented by direct projections from PFC neurons (cues) to thalamocortical relay neurons (actions) to be of strength 0.3 (see [Kim et al., 2017b] for details). In each context we use only two cues and two actions. Then the virtual experiment has the following protocol:

1. Baseline: for 50 trials of the simulation, cue C1 is activated.
2. Adaptation: for 50 (or 100) trials, a visual perturbation is applied.
3. Post-effect: Finally, for the last 50 trials the conditions were returned to the ones used during first 50 trials.

The visual perturbation is fixed for every context and can be of the following types: 1) shift perturbation in reaching context: a switch of the presented cue from C1 to C2 (or C2') while keeping the vector error unaltered directly. 2) Rotation perturbation: a switch of the presented cue from C1 to C2 (or C2') and CCW rotation by 30 (or 90) degrees is applied to the movement endpoint (hence also to the vector error). 3) X-reflection perturbation: a switch of the presented cue from C1 to C2 and x-coordinate reversal in the movement endpoint (which, in turn, affects the vector error in the similar manner).

We label the protocols in the following way.

- Mild shift context: 3-phase context with 30 degrees shift perturbation during 50 trials;
- Mild rotation context: 3-phase context with 30 degrees rotation during 50 trials;
- Strong shift context: 3-phase context with 90 degrees shift perturbation during 100 trials;
- Strong rotation context: 3-phase context with 90 degrees rotation during 100 trials;
- Reflection context: 3-phase context with reflection perturbation during 100 trials.

For all these contexts, the pool of possible actions consisted of 100 reaching movements with endpoints uniformly distributed from 20 to 155 degrees on a circle with 20 cm radius and a centre at the initial hand position. When the reinforcement learning was enabled, the reward of magnitude 3 was given for reaching within a circular target spot of 4 cm size, centered at the target. The initially reward was set to 3 as well, assuming that subjects were pretrained to perform reaching movements and getting reward of this particular magnitude for reaching within the target spot. For strong shift, strong rotation and reflection perturbations we used the cerebellar learning rate = 2.

Schlerf et al. context [Schlerf et al., 2013] The context uses 13 cues with habitual associations with movement to directions from 60 to 90 degrees with a 2.5-degree step, to represent different targets used in the actual experiment. At each trial one of these cues was randomly activated. To simulate trials without visual feedback we did not update the CB correction matrix after them.

The endpoint noise had $SD = 1.05$cm. For the controls the learning rate was set to 1. The pool of actions was 100 reaching movements in directions uniformly distributed from 58 to 92 degrees.

To replicate the altered adaptation observed in ataxic patients, we increased the degradation coefficient γ in the CB model from 0.04 to 0.16 to replicate the experimentally observed asymptotic error. The arm endpoint noise SD was increased to 0.013 to replicate the larger movement variability observed in cerebellar patients.

Simulation with critic active For the version of the context with critic active, the critic thresholds were $t_{low} = 2, t_{high} = 3.5$ and the speeds $speed_{up} = slow_{down} = 1.3$. The optimality coefficient was $a_{opt} = 0.2$, to limit the learning rate values-for this setup due to constantly changing targets and large noise the learning rates tend to get high values for our critic parameters. The movement variability estimate was equal to the noise SD.

We have used several modifications of parameters for controls, which can be regarded as a study of how critic performance depends on its parameters. Increasing both noise to 0.013 and estimate of the variability to 0.013 does not lead to any change both in rates and in adaptation levels. If we additionally increase degradation to 0.016 (like in the constant rate case), we reduce the adaptation level, but the rates stay the same. In fact, for our parameter values increasing noise to 0.013 without adjusting movement variability does not have a strong effect. Though if we do the same, while decreasing the estimate of the variability, the effect is significant. The same happens if we apply stronger noise (not shown). Keeping the noise size at controls level, one can decrease the adaptation levels by reducing the optimality coefficient to 0.05 or the movement variability estimate alone to 0.88cm. In general, increasing noise and decreasing the movement variability estimate has the same qualitative effect (one can get a quantitative version of it too). The fact that increasing the noise to 0.013 did not produce a strong effect only reflects the fact that we increased the noise not strongly enough to match the decrease of the movement variability estimate to 0.88cm (which can be translated into sizes of thresholds t_{low} and t_{high}). For the control noise value changing speed $_{up}$ and $slow_{down}$ within large ranges did not produce almost any effect, but for the increased level of noise (without movement variability adjustment) one can get reduced adaptation levels by either reducing the speed$_{up}$ to 1.12, or increasing the slow to 1.7.

Like in the constant rate case, the adaptation levels the model produce are rather stable to the learning rate amplitude alterations (the most direct analog of the strongly reduced constant learning rate is strongly reduced optimality coefficient), but final adaptation level takes longer to achieve for smaller rates. Not surprisingly, unlike degradation increase, critic parameter modifications affect adaptation to single step perturbations stronger than multistep adaptation, because the model encounters stronger perceptual perturbations. Even though during the multi-step perturbation the perturbations reaches the same magnitude, the formula for the critic we use works with changes of the error rather than with the error itself, otherwise it would not be able to adapt to perturbations strong in magnitude, but not changing perception (e.g. large shifts).

Gutierrez-Garralda at al. context [Gutierrez-Garralda et al., 2013] We used two sensory cues (C1 and C2) corresponding to actions A1, A2 of reaching to targets T1 at 70 degrees and T2 at 110 degrees, respectively.

We set the habitual associations for both cues (C1-A1 and C2-A2) represented by connections from corresponding PFC neurons to thalamocortical relay neurons with initial strength 0.55 (see [Kim et al., 2017b] for details). During the simulation, the target T1 corresponded to the actual target position, and the target T2 corresponds to the distorted target position as perceived during the visual perturbation.

For the first 25 trials, the cue $C1$ was activated, and reward was provided for reaching to the target T1. Then for trials 26 to 50 , the activated cue was changed to $C2$, but the reward was still provided for reaching to the target T1. Finally, in the last 25 trials the activated cue was changed back to C1, and reward was provided for reaching to the target Tl again. The rewarding spot radius ($D_{\max}$) was 4cm. The pool of actions consisted of 100 reaching movements in directions uniformly distributed from 10 to 170 degrees. The reward value was set to be equal to

$$R = -22|\vec{e}|^{1.5} + 0.7\lambda.$$

The initial reward expectation for both cues was set to 0.

In the experiment [Gutierrez-Garralda et al., 2013] the participants were aware of the perturbation. Therefore, we reset the error history when we introduced or removed the perturbation (i.e. the critic did not update its output on a trial immediately following the change in conditions).

In this simulation context we used the following parameters for the critic component:

$t_{low} = 1.9, t_{high} = 2.1$, movement endpoint variability estimate $\kappa = 0.005$, the learning rate update parameters speed $_{up} = 2$ and $slow_{down} = 2.8$, and $a_{opt} = 0.8$.

Parkinson's (PD) and Huntington's Disease (HD) conditions were simulated as described in [Kim et al., 2017b]. Specifically, PD condition involved a 90% reduction in the rates of reinforcement learning in BG to mimic degeneration of dopaminergic neurons in SNc. HD condition was simulated as a 90% reduction in output of the indirect pathway striatal medium spiny neuron population.

