COMPUTATIONAL PSYCHIATRY: A PRIMER

Editor: Peggy Seriès

Contributors:

X.J. Wang, P. Dayan, T. Braver, R. Adams,
M. Browning, D. Redish, T. Maia, M. Ferrante
and others

Preface

I used to work in visual computational neuroscience and gradually moved towards computational psychiatry. This move was partly motivated by my own struggles but ultimately by witnessing the suffering in some of our students, and even the suicide of two of them.

I felt we did not understand them, and what really mattered for all of us. In terms of our research in computational neuroscience, I felt that we were not putting the efforts where they mattered most, and that our research could be immensely more useful in the long term if we did. I became interested in understanding how mental illness is described and how to bridge advances in neurobiology, computational cognitive science and psychiatric disorders. I was attracted as well by the idea that mental health and illness lie on a continuum, that it concerns all of us, to various degrees, as we are all potentially at risk that our suffering can become overwhelming.

I wanted to provide an accessible book for students starting in this new emerging field, coming from a wide variety of backgrounds. I am well aware though that there are no solid truths in this field yet. The field of computational psychiatry is being met with great enthusiasm and hopes for clinical usefulness but is still in its infancy. This book is very imperfect in that it only addresses a subset of questions, and often leads to more questions than answers. Still, I think it shows that computational psychiatry can provide important new insights and help bridge neuroscience and clinical applications.

I am immensely grateful to the brilliant contributors of this book, all international leaders in the field, who trusted me in this process, despite this being my first book and them being all more established than I am.

I have been deeply inspired by my postdoctoral stay at Gatsby Computational Unit and the research led by Peter Dayan with Nathaniel Daw, Yael Niv and Quentin Huys at that time. I'm also very grateful to Eero Simoncelli for teaching me about scientific humility and our role about making our research accessible. More recently I am very grateful for discussions with my collaborators, particularly Stephen Lawrie and Douglas Steele, Renaud Jardri, Sophie Deneve, Jonathan Roiser, Phil Corlett, Paul Fletcher, Andrew McIntosh, Andy Clark and many others.

I would like to thank my very talented students who chose a project in this field, at a time when it was new to all of us: Vincent Valton, James Raymond, Aleks Stolicyn, Aistis Stankevicius, Frank Karvelis, Samuel Rupprechter, Andrea D'Olimpio.

Many thanks to all the people who have supported me from close or far, my mother for the inspiration towards academia and psychiatry, my sister Emma, all my colleagues and friends particularly Aaron Seitz, Matthias Hennig, Isabelle Duverne, Laetitia Pichevin, special thanks to Robert Hamilton who assisted with many steps of the way and to Grant Creegan who wisely came round when everything was already done.

I am hoping the book can be useful as a textbook in this new field and inspire a generation of students who can make a difference.

Chapter 1: Introduction: Toward a Computational Approach to Psychiatry

Janine M. Simmons, Bruce Cuthbert, Joshua A. Gordon, & Michele Ferrante.
National Institute of Mental Health (NIMH)

1.1 A Brief History of Psychiatry: Clinical Challenges & Treatment Development

1.1.1 Clinical Burden

Mental health disorders affect up to one in five adults in the USA and contribute substantially to worldwide morbidity and mortality. The lives of individuals with mental disorders are cut short by ten years on average (Walker, McGee and Druss 2015). Mental illness accounts for 7-13% of all-cause Disability Adjusted Life Years (DALYs). Because of the young age at which they strike, their chronicity and resistance to treatment, mental disorders account for an even greater proportion of all-cause Years Lived with Disability (21-32% YLDs; Whiteford, et al. 2015, Vigo, Thornicroft and Atun 2016). To take just one example, Major Depressive Disorder has become the second leading cause of non-communicable disease disability worldwide (Mrazek, Hornberger, Altar and Degitiar 2014). To lessen these burdens, we need new ways to understand and treat mental illnesses. Achieving this goal represents a substantial challenge. Therefore, it is imperative to utilize all tools at our disposal to make progress in psychiatric research.

1.1.2 Diagnostic Complexity

Although the behaviors associated with mental illness have been described for millennia (*i.e.*, a description of depression and dementia can be found in the "Ebers Papyrus," written in Egypt circa 1550 BC), psychiatry as a medical specialty emerged less than 150 years ago (Wilson 1993; Fisher 2012). As in all medical fields, clinicians and clinical researchers in psychiatry have attempted to establish discrete diagnostic categories to guide treatment and inform prognosis. However, the multi-faceted nature of mental disorders has made this task extremely complex. At the turn of the 20th Century, Emil Kraepelin's (1856-1926) work captured the extent of this challenge in ways that are still echoing today.