Shmuelof et al. context [Shmuelof et al., 2012] We have used a single cue C1 associated with reaching movement to the North (90 degrees, see Fig 1B). The initial strength of the connection from the PFC neuron representing this cue to the thalamocortical neuron corresponding to this action was set to 0.1.

During the asymptote phase, to imitate absent visual feedback, we set the critic output to zero. We simulated the error clamp phase by setting the vector error to zero and providing the reward regardless of whether the target spot was reached or not. The pool of possible actions was 100 reaching movements in directions uniformly distributed from 40 to 110 degrees.

The participants have received explicit reward in [Shmuelof et al., 2012], so we modelled the reward in the following way:

$$R = 2\Theta(0.028 - |\vec{e}|) + 0.65 \cdot \lambda$$

where $\Theta(.)$ is Heaviside step function. The endpoint noise standard deviation was 0.008m. The critic component parameters were $t_{low} = 2, t_{high} = 3.5$, the movement variability $\kappa = 0.008$ $a_{opt} = 0.2$, the rate update factors speed $_{up} = 1.3$ and slow $_{\text{down}} = 1.3$.

6

DISCUSSION

6.1 Significance

The following is a brief discussion of the potential applications and overall significance of the present work.

6.1.1 Clinical relevance

Parkinson's disease (PD) is among the most common neurodegenerative disorders with an estimated 10 million sufferers worldwide [Rocca, 2018]. Neurologically-linked motor disorders like Parkinson's disease (PD) are notoriously difficult to accurately diagnose. Moreover, the development of a laboratory (e.g. genetic) test is unlikely, as the onset of PD depends on complex interactions between genetics and the environment [de Lau and Breteler, 2006]. Clinical diagnosis of PD is further complicated by the existence of other motor diseases— e.g. Huntington's disease (HD), ataxia, and Tourette syndrome— which can also cause Parkinsonian motor symptoms (e.g. tremors, spasms, excessive involuntary movement). Differential diagnosis of PD requires that these confounding disorders and other potential causes of Parkinsonian symptoms— e.g. drug abuse, stroke— are ruled out before a PD diagnosis can be confirmed [NICE, 2017]. The lack of a reliable diagnostic procedure presents a major hurdle that prevents early identification and preventive treatment of PD. However, ongoing research into the physiopathology of neurodegenerative motor disorders has yielded some consensus understanding of the underlying neurophysiological mechanisms that contribute to the onset and progression of motor dysfunction. Damage to brain regions primarily associated with motor control— such as the basal ganglia and cerebellum— has been implicated as a key factor in the loss of motor adaptation, a common feature of PD symptomatology.

In particular, the role of dopaminergic reinforcement learning in the basal ganglia has been the focus of numerous PD-centered studies. Administration of levodopa (L-DOPA), a precursor to dopamine, is the most common initial treatment prescribed for PD patients. Furthermore, the use of deep brain stimulation to target damaged nuclei in the basal ganglia is among the most promising new treatments currently being tested in clinical trials [Bronstein et al., 2011]. Moreover, innovations in brain-machine interface technologies— such as the recently developed ”neural lace” system from the biotech startup Neuralink— could soon enable the restoration of sensory and motor function for PD patients [Musk and Neuralink, 2019].

Despite these medical advances, recent evidence suggests that the underlying mechanisms involved in these pathologies could be more nuanced than previously thought. Patient symptomatology can vary on a case-by-case basis, even between patients diagnosed with the same disease. Some of this variation can be explained by the disruption of particular neuronal species within heterogeneous neuronal populations (e.g. striatal cholinergic interneurons). Therefore, it is important to better understand the underlying physiology that could yield clinically relevant outcomes for improved diagnostic and preventive care to reduce the burden that PD causes for patients and the healthcare system.

6.1.2 Computational methods & neurofunctional theory

In this report, a new ”complete” neurofunctional theory was introduced. The fundamental principles behind this theory are motivated by the goal to fully describe the essential mechanisms underlying neurobehavioral phenomena.

This goal is accomplished largely by the conceptualization of ”neurofunction.” Bridging this gap between biophysics and behavior yields insights into the functional dependence of behavioral output on localized heterogeneous neural circuitry. From this localized circuitry can emerge large-scale, inter-network interactions— e.g. the multi-network neural interactions that form the cerebro-cortico-striatal loop, which could play a role in motor adaptation [Caligiore et al., 2016a, Doya, 2000a, Haber, 2016, Houk and Wise, 1995].

Neurocomputational network models can be derived from statistical measurements—

e.g. from behavioral experiments. Prior knowledge of the relevant neural circuitry— e.g. ionic timescales, neurotransmitters, cell types— can also ease the model construction process. When constructing a neurocomputational model, biophysical parameters are chosen to replicate the known properties of the relevant neurophysiology, while simultaneously replicating the relevant statistical observations. Typically this is an iterative tuning process that continues until the model statistically replicates the observed behavior— e.g. motor adaptation with a specific learning rate.

Using this neurocomputational approach, complex hypotheses can be posed as computer simulations that attempt to replicate essential statistical features of the original behavioral experiments. Essentially, this approach seats Bayesian inference problems in a neurofunctional context that can be more easily interpreted, which addresses a frequently cited disadvantage of popular Bayesian techniques like artificial neural networks. A neurocomputational model can quantitatively replicate phenomenological outcomes— e.g. the success or failure to adapt to a visual perturbation— by Bayesian inference of model parameters, which can be formulated with varying degrees of biophysical details— depending on the specific goal of the neurocomputational experiment.

Neurocomputational modeling can be used as a flexible quantitative tool— capable of testing complex neurofunctional hypotheses, especially those that would otherwise be intractable (e.g. mechanistic hypotheses involving the temporal dynamics of multiple overlapping neural circuits with unknown connectivity). If implemented appropriately, interpretation of simulation results can be accomplished by employing straightforward empirical logic. Initially, it can be assumed that the biophysical model adequately characterizes the essential neurofunctional features of the learning system that produced the empirical data. Then, given that the biophysical parameter **X** is manipulated, the simulation should produce **Y**— this kind of causal relationship can be statistically inferred from the data, e.g. by correlation between two variables. Thus, if the expected correlation cannot be replicated, the null hypothesis is confirmed— the neurofunctional system cannot be characterized by the proposed model. Otherwise, assuming sufficient statistical power, the biophysical relationship defined by the model parameters could explain the essential neurofunctional dynamics of the system.

6.2 Results Summarized

6.2.1 The functional role of striatal cholinergic interneurons in reinforcement learning from computational perspective

chapter 4 describes an investigation into the neurofunctional role played by striatal cholinergic interneurons (”tonically active neurons,” TANs) during reinforcement-mediated motor adaptation. The experimental results presented illustrate that TANs can affect the timing and encoding of reward during RL-mediated adaptation by selectively modulating dopamine release from D2 neurons.

Moreover, the presented neurocomputational simulation results suggest that the inability of L-DOPA to completely restore non-error-based learning in Parkinson’s patients could be explained by the specific mechanism by which L-DOPA increases extracellular dopamine in the striatum. Essentially, L-DOPA restores the tonic dopamine concentration to healthy levels, which restores some of the D2-TAN encoding of reward. However because the phasic release of dopamine is unaffected by L-DOPA, D2-TAN encoding is not fully recovered.

6.2.2 The interplay between cerebellum and basal ganglia in motor adaptation: a modeling study

A key takeaway from this publication is that the cerebellum and basal ganglia can operate via neurofunctionally distinct learning strategies that cannot be concurrently active, and therefore must be coordinated via some latent neurofunctional mechanism.

Using model-based (i.e. supervised) learning, the cerebellum can leverage real-time visual feedback to adaptively correct motor control— in the model, this is represented by an architecture similar to a supervised artificial neural network. In contrast, the basal ganglia can learn cue-action associations via dopamine-based RL, which can be characterized as non-error-based (i.e. model-free) learning— striatal dopamine encodes discrete rewards/punishments without depending on real-time visual feedback.

These distinct learning strategies have different advantages and disadvantages. The cerebellar error-based learning strategy can quickly adapt to small visual perturbations but fails to adapt to larger perturbations. Meanwhile the basal ganglia's non-error-based learning strategy is more robust to large perturbations but adapts to small perturbations less quickly than the cerebellum.

A key conclusion of this study is that a neurofunctional cortico-cerebro-striatal control circuit could be responsible for coordinating between these two distinct learning strategies— a result that is validated by the computational model. This conclusion could be clinically relevant for the diagnosis of Parkinsonian symptoms, particularly in quantitatively distinguishing between different neurofunctional motor disorders.

6.3 Robert Capps: Contributions

In this section, the author's contributions are listed for the relevant experiments described in chapter 4 and chapter 5.

6.3.1 The functional role of striatal cholinergic interneurons in reinforcement learning from computational perspective

For the publication *The functional role of striatal cholinergic interneurons...*, the author's contributions include:

Validation of the computational model— Validated the model by iteratively researching the individual biophysical processes relevant to TAN-dopamine interactions. Compiled relevant quantitative details from the literature to formulate the computational model of TAN-dopamine interactions.

Formal analysis and software development— Implemented the TAN-dopamine release model in the Python programming language. Explored the model's sensitivity to various perturbations and initial conditions using a variety of optimization techniques and visualization methods.

Investigation— Reviewed the literature, explored quantitative behavior of the computational model.

Data curation— Amalgamated data from tables and figures in the literature, which was used for statistical fitting procedures during model development.

Writing (original draft preparation, review and editing).

Visualization— Programmatically created the time series figures (using the Python visualization library Matplotlib), and digitally created the mechanistic diagram of TAN-dopamine interactions (Figure 1) using the open-source vector editing software Inkscape.

6.3.2 The interplay between cerebellum and basal ganglia in motor adaptation: a modeling study

For the publication *The interplay between cerebellum and basal ganglia...*, the author's contributions include:

Conceptualization— Discussed relevant literature, drafted visual diagrams (e.g. on the whiteboard) during lab meetings to explore new avenues of thought relevant to the project.

Methodology— Contributed existing knowledge of computational learning models. Assisted with technical troubleshooting during model development, implementation (e.g. optimization, numerical integration).

Writing (original draft preparation, review and editing).

Visualization— Assisted with making the figures publication-ready (in particular, the figures that show simulation results).

6.4 Future Directions

Finally, this section describes some interesting potential future directions related to the present work. In particular, some possible next steps in the development of a ”complete” neurofunctional theory are discussed. Potential applications of neurofunctional theory are also listed and described briefly.

6.4.1 Conservation of neurofunction

The continued development of neurofunctional theory is an obvious next step in this avenue of research. Specifically a ”complete” neurofunctional theory as defined in chapter 1 requires a more rigorous mathematical formalization of the computational principles underlying neurofunctional theory.

One possible approach involves the formulation of a conservation law of neurofunction. The proposed conservation law could state that the neurofunctional characteristics of an observed behavior (or behaviors) are symmetric in spacetime— i.e. neurofunctional characteristics are conserved (invariant) with respect to spacetime, which implies that observed neurofunction is also preserved between individual computational agents, e.g. human participants in a behavioral study.

Future research on this topic would aim to formalize this law, which could result in the derivation of a physical unit of measure— e.g. by relating the estimated neurofunctional tensor $\hat{\mathbf{N}}$ to the Lagrangian $\mathbf{L}[\mathbf{q}(\mathbf{t}), \dot{\mathbf{q}(\mathbf{t})}, t]$, a functional of the generalized coordinates $\mathbf{q}(t)$ and the times t_1, t_2 is obtained, which can be expressed as $\hat{\mathbf{N}}[\mathbf{q}, t_1, t_2] = \int_{t1}^{t2} \mathbf{L}[\mathbf{q}(t), \dot{\mathbf{q}}(t), t]$, an application of the principle of least action [Feynman, 1965].

6.4.2 Quantitative ethics

The future application of neurofunctional principles— particularly in concert with neurocomputational methods— necessitates serious ethical consideration [Char et al., 2018, Verghese et al., 2018]. The integration of psychology, neuroscience, and neurocomputational methods could provide deeper quantitative insights in fields that straddle the arts and sciences, such as law, philosophy, economics, and sociology. Although qualitative or subjective reasoning is commonly applied in these contexts, neurocomputational techniques could be used to quantitatively validate previously intractable problems in these fields.

6.5 Points Summarized

Section	**Description**
(section 1.1)	A new Neurofunctional theory of learning is introduced, potential applications of the theory are discussed.
(chapter 2)	A synthetic literature review is presented as a chronological narrative. The review summarizes some key discoveries in the cognitive sciences— namely those that motivate the need for a ”complete” Neurofunctional theory— e.g. behaviorism, the cognitive revolution, and the inception of computational neuroscience are discussed.
(chapter 4, chapter 5)	Two of the author’s publications are presented in the context of Neurofunctional theory.
(chapter 6)	Results from the included publications are presented in the context of Neurofunctional theory, and specific applications of these conclusions are discussed.

REFERENCES

[Antony and Roemer, 2003] Antony, M. and Roemer, L. (2003). Behavior therapy. In *Essential psychotherapies: Theory and practice*, pages 182–223. Guilford Press, 2nd edition.

[Aosaki et al., 2010] Aosaki, T., Miura, M., Suzuki, T., Nishimura, K., and Masuda, M. (2010). Acetylcholine-dopamine balance hypothesis in the striatum: An update. *Geriatrics & Gerontology International*, 10:S148–S157.

[Aosaki et al., 1994a] Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A., and Kimura, M. (1994a). Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *The Journal of Neuroscience*, 14(6):3969–3984.

[Aosaki et al., 1994b] Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A., and Kimura, M. (1994b). Responses of tonically active neurons in the primate’s striatum undergo systematic changes during behavioral sensorimotor conditioning. *The Journal of Neuroscience*, 14(6):3969–3984.

[Apicella et al., 2011] Apicella, P., Ravel, S., Deffains, M., and Legallet, E. (2011). The role of striatal tonically active neurons in reward prediction error signaling during instrumental task performance. *Journal of Neuroscience*, 31(4):1507–1515.