Importantly, he noted the difficulty of creating a complete nosology in the face of diverse, non-specific clinical presentations, and in the absence of a clear understanding of natural causes (Kendler and Jablensky 2011). Kraepelin's approach emphasized detailed, longitudinal clinical assessments with a focus on *syndromes* of commonly co-occurring symptoms. Through this process, he made one of the first diagnostic classifications in psychiatry. Specifically, he differentiated manic-depressive illness (bipolar disorder, in today's terminology) from 'dementia praecox' (Fisher, 2012). Subsequently, Bleuler (1857-1939) re-characterized and re-named dementia praecox as schizophrenia (Maatz, Hoff and Angst 2015). Although the diagnostic criteria and sub-types of these disorders have evolved over time, Kraepelin and Bleuler's fundamental clinical characterizations of different types of psychosis remain in use today.

In 1970, Robins and Guze sought to update the work of Kraepelin and Bleuler by establishing a method for achieving a more rigorous classification and improved diagnostic validity in psychiatry (Robins and Guze 1970). They recommended that psychiatric diagnoses be based upon five components. The first three of these re-emphasized the features of a thorough clinical assessment: 1. symptoms, demographics, and precipitating factors; 2. longitudinal course; 3. family history; while two additional elements would aid in the creation of homogenous diagnostic sub-groups: 4. laboratory studies and psychological tests; 5. exclusion criteria. This work proved hugely influential, even though the fourth component was not actually available. As the authors note themselves: "Unfortunately, consistent and reliable laboratory findings have not yet been demonstrated in the more common psychiatric disorders" (Robins and Guze 1970). In the 21st century, psychiatry still lacks objective and robust laboratory testing.

Following on these early efforts, the publication of the 3rd edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) in 1980 marked the modern age of psychiatric nosology (American Psychiatric Association 1980; Spitzer, Williams and Skodol 1980; Wilson 1993; Hyman 2010; Fisher 2012). Facing the challenges of a field without objective diagnostic testing, Spitzer and the American Psychiatric Association (APA) recognized that diagnostic validity might be out of reach. Given the extent of the knowledge gap, they sought to address the critical needs by: 1. Defining boundaries of mental disorders; 2. Stimulating progress in research and treatment development; and, 3. Increasing diagnostic reliability across research and treatment settings. DSM-III intentionally and

explicitly adopted an *a-theoretical*, descriptive approach that continues to serve as the bedrock of psychiatric diagnosis today (including DSM-5, (American Psychiatric Association, 2013)). In the DSM, each disorder is characterized by a list of possible symptoms and a minimum number of these symptoms are required to provide a diagnosis. For example, to meet criteria for Major Depressive Disorder, at least five of nine possible symptoms must be present concurrently. The DSM has become the *de facto* guide of both clinical psychiatry and psychiatric research. Newer versions of the DSM have largely achieved APA's goals of providing clinicians with a common clinical language, improving diagnostic reliability, and allowing rough classification of patients for treatment (Hyman 2010).

The DSM has established a common framework within which clinical psychiatrists can operate. However, the current diagnostic system poses challenges for multiple reasons. The DSM groups patients into diagnostic categories based on subsets of symptoms selected from a longer checklist. Patient groups become heterogeneous because individuals grouped into one category can have different (and sometimes non-overlapping) constellations of symptoms (**Table 1.1**). Moreover, because many symptoms are shared by more than one syndrome, comorbidity becomes the rule rather than the exception. As a result, patients frequently receive multiple diagnoses. It remains unclear whether this apparent comorbidity arises from underlying biology or simply reflects a classification system that is ill suited to capture the full complexity of human brain and behavior (Lilenfeld, Waldman, and Israel 1994, Maj 2005, Kaplan, et al. 2001, Sanislow et al. 2010). In addition, progress in genomics and neurobiology has revealed that different DSM diagnostic categories often share risk genes and so far, cannot be differentiated by neuroimaging (Farah and Gillihan 2012; Cross-Disorder Group of the Psychiatric Genomics Consortium 2013; Mayberg 2014; Simmons and Quinn 2014). Ideally, our diagnostic nosology should be informed by a deeper understanding of pathophysiology. In the USA, the National Institute of Mental Health (NIMH) has sought to address these problems through the RDoC initiative launched in 2010 (see **Section 1.4**).

< Table 1.1 near here: MDD symptom heterogeneity >

1.1.3 Treatment Development

In other areas of medicine, clinical advances have followed the availability of objective diagnostic tests, increased understanding of pathophysiological mechanisms, and the development of appropriate animal and computational models to rigorously test potential treatments. A prime example of this process can be seen in the improved treatments for cardiac arrhythmias resulting from identification and modeling of relevant cardiac ion channels (Bartos, Grandi, and Ripplinger 2015; Gomez, Cardona, and Trenor 2015). In psychiatry, treatment development has not yet followed such a path. Modern treatment options remain closely linked to psychotherapeutic and pharmacological approaches developed or discovered over fifty years ago.