[Ashby and Crossley, 2011] Ashby, F. G. and Crossley, M. J. (2011). A computational model of how cholinergic interneurons protect striatal-dependent learning. *Journal of Cognitive Neuroscience*, 23(6):1549–1566.

[Barto, 1995] Barto, A. (1995). *Adaptive Critics and the Basal Ganglia.*

[Barto et al., 1982] Barto, A., Anderson, C., and Sutton, R. (1982). Synthesis of nonlinear control surfaces by a layered associative search network. *Biological Cybernetics*, 43(3):175–185.

[Barto and Sutton, 1981] Barto, A. and Sutton, R. (1981). Landmark learning: An illustration of associative search. *Biological Cybernetics*, 42(1):1–8.

[Barto et al., 1981] Barto, A., Sutton, R., and Brouwer, P. (1981). Associative search network: A reinforcement learning associative memory. *Biological Cybernetics*, 40(3):201–211.

[Bastian, 2006] Bastian, A. (2006). Learning to predict the future: The cerebellum adapts feedforward movement control. *Current Opinion in Neurobiology*, 16(6):645–649.

[Beigi et al., 2016] Beigi, M., Wilkinson, L., Gobet, F., Parton, A., and Jahanshahi, M. (2016). Levodopa medication improves incidental sequence learning in parkinson's disease. *Neuropsychologia*, 93:53–60.

[Bennett et al., 2000] Bennett, B. D., Callaway, J. C., and Wilson, C. J. (2000). Intrinsic membrane properties underlying spontaneous tonic firing in neostriatal cholinergic interneurons. *The Journal of Neuroscience*, 20(22):8493–8503.

[Bowersox and Closs, 1989] Bowersox, D. and Closs, D. (1989). Simulation In Logistics: A Review Of Present Practice And A. *Journal of Business Logistics*, 10(1):133. Publisher: ProQuest Central.

[Bronstein et al., 2011] Bronstein, J. M., Tagliati, M., Alterman, R. L., Lozano, A. M., Volkmann, J., Stefani, A., Horak, F. B., Okun, M. S., Foote, K. D., Krack, P., Pahwa, R., Henderson, J. M., Hariz, M. I., Bakay, R. A., Rezai, A., Marks, W. J., Moro, E., Vitek, J. L., Weaver, F. M., Gross, R. E., and DeLong, M. R. (2011). Deep Brain Stimulation for Parkinson Disease. *Archives of neurology*, 68(2):165–171.

[Brooks, 2008] Brooks, D. (2008). Optimizing levodopa therapy for parkinsons disease with levodopa/carbidopa/entacapone: implications from a clinical and patient perspective. *Neuropsychiatric Disease and Treatment*, page 39.

[Brown et al., 2012] Brown, M. T. C., Tan, K. R., O'Connor, E. C., Nikonenko, I., Muller, D., and Lüscher, C. (2012). Ventral tegmental area GABA projections pause accumbal cholinergic interneurons to enhance associative learning. *Nature*, 492(7429):452–456.

[Butcher et al., 2017] Butcher, P. A., Ivry, R. B., Kuo, S.-H., Rydz, D., Krakauer, J. W., and Taylor, J. A. (2017). The cerebellum does more than sensory prediction error-based learning in sensorimotor adaptation tasks. *Journal of Neurophysiology*, 118(3):1622–1636.

[Cachope et al., 2012] Cachope, R., Mateo, Y., Mathur, B. N., Irving, J., Wang, H.-L., Morales, M., Lovinger, D. M., and Cheer, J. F. (2012). Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: Setting the tone for reward processing. *Cell Reports*, 2(1):33–41.

[Calabresi et al., 2000] Calabresi, P., Centonze, D., Gubellini, P., Pisani, A., and Bernardi, G. (2000). Acetylcholine-mediated modulation of striatal function. *Trends in Neurosciences*, 23(3):120–126.

[Caligiore et al., 2016a] Caligiore, D., Pezzulo, G., Baldassarre, G., Bostan, A., Strick, P., Doya, K., Helmich, R., Dirkx, M., Houk, J., Jörntell, H., Lago, A., Galea, J., Miall, R., Popa, T., Kishore, A., Verschure, P., Zucca, R., and Herreros, I. (2016a). Consensus Paper: Towards a Systems-Level View of Cerebellar Function: The Interplay Between Cerebellum, Basal Ganglia, and Cortex. *The Cerebellum*, 16.

[Caligiore et al., 2016b] Caligiore, D., Pezzulo, G., Baldassarre, G., Bostan, A. C., Strick, P. L., Doya, K., Helmich, R. C., Dirkx, M., Houk, J., Jörntell, H., Lago-Rodriguez, A., Galea, J. M., Miall, R. C., Popa, T., Kishore, A., Verschure, P. F. M. J., Zucca, R., and Herreros, I. (2016b). Consensus paper: Towards a systems-level view of cerebellar function: the interplay between cerebellum, basal ganglia, and cortex. *The Cerebellum*, 16(1):203–229.

[Castro et al., 2014] Castro, L. N. G., Hadjiosif, A. M., Hemphill, M. A., and Smith, M. A. (2014). Environmental consistency determines the rate of motor adaptation. *Current Biology*, 24(10):1050–1061.

[Centonze et al., 2003] Centonze, D., Gubellini, P., Pisani, A., Bernardi, G., and Calabresi, P. (2003). Dopamine, acetylcholine and nitric oxide systems interact to induce corticostriatal synaptic plasticity. *Reviews in the Neurosciences*, 14(3).

[Char et al., 2018] Char, D., Shah, N., and Magnus, D. (2018). Implementing machine learning in health care—Addressing ethical challenges. *The New England Journal of Medicine*, 378(11):981.

[Chen et al., 2016] Chen, J., Ho, S.-L., Lee, T. M.-C., Chang, R. S.-K., Pang, S. Y.-Y., and Li, L. (2016). Visuomotor control in patients with parkinson's disease. *Neuropsychologia*, 80:102–114.

[Chomsky, 1959] Chomsky, N. (1959). Review of verbal behavior by B. In *F. Skinner. The History of Psychology: Fundamental Questions*, pages 408–429.

[Corsini et al., 2008] Corsini, R., Wedding, D., and Dumont, F. (2008). *Current psychotherapies.* Thomson Brooks/Cole; /z-wcorg/.

[Cragg, 2006] Cragg, S. J. (2006). Meaningful silences: how dopamine listens to the ACh pause. *Trends in Neurosciences*, 29(3):125–131.

[Dautan et al., 2014] Dautan, D., Huerta-Ocampo, I., Witten, I. B., Deisseroth, K., Bolam, J. P., Gerdjikov, T., and Mena-Segovia, J. (2014). A major external source of cholinergic innervation of the striatum and nucleus accumbens originates in the brainstem. *Journal of Neuroscience*, 34(13):4509–4518.

[de Lau and Breteler, 2006] de Lau, L. M. and Breteler, M. M. (2006). Epidemiology of Parkinson’s disease. *The Lancet Neurology*, 5(6):525–535.

[Deng et al., 2007] Deng, P., Zhang, Y., and Xu, Z. C. (2007). Involvement of ih in dopamine modulation of tonic firing in striatal cholinergic interneurons. *Journal of Neuroscience*, 27(12):3148–3156.

[Ding et al., 2010] Ding, J. B., Guzman, J. N., Peterson, J. D., Goldberg, J. A., and Surmeier, D. J. (2010). Thalamic gating of corticostriatal signaling by cholinergic interneurons. *Neuron*, 67(2):294–307.

[Doig et al., 2014] Doig, N. M., Magill, P. J., Apicella, P., Bolam, J. P., and Sharott, A. (2014). Cortical and thalamic excitation mediate the multiphasic responses of striatal cholinergic interneurons to motivationally salient stimuli. *Journal of Neuroscience*, 34(8):3101–3117.

[Donchin et al., 2012] Donchin, O., Rabe, K., Diedrichsen, J., Lally, N., Schoch, B., Gizewski, E. R., and Timmann, D. (2012). Cerebellar regions involved in adaptation to force field and visuomotor perturbation. *Journal of Neurophysiology*, 107(1):134–147.

[Doya, 1996] Doya, K. (1996). *Temporal Difference Learning in Continuous Time and Space.*

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

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

[Doya, 2002] Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, 15(4–6):495–506.

[Doyon et al., 2003] Doyon, J., Penhune, V., and Ungerleider, L. (2003). Distinct contribution of the cortico-striatal and cortico-cerebellar systems to motor skill learning. *Neuropsychologia*, 41(3):252–262.

[Dreyfus, 1990] Dreyfus, S. E. (1990). Artificial neural networks, back propagation, and the kelley-bryson gradient procedure. *Journal of Guidance, Control, and Dynamics*, 13(5):926–928.

[Feynman, 1965] Feynman, R. P. (1965). *The character of physical law.* M.I.T. Press, Cambridge, Mass.

[Fodor, 1983] Fodor, J. (1983). *The Modularity of Mind.* MIT Press.

[Frank et al., 2007] Frank, M., Samanta, J., Moustafa, A., and Sherman, S. (2007). Hold Your Horses: Impulsivity, Deep Brain Stimulation, and Medication in Parkinsonism. *Science*, 318(5854):1309–1312.

[Frank, 2005] Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated parkinsonism. *Journal of Cognitive Neuroscience*, 17(1):51–72.

[Frank, 2006] Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, 19(8):1120–1136.

[Franklin and Wolpert, 2008] Franklin, D. W. and Wolpert, D. M. (2008). Specificity of reflex adaptation for task-relevant variability. *Journal of Neuroscience*, 28(52):14165–14175.

[Franklin and Frank, 2015a] Franklin, N. and Frank, M. (2015a). A cholinergic feedback circuit to regulate striatal population uncertainty and optimize reinforcement learning. *ELife*, 4.

[Franklin and Frank, 2015b] Franklin, N. T. and Frank, M. J. (2015b). A cholinergic feedback circuit to regulate striatal population uncertainty and optimize reinforcement learning. *eLife*, 4.

[Galarraga et al., 1999] Galarraga, E., Hernández-López, S., Reyes, A., Miranda, I., Bermudez-Rattoni, F., Vilchis, C., and Bargas, J. (1999). Cholinergic modulation of neostriatal output: A functional antagonism between different types of muscarinic receptors. *The Journal of Neuroscience*, 19(9):3629–3638.

[Galea et al., 2011] Galea, J., Vazquez, A., Pasricha, N., Xivry, J.-J., and Celnik, P. (2011). Dissociating the Roles of the Cerebellum and Motor Cortex during Adaptive Learning:

The Motor Cortex Retains What the Cerebellum Learns. *Cerebral Cortex*, 21(8):1761–1770.

[Galea et al., 2010] Galea, J. M., Vazquez, A., Pasricha, N., de Xivry, J.-J. O., and Celnik, P. (2010). Dissociating the roles of the cerebellum and motor cortex during adaptive learning: The motor cortex retains what the cerebellum learns. *Cerebral Cortex*, 21(8):1761–1770.

[Gao et al., 1996] Gao, J.-H., Parsons, L., Bower, J., Xiong, J., Li, J., and Fox, P. (1996). Cerebellum implicated in sensory acquisition and discrimination rather than motor control. *Science*, 272(5261):545–547.

[Gellman et al., 1985] Gellman, R., Gibson, A. R., and Houk, J. C. (1985). Inferior olivary neurons in the awake cat: detection of contact and passive body displacement. *Journal of Neurophysiology*, 54(1):40–60.

[Gellman et al., 1983] Gellman, R., Houk, J. C., and Gibson, A. R. (1983). Somatosensory properties of the inferior olive of the cat. *The Journal of Comparative Neurology*, 215(2):228–243.

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

[Graybiel, 2008] Graybiel, A. M. (2008). Habits, rituals, and the evaluative brain. *Annual Review of Neuroscience*, 31(1):359–387.

[Gutierrez-Garralda et al., 2013] Gutierrez-Garralda, J. M., Moreno-Briseño, P., Boll, M.-C., Morgado-Valle, C., Campos-Romo, A., Diaz, R., and Fernandez-Ruiz, J. (2013). The effect of parkinson's disease and huntington's disease on human visuomotor learning. *European Journal of Neuroscience*.

[Haber, 2016] Haber, S. (2016). Corticostriatal circuitry. *Dialogues in Clinical Neuroscience*, 18(1):7–21.

[Harischandra and Örjan Ekeberg, 2008] Harischandra, N. and Örjan Ekeberg (2008). System identification of muscle–joint interactions of the cat hind limb during locomotion. *Biological Cybernetics*, 99(2):125–138.

[Hayes, 1991] Hayes, S. (1991). A relational control theory of stimulus equivalence. In *Dialogues on verbal behavior: The First International Institute on Verbal Relations*, pages 19–40. Context Press.

[Heffley et al., 2018] Heffley, W., Song, E. Y., Xu, Z., Taylor, B. N., Hughes, M. A., McKinney, A., Joshua, M., and Hull, C. (2018). Coordinated cerebellar climbing fiber activity signals learned sensorimotor predictions. *Nature Neuroscience*, 21(10):1431–1441.

[Hikosaka et al., 2002] Hikosaka, O., Nakamura, K., Sakai, K., and Nakahara, H. (2002). Central mechanisms of motor skill learning. *Current Opinion in Neurobiology*, 12(2):217–222.

[Hodgkin and Huxley, 1952] Hodgkin, A. and Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544.

[Hollerman and Schultz, 1998] Hollerman, J. R. and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1(4):304–309.

[Houk and Barto, 1992] Houk, J. and Barto, A. (1992). Distributed sensorimotor learning. *Advances in Psychology*, 87:71–100.

[Houk et al., 1996] Houk, J., Buckingham, J., and Barto, A. (1996). Models of the cerebellum and motor learning. *Behavioral and Brain Sciences*, 19(3):368–383.