Although there are a multitude of psychotherapy sub-types, three fundamental psychological models predominate. First, Sigmund Freud's (1856-1939) psychoanalytic theory emphasized the importance of the unconscious mind (see topographical model in **Figure 1.1**). Freud and his followers proposed that intra-psychic conflict led to mental illness. Therefore, psychodynamic psychotherapy seeks to bring unconscious material to conscious awareness, uncover unexpressed emotions, and resolve past experiences (Freud 1966; Blagys and Hilsenroth, 2000; Rawson 2005; Gabbard 2007).

Second, behavior therapy grew from the early 20th century psychological tradition of behaviorism (Watson 1913; Skinner 1938). In contrast to the psychodynamic focus on internal states, behaviorism prioritizes observable actions and proposes that all behavior is fundamentally a learned response to environmental stimuli. Which behaviors are expressed depends on prior experience with environmental contingencies, and mental illnesses consist of maladaptive learned responses (Mowrer 1947; Foa and Kozak 1986; Foa 2011). Behavior therapy seeks to eliminate psychiatric symptoms by disconnecting maladaptive behaviors from their environmental triggers or by forming new, more adaptive responses. For example, behavior therapists commonly use exposure therapy to treat anxiety disorders such as phobias, OCD, and PTSD (Foa and Kozak 1986; Foa 2011).

< Figure 1.1 near here: topographical model>

Third, in the 1950's and 1960's, Albert Ellis and Aaron Beck proposed new models for psychotherapy that integrated information processing and cognitive psychology (Ellis 1957; Beck 1991). In these models, automatic thoughts and core beliefs underlie emotions and behaviors (**Figure 1.2**). Depressive mood and anxiety disorders result from irrational beliefs, distorted perceptions, and automatic negative thoughts. Therefore, the goal of this type of therapy is to identify and correct these cognitive distortions. Cognitive and behavioral therapy techniques are often combined as CBT (Blagys and Hilsenroth 2000; Blagys and Hilsenroth 2002).

Whatever the method, psychotherapy has been shown to significantly reduce psychiatric symptoms and improve mental well-being over the long-term, with multiple meta-analyses demonstrating large effect sizes[1]. For psychodynamic psychotherapy, median effect sizes range from 0.69 to 1.8, depending on the targeted symptoms and length of treatment (Shedler 2010). A meta-analytic review of prolonged exposure therapy for PTSD demonstrated mean effect sizes of 1.08 for PTSD-specific symptoms and 0.77 for general symptoms of distress (Powers, Halpern, Ferenschack, Gillihan, and Foa 2010). Because CBT has been standardized, its benefits for depression and anxiety have been rigorously studied. Effect sizes are moderate-to-large, ranging from 0.58 to 1.0 (Shedler 2010). It should be noted, however, that very few psychotherapy outcome studies adequately assess the quality and fidelity of psychotherapy, even in research settings (Perepletchikova, Treat, and Kazdin 2007; Cox, Martinez, and Southam-Gerow 2019). In 2015, the Institute of Medicine found that psychosocial interventions proven effective in research settings have not been routinely integrated into clinical practice (IOM (Institute of Medicine), 2015).

< Figure 1.2 near here: Beck's Theory>

Whereas psychotherapeutic techniques have deep roots in historical theories of mental and/or behavioral processes, most breakthrough developments in psychiatric medications have occurred purely

[1] **Effect size** (d) is a statistical concept that measures the strength of the relationship between two variables on a numeric scale. Effect size is most commonly computed as the standardized difference between two means. The values of the effect

by chance (Preskorn 2010a). Serendipitous discoveries between the 1920s-1960s led to early pharmacological treatments for mental disorders, including chlorpromazine and other typical antipsychotics, lithium for bipolar disorder, and tricyclic antidepressants. During the second half of the 20th century, rational pharmaceutical development followed the accumulation of knowledge related to neurotransmitters. The greatest production of compounds targeting specific neurotransmitter systems occurred between the 1960s-1990s (Preskorn 2010b). Fluoxetine, the first selective serotonin reuptake inhibitor, received FDA approval for treatment of depression in 1987. Risperidone, an early "atypical" or second-generation antipsychotic (SGA), came on the market in 1993. Most recently, after a gap of more than 25 years, drugs that act at glutamate receptors have shown potential for acute treatment of depression and suicidality (Zanos et al. 2016; Lener et al. 2016).