[Houk and Wise, 1995] Houk, J. and Wise, S. (1995). Feature Article: Distributed Modular Architectures Linking Basal Ganglia, Cerebellum, and Cerebral Cortex: Their Role in Planning and Controlling Action. *Cerebral Cortex*, 5(2):95–110.

[Huberdeau et al., 2015] Huberdeau, D. M., Krakauer, J. W., and Haith, A. M. (2015). Dual-process decomposition in human sensorimotor adaptation. *Current Opinion in Neurobiology*, 33:71–77.

[Hyland et al., 2002] Hyland, B., Reynolds, J., Hay, J., Perk, C., and Miller, R. (2002). Firing modes of midbrain dopamine cells in the freely moving rat. *Neuroscience*, 114(2):475–492.

[Hyland and Clayton, 1992] Hyland, K. and Clayton, P. T. (1992). Aromatic l-amino acid decarboxylase deficiency: Diagnostic methodology. *Clinical Chemistry*, 38(12):2405–2410.

[Ivry et al., 2002] Ivry, R., Spencer, R., Zelaznik, H., and Diedrichsen, J. (2002). The cerebellum and event timing. *Annals of the New York Academy of Sciences*, 978(1):302–317.

[Izawa et al., 2012] Izawa, J., Criscimagna-Hemminger, S. E., and Shadmehr, R. (2012). Cerebellar contributions to reach adaptation and learning sensory consequences of action. *Journal of Neuroscience*, 32(12):4230–4239.

[Izawa et al., 2008] Izawa, J., Rane, T., Donchin, O., and Shadmehr, R. (2008). Motor adaptation as a process of reoptimization. *Journal of Neuroscience*, 28(11):2883–2891.

[Izawa and Shadmehr, 2011] Izawa, J. and Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLoS Computational Biology*, 7(3):e1002012.

[Joshua et al., 2008] Joshua, M., Adler, A., Mitelman, R., Vaadia, E., and Bergman, H. (2008). Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *Journal of Neuroscience*, 28(45):11673–11684.

[Kalia and Lang, 2015] Kalia, L. V. and Lang, A. E. (2015). Parkinson's disease. *The Lancet*, 386(9996):896–912.

[Kasher, 1998] Kasher, A. (1998). *Pragmatics: Communication, interaction, and discourse.* SUNY Press.

[Kheradmand and Zee, 2011] Kheradmand, A. and Zee, D. (2011). Cerebellum and ocular motor control. *Frontiers in Neurology*, 2:53.

[Kim et al., 2019] Kim, T., Capps, R. A., Hamade, K. C., Barnett, W. H., Todorov, D. I., Latash, E. M., Markin, S. N., Rybak, I. A., and Molkov, Y. I. (2019). The functional role of striatal cholinergic interneurons in reinforcement learning from computational perspective. *Frontiers in Neural Circuits*, 13.

[Kim et al., 2017a] Kim, T., Hamade, K., Todorov, D., Barnett, W., Capps, R., Latash, E., Markin, S., Rybak, I., and Molkov, Y. (2017a). Reward Based Motor Adaptation Mediated by Basal Ganglia. *Frontiers in Computational Neuroscience*, 11.

[Kim et al., 2017b] Kim, T., Hamade, K. C., Todorov, D., Barnett, W. H., Capps, R. A., Latash, E. M., Markin, S. N., Rybak, I. A., and Molkov, Y. I. (2017b). Reward based motor adaptation mediated by basal ganglia. *Frontiers in Computational Neuroscience*, 11.

[Kita, 1993] Kita, H. (1993). Chapter 4 GABAergic circuits of the striatum. In *Progress in Brain Research*, pages 51–72. Elsevier.

[Klopf, 1972] Klopf, A. (1972). *Brain function and adaptive systems: A heterostatic theory.*

[Klopf, 1979] Klopf, A. (1979). Goal-seeking systems from goal-seeking components: Implications for AI. *The Cognition and Brain Theory Newsletter*, 3(2).

[Klopf, 1982] Klopf, A. (1982). *The hedonistic neuron: A theory of memory, learning, and intelligence.* Toxicology-Sci.

[Knowles, 1986] Knowles, W. S. (1986). Application of organometallic catalysis to the commercial production of l-DOPA. *Journal of Chemical Education*, 63(3):222.

[Knuth, 2002] Knuth, D. (2002). *Selected papers on computer languages*. CSLI Publications.

[Koós and Tepper, 1999] Koós, T. and Tepper, J. M. (1999). Inhibitory control of neostriatal projection neurons by GABAergic interneurons. *Nature Neuroscience*, 2(5):467–472.

[Kosillo et al., 2016] Kosillo, P., Zhang, Y.-F., Threlfell, S., and Cragg, S. J. (2016). Cortical control of striatal dopamine transmission via striatal cholinergic interneurons. *Cerebral Cortex*, 26(11):4160–4169.

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

[Lovie and S, 1996] Lovie, P. and S, F. (1996). Charles Edward Spearman. *Notes and Records of the Royal Society of London*, 50(1):75–88.

[Luque et al., 2014] Luque, N. R., Garrido, J. A., Carrillo, R. R., D'Angelo, E., and Ros, E. (2014). Fast convergence of learning requires plasticity between inferior olive and deep cerebellar nuclei in a manipulation task: a closed-loop robotic simulation. *Frontiers in Computational Neuroscience*, 8.

[Markin et al., 2015] Markin, S. N., Klishko, A. N., Shevtsova, N. A., Lemay, M. A., Prilutsky, B. I., and Rybak, I. A. (2015). A neuromechanical model of spinal control of locomotion. In *Neuromechanical Modeling of Posture and Locomotion*, pages 21–65. Springer New York.

[Maurice, 2004] Maurice, N. (2004). D2 dopamine receptor-mediated modulation of voltage-dependent na channels reduces autonomous activity in striatal cholinergic interneurons. *Journal of Neuroscience*, 24(46):10289–10301.

[McCorduck, 2004] McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. A.K. Peters; /z-wcorg/.

[McCulloch and Pitts, 1943] McCulloch, W. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.

[Miall et al., 1993] Miall, R., Weir, D., Wolpert, D., and Stein, J. (1993). Is the Cerebellum a Smith Predictor? *Journal of Motor Behavior*, 25(3):203–216.

[Miller, 2003] Miller, G. (2003). The cognitive revolution: A historical perspective. *Trends in Cognitive Sciences*, 7(3):141–144.

[Mirenowicz and Schultz, 1996] Mirenowicz, J. and Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, 379(6564):449–451.

[Molkov et al., 2009] Molkov, Y. I., Mukhin, D. N., Loskutov, E. M., Feigin, A. M., and Fidelin, G. A. (2009). Using the minimum description length principle for global reconstruction of dynamic systems from noisy time series. *Physical Review E*, 80(4):046207.

[Montague et al., 1996] Montague, P., Dayan, P., and Sejnowski, T. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience*, 16(5):1936–1947.

[Morris et al., 2004] Morris, G., Arkadir, D., Nevet, A., Vaadia, E., and Bergman, H. (2004). Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron*, 43(1):133–143.

[Musk and Neuralink, 2019] Musk, E. and Neuralink (2019). An integrated brain-machine interface platform with thousands of channels. *bioRxiv*, page 703801.