Despite these advances, fundamental pharmacological treatment developments in psychiatry have stagnated (Hyman 2012, Insel 2015). First-line medication fails in approximately half of all patients, and the median effect size of treatment with any psychopharmacological agent is only 0.4 (Luecht et al. 2012). Moreover, for the last 25-30 years, almost all new psychiatric medications have been "me too" drugs-- closely related to the original chemical compound and acting through the same mechanism of action (Fibiger 2012; Harrison et al. 2016). Although newer medications can provide important reductions in associated side effects and greater tolerability, they are not more effective. For example, modern antipsychotics are no more effective than first-generation drugs, according to recent meta-analyses (Geddes, Freemantle, Harrison, and Bebbington 2000; Crossley, Constante, McGuire, and Power 2010). Anti-depressant efficacy remains difficult to differentiate from placebo effects (Khin et al. 2011) and lithium remains the most effective option for bipolar disorder, despite its limited tolerability and unclear mechanisms of action (Harrison et al. 2016).

After more than 100 years of psychological theories, psychopharmacological research, and clinical experience, the challenges of understanding and treating mental illness remain firmly in place. As a medical field, psychiatry faces two inter-related sticking points. The first is diagnostic complexity. Although DSM provides a foundation for clinical care in the face of limited treatment options, heterogeneous categories, individual differences, and comorbidity have stymied development of a principled pathophysiological understanding of psychiatric disorders. The second is stagnation in treatment development. Although both psychotherapeutic and pharmacological treatments have shown

efficacy, morbidity and mortality for people with serious mental illness remains unacceptably high (Insel 2012; Walker, McGee, and Druss 2015; Whiteford et al 2015; Vigo, Thornicroft, and Atun 2016). Solving these extremely difficult problems requires a set of novel conceptual approaches, including the integration of neuroscience findings and computational modeling.

1.4 Toward the Future of Psychiatric Research

In 2010, NIMH proposed a new conceptual model to guide clinical psychiatric research (Insel et al. 2010; Morris and Cuthbert 2012; Simmons and Quinn 2014). The Research Domain Criteria (RDoC) takes a fundamentally different approach than DSM and addresses a different set of proximate questions. The RDoC framework seeks to further our understanding of psychopathology through pathophysiology by building upon ongoing advances in the behavioral and neurobiological sciences. The RDoC model proposes that human behavior can be parsed into fundamental domains of function (currently Negative Valence, Positive Valence, Cognitive Systems, Social Processes, Arousal and Regulatory Processes, and Sensorimotor Systems; **Figure 1.3**). These domains can be further subdivided into core psychological-level constructs (*e.g.*, working memory; see (MacCorquodale and Meehl 1948)). RDoC hypothesizes that construct-level behaviors can be linked to the function of specific neural circuits and other biological processes, but also emphasizes the importance of developmental trajectories and environmental influences upon behavior. Constructs are conceptualized as dimensional, including the full continuum from illness to health, without specific clinical break points. The RDoC matrix provides a framework for investigations across multiple units of biological and behavioral analysis.

< Figure 1.3 near here: RDoC schematic>

RDoC was created as an attempt to move beyond the stagnation in psychiatric diagnosis and treatment development. The intent was to provide a framework based upon mechanisms of dysregulation in normative functioning, thus better aligning psychopathology research with rapidly evolving knowledge about neural systems and behavior. The specific elements of the framework are expected to change as new knowledge accumulates. The over-riding question is whether this framework can help

characterize psychiatric dysfunction more robustly; the long-term goal is to identify underlying mechanisms and specific functions that might serve as targets for treatment.

As a heuristic, RDoC can readily serve as a basis for the emerging field of Computational Psychiatry. RDoC provides a conceptual framework within which specific theories can be applied and quantitative models tested. Rather than considering psychiatric diagnoses as clusters of symptoms, RDoC functional domains and constructs can be conceptualized as resulting from sets of underlying computations taking place across interacting neural circuits. In theory, these neural processes can, in turn, be described by algorithmic representations that describe information processing in the system (Marr 1982; Hofstadter 1985; Hofstader 2008; Churchland and Sejnowski 1994; Damasio 2010; Redish 2013). Questions regarding the underlying neural circuits that perform those computations can then be asked. Stated differently, RDoC constructs can be considered latent constructs linking neurophysiological processes to behavioral observations (Huys, Moutsoussi, and Williams 2011; Maia & Frank 2011; Wang and Krystal 2014; Redish and Gordon 2016). Environmental factors and neurodevelopmental status can also be formally included in these algorithms. Ongoing RDoC experiments have begun to produce results that computational modelers can use as the basis for formalizing models that will better inform clinical practice. As such, applying computational approaches to RDoC-like frameworks may even transcend psychiatry and be used for advancing all kinds of translational neuroscience research (Sanislow et al 2019) e.g., computational neurology, computational vision, computational neuroscience of drug addiction. The goals of computational psychiatry and how computational models might best be applied to questions in behavioral neuroscience and psychiatry will be discussed in the following section.