[Newell et al., 1959] Newell, A., Shaw, J., and Simon, H. (1959). Report on a General Problem Solving Method. *Report on a General Problem Solving Method*, 29.

[NICE, 2017] NICE (2017). Parkinson's disease in adults.

[Oswald et al., 2009] Oswald, M. J., Oorschot, D. E., Schulz, J. M., Lipski, J., and Reynolds, J. N. J. (2009). IHcurrent generates the afterhyperpolarisation following activation of subthreshold cortical synaptic inputs to striatal cholinergic interneurons. *The Journal of Physiology*, 587(24):5879–5897.

[O'Reilly and Frank, 2006] O'Reilly, R. and Frank, M. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation*, 18(2):283–328.

[Paulin, 1993] Paulin, M. (1993). The role of the cerebellum in motor control and perception. *Brain, Behavior and Evolution*, 41(1):39–50.

[Pavlov, 2010] Pavlov, P. (2010). Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex. *Annals of Neurosciences*, 17(3):136.

[Penhune and Steele, 2012] Penhune, V. and Steele, C. (2012). Parallel contributions of cerebellar, striatal and M1 mechanisms to motor sequence learning. *Behavioural Brain Research*, 226(2):579–591.

[Peter Grünwald, 1998] Peter Grünwald (1998). The Minimum Description Length Principle and Reasoning under Uncertainty.

[Piaget et al., 1980] Piaget, J., Chomsky, N., and Piattelli-Palmarini, M. (1980). *Language and learning: The debate between Jean Piaget and Noam Chomsky*. /z-wcorg/.

[Piccinini and Craver, 2011] Piccinini, G. and Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3):283–311.

[Pisani, 2003] Pisani, A. (2003). Targeting striatal cholinergic interneurons in parkinson's disease: Focus on metabotropic glutamate receptors. *Neuropharmacology*, 45(1):45–56.

[Prochazka, 1999] Prochazka, A. (1999). Chapter 11 quantifying proprioception. In *Progress in Brain Research*, pages 133–142. Elsevier.

[Reynolds, 2004] Reynolds, J. N. J. (2004). Modulation of an afterhyperpolarization by the substantia nigra induces pauses in the tonic firing of striatal cholinergic interneurons. *Journal of Neuroscience*, 24(44):9870–9877.

[Rice and Cragg, 2004] Rice, M. E. and Cragg, S. J. (2004). Nicotine amplifies reward-related dopamine signals in striatum. *Nature Neuroscience*, 7(6):583–584.

[Rissanen, 1978] Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5):465–471.

[Rocca, 2018] Rocca, W. A. (2018). The burden of Parkinson’s disease: a worldwide perspective. *The Lancet Neurology*, 17(11):928–929.

[Ruggles and Brodie, 1947] Ruggles, R. and Brodie, H. (1947). An Empirical Approach to Economic Intelligence in World War II. *Journal of the American Statistical Association*, 42(237):72–91.

[Scherman et al., 1989] Scherman, D., Desnos, C., Darchen, F., Pollak, P., Javoy-Agid, F., and Agid, Y. (1989). Striatal dopamine deficiency in parkinson's disease: Role of aging. *Annals of Neurology*, 26(4):551–557.

[Schiff et al., 1996] Schiff, S., So, P., Chang, T., Burke, R., and Sauer, T. (1996). Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Physical Review E*, 54(6):6708–6724.

[Schlerf et al., 2013] Schlerf, J. E., Xu, J., Klemfuss, N. M., Griffiths, T. L., and Ivry, R. B. (2013). Individuals with cerebellar degeneration show similar adaptation deficits with large and small visuomotor errors. *Journal of Neurophysiology*, 109(4):1164–1173.

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

[Schultz, 1999] Schultz, W. (1999). The reward signal of midbrain dopamine neurons. *Physiology*, 14(6):249–255.

[Schultz, 2015] Schultz, W. (2015). Neuronal Reward and Decision Signals: From Theories to Data. *Physiological Reviews*, 95(3):853–951.

[Schultz, 2016] Schultz, W. (2016). Reward functions of the basal ganglia. *Journal of Neural Transmission*, 123(7):679–693.

[Schultz et al., 1997] Schultz, W., Dayan, P., and Montague, P. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306):1593–1599.

[Schulz et al., 2011] Schulz, J. M., Oswald, M. J., and Reynolds, J. N. J. (2011). Visual-induced excitation leads to firing pauses in striatal cholinergic interneurons. *Journal of Neuroscience*, 31(31):11133–11143.

[Schulz and Reynolds, 2013] Schulz, J. M. and Reynolds, J. N. (2013). Pause and rebound: sensory control of cholinergic signaling in the striatum. *Trends in Neurosciences*, 36(1):41–50.

[Shin et al., 2017] Shin, J. H., Adrover, M. F., and Alvarez, V. A. (2017). Distinctive modulation of dopamine release in the nucleus accumbens shell mediated by dopamine and acetylcholine receptors. *The Journal of Neuroscience*, 37(46):11166–11180.

[Shin et al., 2015] Shin, J. H., Adrover, M. F., Wess, J., and Alvarez, V. A. (2015). Muscarinic regulation of dopamine and glutamate transmission in the nucleus accumbens. *Proceedings of the National Academy of Sciences*, 112(26):8124–8129.

[Shmuelof et al., 2012] Shmuelof, L., Huang, V. S., Haith, A. M., Delnicki, R. J., Mazzoni, P., and Krakauer, J. W. (2012). Overcoming motor ”forgetting” through reinforcement of learned actions. *Journal of Neuroscience*, 32(42):14617–14621a.

[Skinner, 1958] Skinner, B. (1958). Reinforcement today. *American Psychologist*, 13(3):94–99.

[Skinner, 1969] Skinner, B. (1969). Contingencies of Reinforcement. *Contingencies of Reinforcement*, 311.

[Smith et al., 1998] Smith, Y., Bevan, M. D., Shink, E., and Bolam, J. P. (1998). Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience*, 86(2):353–387. Place: United States.

[Straub et al., 2014] Straub, C., Tritsch, N. X., Hagan, N. A., Gu, C., and Sabatini, B. L. (2014). Multiphasic modulation of cholinergic interneurons by nigrostriatal afferents. *Journal of Neuroscience*, 34(25):8557–8569.

[Sulzer et al., 2016] Sulzer, D., Cragg, S. J., and Rice, M. E. (2016). Striatal dopamine neurotransmission: Regulation of release and uptake. *Basal Ganglia*, 6(3):123–148.

[Suri and Schultz, 2001] Suri, R. and Schultz, W. (2001). Temporal Difference Model Reproduces Anticipatory Neural Activity. *Neural Computation*, 13(4):841–862.

[Sutton, 1984] Sutton, R. S. (1984). *Temporal Credit Assignment in Reinforcement Learning.* [PhD Thesis]., University of Massachusetts Amherst.

[Sutton, 1988] Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, 3(1):9–44.

[Sutton and Barto, 1981] Sutton, R. S. and Barto, A. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, 88(2):135.

[Tan and Bullock, 2008] Tan, C. O. and Bullock, D. (2008). A dopamine–acetylcholine cascade: Simulating learned and lesion-induced behavior of striatal cholinergic interneurons. *Journal of Neurophysiology*, 100(4):2409–2421.

[Teka et al., 2017a] Teka, W. W., Hamade, K. C., Barnett, W. H., Kim, T., Markin, S. N., Rybak, I. A., and Molkov, Y. I. (2017a). From the motor cortex to the movement and back again. *PLOS ONE*, 12(6):e0179288.

[Teka et al., 2017b] Teka, W. W., Hamade, K. C., Barnett, W. H., Kim, T., Markin, S. N., Rybak, I. A., and Molkov, Y. I. (2017b). From the motor cortex to the movement and back again. *PLOS ONE*, 12(6):e0179288.

[Telgen et al., 2014] Telgen, S., Parvin, D., and Diedrichsen, J. (2014). Mirror reversal and visual rotation are learned and consolidated via separate mechanisms: Recalibrating or LearningDe novo? *The Journal of Neuroscience*, 34(41):13768–13779.

[Tepper et al., 2018] Tepper, J. M., Koós, T., Ibanez-Sandoval, O., Tecuapetla, F., Faust, T. W., and Assous, M. (2018). Heterogeneity and diversity of striatal GABAergic interneurons: Update 2018. *Frontiers in Neuroanatomy*, 12.

[Tepper et al., 2010] Tepper, J. M., Tecuapetla, F., Koós, T., and Ibáñez-Sandoval, O. (2010). Heterogeneity and diversity of striatal GABAergic interneurons. *Frontiers in Neuroanatomy*, 4.

[Thorndike, 1898] Thorndike, E. (1898). *Some Experiments on Animal Intelligence*. Zenodo.

[Thoroughman and Shadmehr, 2000] Thoroughman, K. A. and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature*, 407(6805):742–747.

[Threlfell et al., 2012] Threlfell, S., Lalic, T., Platt, N. J., Jennings, K. A., Deisseroth, K., and Cragg, S. J. (2012). Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron*, 75(1):58–64.

[Todorov et al., 2019] Todorov, D., Capps, R., Barnett, W., Latash, E., Kim, T., Hamade, K., Markin, S., Rybak, I., and Molkov, Y. (2019). The interplay between cerebellum and basal ganglia in motor adaptation: A modeling study. *PLOS ONE*, 14(4):0214926.

[Toni and Passingham, 1999] Toni, I. and Passingham, R. (1999). Prefrontal-basal ganglia pathways are involved in the learning of arbitrary visuomotor associations: A PET study. *Experimental Brain Research*, 127(1):19–32.

[Tsai et al., 2009] Tsai, M., Kohlenberg, R., Kanter, J., Kohlenberg, B., Follette, W., and Callaghan, G. (2009). *A Guide to Functional Analytic Psychotherapy: Awareness, Courage, Love, and Behaviorism*. Springer US.

[Turing, 1951] Turing, A. M. (1951). *The chemical basis of morphogenesis*.

[Vaswani et al., 2015] Vaswani, P. A., Shmuelof, L., Haith, A. M., Delnicki, R. J., Huang, V. S., Mazzoni, P., Shadmehr, R., and Krakauer, J. W. (2015). Persistent residual errors in motor adaptation tasks: Reversion to baseline and exploratory escape. *Journal of Neuroscience*, 35(17):6969–6977.

[Verghese et al., 2018] Verghese, A., Shah, N., and Harrington, R. (2018). What this computer needs is a physician: Humanism and artificial intelligence. *Jama*, 319(1):19–20.

[Wade and Katzman, 1975] Wade, L. A. and Katzman, R. (1975). SYNTHETIC AMINO ACIDS AND THE NATURE OF l-DOPA TRANSPORT AT THE BLOOD-BRAIN BARRIER. *Journal of Neurochemistry*, 25(6):837–842.

[Wagner et al., 2017] Wagner, M. J., Kim, T. H., Savall, J., Schnitzer, M. J., and Luo, L. (2017). Cerebellar granule cells encode the expectation of reward. *Nature*, 544(7648):96–100.

[Wall et al., 2013] Wall, N. R., Parra, M. D. L., Callaway, E. M., and Kreitzer, A. C. (2013). Differential innervation of direct- and indirect-pathway striatal projection neurons. *Neuron*, 79(2):347–360.

[Wang et al., 1987] Wang, J.-J., Kim, J. H., and Ebner, T. J. (1987). Climbing fiber afferent modulation during a visually guided, multi-joint arm movement in the monkey. *Brain Research*, 410(2):323–329.

[Watson and Rayner, 1920] Watson, J. and Rayner, R. (1920). CONDITIONED EMOTIONAL REACTIONS. In *CONDITIONED EMOTIONAL REACTIONS*, page 14.

[Watson, 1913] Watson, J. B. (1913). *Psychology as the Behaviorist Views it.* Classics in the History of Psychology—Watson.

[Wickens, 1980] Wickens, C. (1980). The Structure of Attentional Resources. In *Attention and Performance*, volume VIII, 8, page 239.

[Williams, 1992] Williams, R. (1992). *Simple statistical gradient-following algorithms for connectionist reinforcement learning.*

[Wilson, 2005] Wilson, C. J. (2005). The mechanism of intrinsic amplification of hyperpolarizations and spontaneous bursting in striatal cholinergic interneurons. *Neuron*, 45(4):575–585.

[Xenias et al., 2015] Xenias, H. S., Ibanez-Sandoval, O., Koos, T., and Tepper, J. M. (2015). Are striatal tyrosine hydroxylase interneurons dopaminergic? *Journal of Neuroscience*, 35(16):6584–6599.

[Yager et al., 2015] Yager, L., Garcia, A., Wunsch, A., and Ferguson, S. (2015). The ins and outs of the striatum: Role in drug addiction. *Neuroscience*, 301:529–541.

[Zeeuw, 1998] Zeeuw, C. D. (1998). Microcircuitry and function of the inferior olive. *Trends in Neurosciences*, 21(9):391–400.

[Zhang and Cragg, 2017] Zhang, Y.-F. and Cragg, S. J. (2017). Pauses in striatal cholinergic interneurons: What is revealed by their common themes and variations? *Frontiers in Systems Neuroscience*, 11.

[Zhang et al., 2018] Zhang, Y.-F., Reynolds, J. N., and Cragg, S. J. (2018). Pauses in cholinergic interneuron activity are driven by excitatory input and delayed rectification, with dopamine modulation. *Neuron*, 98(5):918–925.e3.

[Zhou et al., 2002] Zhou, F.-M., Wilson, C. J., and Dani, J. A. (2002). Cholinergic interneuron characteristics and nicotinic properties in the striatum. *Journal of Neurobiology*, 53(4):590–605.

[Zsigmond et al., 2014] Zsigmond, P., Nord, M., Kullman, A., Diczfalusy, E., Wårdell, K., and Dizdar, N. (2014). Neurotransmitter levels in basal ganglia during levodopa and deep brain stimulation treatment in parkinson's disease. *Neurology and Clinical Neuroscience*, 2(5):149–155.