1.2 Computational Approaches in Neuroscience & Psychiatry

1.2.1. Computational Neuroscience

Computational neuroscience formalizes the biological structures and mechanisms of the nervous system in terms of information processing. Computational neuroscience is a highly interdisciplinary field at the intersections of fields such as neuroscience, cognitive science, psychology, engineering, computer science, mathematics, and biophysics. The last 25 years have seen significant growth in this field. From 1991 to 2016, the field grew more than 200 times, from 2 peer-reviewed scientific articles published per year to more than 400 publications per year. During the mid-1980s, two key factors led to

this booming growth (Abbott 2008). The first factor was linked to the implementation and wide adoption of the back-propagation algorithm in artificial neural networks (Rumelhart and McClelland 1986, see **Chapter 2.1**). Adopting back-propagation led to a great expansion in the number of tasks that neural network models could handle, and consequently, in the number of scientists interested in questions answerable with these techniques. The second factor involved the translation of key concepts and mathematical approaches from physics into neuroscience (see **Chapter 2**). For instance, in the 1980s, physicists like John Hopfield and Daniel Amit elegantly showed how a memory model could be further analyzed using statistical techniques originally developed to address theoretical issues related to disordered magnets (Amit, Gutfreund, and Sompolinsky 1985; Hopfield 1982).

Mathematical models, such as those adopted from physics, have clear advantages over more abstract schematics and word descriptors. They force the modeler to be as precise, self-consistent, and as complete as possible in deriving the implications of the model. Such models can be used for different purposes:

- To describe the available data in a concise and synthetic way, possibly unifying different sets of data in the same formalism (*i.e.*, answer the question "what?" as in, what are the fundamental properties of the phenomenon studied?)
- To link the observed data to possible underlying mechanisms (*i.e.*, answer the question "how?" as in, how do the necessary and sufficient conditions for a phenomenon emerge?)
- To understand "why?" observed behaviors emerge as a consequence of a principle that can be justified theoretically (*e.g.*, through understanding the process of individual optimization under some biological, environmental, or developmental constraints).

In describing the question that a model can answer, it is also common to refer to Marr's levels of analysis (Marr 1982; **Figure 1.4**). Marr proposed that information processing systems can be understood at three distinct, complementary levels of analysis: computational, algorithmic, and implementation. The computational level specifies the problem to be solved in terms of some generic input-output mapping (e.g. 'list sorting'). The algorithm specifies how the problem can be solved, what representations it uses, and what processes it employs to build and manipulate the representations (e.g. 'Quick-sort', 'Bubble-sort'). Implementation is the level of physical parts and their organization (e.g. specific programming language). It describes the physical mechanisms that carry out the algorithm. These levels function

mostly independently and can often be described using different mathematical models. The challenge is often to bridge these levels of description and to understand how they might constrain each other.

<Figure 1.4 near here: Levels of Marr>

Computational approaches have dramatically changed how basic neurobiological phenomena are described. Examples of successful and influential theoretical models include: Lapicque's integrate-and-fire model (Lapicque 1907; Lapicque 1926), which provides a simple model of neuronal changes in voltage and activity (see **Section 2.1.5**); Hodgkin-Huxley's model of action potential generation and propagation (Hodgkin & Huxley, 1952), which provides a detailed description of the dynamics of sodium and potassium channels in the initiation of the action potential; Rall's cable theory (Rall 1977), which provides a description of how the neuronal voltage relates to various morphological properties of neuronal processes (e.g., axons and dendrites); and Hebb's plasticity rules (Hebb 1949), which provide an algorithm to update the strength of neuronal connections within neural networks (*e.g.*., the synaptic plasticity, as a consequence of learning).

Fundamental theoretical advances in neuroscience have also changed how we view information encoding and computation in the brain. This can be seen, for example, through the application of information theory in Barlow's hypothesis of predictive coding (see also **Section 2.4.6**). A prominent figure of theoretical neuroscience Horace Barlow (1985) suggested that the hierarchic organization of sensory systems reflects two imperatives: (1) to take in a maximum of new information to detect statistical regularities in the environment; and (2) to exploit these learned regularities to construct predictions about the environment. Those predictions are then used to guide adaptive behavior. In other words, he proposed that the brain evolved to efficiently code sensory information by using information-processing strategies optimized to the statistics of the perceptual environment (Olshausen & Field, 1997). This framework suggests that the brain functions as a predictive engine rather than a purely reactive sensory organ, using an internal model of the statistics of the world to continuously and automatically infer what environment it is placed in, and what best action to take.

More recently, it has been proposed that computational models of cognitive function could be used to explain psychopathology. For example, impairments in the processes involved in predictive coding could explain a variety of observations, ranging from impoverished theory of mind in autism to

abnormalities of smooth pursuit eye movements in schizophrenia (see **Section 2.4.6 & Chapter 6**). The development of such ideas marked the birth of the field of "Computational Psychiatry."

1.2.2 Computational Psychiatry

Simply defined, computational psychiatry consists of applying computational modeling and theoretical approaches to psychiatric questions. Although very young, computational psychiatry is already an extremely diverse field, leveraging concepts from psychiatry, psychology, computer science, neuroscience, electrical/chemical engineering, mathematics, and biophysics. Computational psychiatry seeks to understand how and why the nervous system may process information in dysregulated ways, thereby giving rise to the full spectrum of psychopathological states and behaviors. It seeks to elucidate how psychiatric dysfunctions may mechanistically emerge, be classified, predicted, and clinically addressed. Computational psychiatry models can also be used to connect distinct levels of analysis through biologically grounded theories and rigorous analytical methods.

Integrating computational modeling into psychiatry can aid research in several fundamental ways. First, in a formalized computational model, all assumptions underlying a clinical characterization, moderating factors, and experimental hypotheses must be made explicit. Both manipulated (independent) and measured (dependent) variables can be included as factors in mathematical formulas. The extent to which experimental results match model predictions can qualitatively and quantitatively inform our mechanistic understanding and guide future experiments. In this way, developing and testing computational models provides a clear, iterative approach to increased understanding of psychopathological complexity. Additionally, computational models can explicitly incorporate time, enriching our ability to understand how functional neuro-cognitive architectures develop and to identify critical temporal windows associated with abnormal developmental trajectories.

Similar to their broader role in computational neuroscience, computational models can help psychiatric researchers answer three fundamental questions regarding the differences in neural information processing that may characterize psychopathology:

- *What* are the main biological components involved in psychopathology and what are the mathematical relationships between these components? Computational approaches require clear and precise definitions of the basic building blocks of cognitive functions and their impairment.
- *How* do dysfunctions in the individual biological units or in their interactions lead to the behavioral changes seen in mental illness? Answers to this question may allow targeted and dynamic manipulations of the system to treat the emotional, cognitive, and behavioral problems associated with psychiatric disorders.
- *Why* have these changes occurred? Understanding etiology in a dynamical system is most challenging, because early, initial changes may have had downstream effects on several nodes. Full investigation therefore requires integrating time into computational models and testing their predictive value using longitudinal designs across various neurodevelopmental trajectories.

This approach, which seeks to bridge the gap between neuroscience and psychopathology, is consistent with the RDoC research framework because it conceptualizes psychopathology with reference to specific neural circuits (what), seeks to understand the relationships between psychological constructs and neurobiological function (how), and explicitly considers the impact of both biological and environmental etiological factors on neurodevelopmental trajectories (why). Progress about these questions should open up a range of potential preventative approaches to mental illness.

1.2.3 Data-driven approaches

While a wide spectrum of computational approaches exists, computational models in psychiatry can be divided into two broad groups: data-driven models and theory-driven approaches (Huys, Maia, & Frank, 2016). This book focuses on theory-driven models. However, the two approaches are complementary, equally promising and can be combined.

The data-driven approach to computational psychiatry can be described as the application of machine learning techniques to vast amount of data related to psychiatric patients, without explicit reference to current psychological or neurobiological theory. The goal is to find new statistical relations between high-dimensional datasets (*e.g.*, genetics, neuroimaging findings, behavioral performances, self-report questionnaires, response to treatment, etc.) that could be meaningful for classification,

treatment selection or prediction of treatment outcome. Here, the assumption is that our understanding of mental illness will improve primarily through improvements in data quality, quantity, and analytics. This "blind" or "brute force" approach may allow researchers to generate new theories based purely on multi-faceted clinical data rather than on potentially outmoded historical perspectives (Huys, Maia, and Frank 2016).

Figure 1.5 illustrates how a data-driven approach might lead to new descriptions and classifications, going beyond traditional symptom-based categories of mental disorder. Consider a population of patients with a range of mood disorders (*e.g.*, major depressive disorder, dysthymia, and bipolar disorder). Machine learning techniques applied to a range of genetic, physiological, brain activity, behavioral data and social, cultural and environmental factors related to those patients, might lead to the discovery of new unbiasedly derived bio-behavioral clusters. Such clusters might form groups that are more homogeneous than the original DSM classification and might connect more directly with the underlying causes of the illness.

< Figure 1.5 near here: Data drive Computational Psychiatry >

1.2.4 Theory-driven approaches

Theory-driven models are the focus of this book and can be described as the application of "classical" computational neuroscience approaches to psychiatric questions. In general, these models mathematically describe the relationships between observable variables (*e.g.*, behaviors) and theoretically relevant, but potentially unobservable, biological mechanisms. Theory-driven models commonly incorporate known experimental knowledge of brain anatomy and/or physiology or of higher-level functions for which basic theories have been developed (*e.g.*, perception, learning or decision making). These models are particularly useful when the cognitive/behavioral function of a neurobiological network is known and/or when accurate and detailed experimental data are available to constrain the model. Theory-driven models can span across multiple levels of analysis and abstraction, from molecules to complex behaviors. They can show whether existent data are sufficient to explain the measured physiological behavior of the circuit; they can also highlight whether unaccounted biological mechanisms could better explain the data, and they can point to gaps in knowledge.

Three representative exemplars of theory-driven models are discussed below: biophysically realistic neural-network models, reinforcement learning models, and Bayesian models (Huys, Maia, and Frank 2016).

Biophysical models are commonly used to elucidate how biological abnormalities found in mental disorders affect neuro-behavioral dynamics. Biophysical models rely on the theoretical assumption that the essential computations of single neurons and synapses can be captured by sets of first order differential equations of the type proposed by Hodgkin and Huxley (1952). *Synthetic* computational models recapitulate biophysically realistic properties of neurons and can be used to test the proposed input-output properties of neurophysiological systems in a behavioral context (Wang and Krystal 2014; Ferrante et al 2016; Huys, Maia, and Frank 2016; Shay, Ferrante, Chapman, and Hasselmo, 2016). Biological models are most appropriate when our biological knowledge base is well established and could help identify biological mechanisms that best explain natural variance in patient populations. Biologically realistic models vary in complexity, and the optimal degree of biological detail depends upon the scientific question asked. Simpler models can be more generalizable, while complex models may lose their reductionist appeal as they increase their biological realism. Because biological realism tends to be computationally expensive, these models are most easily implemented when the network is relatively small and/or when the relevant biological parameters are relatively few. A reductionist approach incorporating the fundamental biological features of a complex system can be the simplest possible framework to elucidate the relationship among biological mechanism, neural computations, and functional output. Examples include models of dopamine signals linked to reward-prediction error, working memory internally represented as sustained neural activity, and neural integrators in perceptual decision-making tasks, all of which are relevant to our understanding of some of the cognitive impairments observed in psychiatry (see **Chapters 2-4**). On the other hand, using biologically realistic models might be premature for explaining other psychiatric symptoms, such as psychosis, where a clear neurophysiological characterization at the cellular and systems level is still lacking (Wang and Krystal 2014, see also **Chapter 6**).

Reinforcement learning (RL) as a research field lays at the intersection between mathematical psychology, artificial intelligence, and control theory. It addresses how systems of any sort, be they artificial or natural, can learn to gain rewards and avoid punishments in what might be very complicated environments involving states (such as locations in a maze) and transitions between states. They

describe how an agent 'should' behave under some explicit notion of what that agent is trying to optimize. In that sense, they offer a normative framework to understand behavior.

Reinforcement learning was born from the combination of two long and rich research traditions, which had previously been pursued independently (Sutton and Barto 2018). The first thread concerns optimal control and solutions using value functions and dynamic programming. Optimal control relates to mathematical techniques that deal with the problem of finding a control law for a given system such that a certain optimality criterion is achieved. The second thread concerns learning by trial and error. This thread finds its origin in the psychology of animal learning, particularly in the scientific exploration of Pavlovian (classical) and instrumental (operant) conditioning. Classical conditioning is a form of learning whereby a neutral stimulus (called the *conditioned stimulus*, CS) becomes associated with an unrelated rewarding or punishing stimulus (called the *unconditioned stimulus*, US) in order to produce a behavioral response (called *conditioned response,* CR*)*. In the famous example studied by Pavlov, the repeated pairing between a bell (the CS) and food (the US) would lead to dogs salivating (the CR) when the bell was presented alone. Instrumental conditioning relates to learning associations between actions and outcomes. B. F. Skinner showed that behaviors followed by positive reinforcement are more likely to be repeated, while behaviors followed by negative reinforcement are more likely to be extinguished. Pavlovian conditioning, instrumental conditioning, and subsequent research show that animals and humans naturally learn the associations between objects, actions, and reinforcement contingencies in their environment and use this learning to predict future outcomes. Learning occurs to optimise those predictions (or reduce prediction errors). Interestingly, studies of operant conditioning also form a basis for some modern psychotherapies, particularly behavioral psychotherapies, which offer methods designed to reinforce desired behaviors and eliminate undesired behaviors. As such, such models relate naturally with psychiatry.

Converging evidence from lesion studies, pharmacological manipulations, and electrophysiological recordings in behaving animals, as well as fMRI signals in humans, have provided links between RL models and neural structures. In particular, a significant body of literature suggests that the neuromodulator dopamine provides a key reinforcement signal: the temporal difference reward prediction error (see **Sections 2.3 and 5.3**). Dopamine dependent temporal difference models provide a key link between neuromodulation (often hypothesized to be dysregulated in mental illness), pharmaceutical treatments, substances of abuse, and learning systems.

Finally, a prominent idea in modern computational neuroscience is that the brain maintains and updates internal probabilistic models of the world that serve to interpret the environment and guide our actions. In doing so, it uses calculations akin to the well-known statistical methods of Bayesian inference **(see Section 2.4)**. Bayesian inference methods are used to update the probability for a hypothesis, as more evidence or information becomes available. When applied to psychiatry, this approach conceptualizes mental illness as the brain trying to interpret the world through distorted internal probabilistic models, or incorrectly combining such internal models with sensory information, generating maladaptive beliefs.

Bayesian models can be particularly useful in predicting expected behaviors (what would be the optimal thing to do in a given task), quantifying the severity of dysfunctional behavior as the 'distance' from optimality, and understanding how maladaptive beliefs can arise. Traditionally, Bayesian inference has been applied primarily to behavioral data, but more recently there has been an effort to integrate behavioral data with neural or fMRI data (Fischer and Peña 2011; Turner et al 2013).

These main types of theory-driven models - biophysical models, RL, predictive coding and Bayesian models - will be described in more detail in **Chapter 2**.

Of course, theory- and data-driven models are not mutually exclusive. Theory-driven models are often heavily grounded to and validated by experimental data. Similarly, fully unbiased methods of collecting and analyzing data do not exist and often incorporate hypotheses that can be formulated as theories. Both approaches are complementary. Ultimately, they will need to be combined to provide precise diagnostic classifications, predictions, and explanations of mechanistic neurobehavioral trajectories.

A number of initiatives have been created that encourage the development of both types of approaches. As discussed above, NIMH's RDoC initiative encourages psychiatric researchers to study focused aspects of dysfunction that may cut across current diagnostic categories and link mechanistic explanations across different levels of biological analysis. The BRAIN Initiative has fostered the development of innovative neuro-technologies able to record simultaneously from large numbers of cells and to stimulate brain activity with high spatio-temporal precision. Together, these initiatives are generating large, complex, multimodal datasets that will provide fertile ground for cutting-edge computational modeling.

Computational Psychiatry is undoubtedly rising. The first article to use the term computational psychiatry was published in 2007 (Montague 2007). In the following 10 years, the field has rapidly expanded, with 220 publications and several technical books (Parks, Levine, and Long 1999; Sun 2008; Redish and Gordon 2016; Anticevic and Murray 2017; Heinz 2017; Wollace 2017). Groups interested in such questions, as well as summer schools and workshops, have also recently blossomed. However, the field is still in its infancy and comprehensive models able to explain psychopathology at the individual level still need to be implemented. We hope that this book will inspire a new cohort of scientists and help towards a new understanding and treatment of mental illness.

1.3 Structure of the book

In the next section, we survey the main methods of theory-driven computational psychiatry. We will cover neural networks and connectionist methods, drift-diffusion models, reinforcement learning models, predictive coding, and Bayesian models as well as methods related to fitting computational models to behavioral data.

In the spirit of RDOC (cf. **Figure 1.3**), the following section describes models relevant to the dimensions of behavioral functioning, focusing on models of healthy function with an emphasis on cognitive systems and positive and negative valence systems. **Chapter 3** describes biologically detailed models of working memory and decision-making. **Chapter 4** describes models of Cognitive Control. **Chapter 5** focuses on reinforcement systems. The following chapters then illustrate the application of computational approaches to schizophrenia (**Chapter 6**), depression (**Chapter 7**), anxiety (**Chapter 8**), addiction (**Chapter 9**) and the example of a tic disorder (Tourette's Syndrome, **Chapter 10**). In **Chapter 11**, we offer additional pointers on disorders not covered in **Chapters 5-10** and offer some guidelines for future research.

1.4 Chapter Summary

- The burden of mental health diseases is enormous in terms of suffering, life expectancy, and economic cost.

- There has been stagnation in the discovery of new pharmacological drugs and treatments in the last decades.
- The definition and diagnosis of psychiatric disorders has been problematic for centuries. It is likely that, for most disorders, it will be impossible to pin down a single cause, a single organic substrate, or a single time course. The current categorical classification of mental disorders, known as the DSM, has proved to be clinically very valuable, but the heterogeneous phenotypes associated with DSM-based diagnoses and the Manual's a-theoretical structure make it difficult to consider biological mechanisms that could lead to more effective treatments.
- New approaches aim to move from the description of mental illnesses as collections of symptoms toward methods to bridge neuroscience and cognitive modeling with psychopathology. The NIMH RDoC initiative encourages this approach and computational modeling can provide a useful approach to solve some challenges highlighted by RDoC (*e.g.*, causally linking distinct units of analysis, modeling temporal trajectories and dynamic interactions between specific constructs across neurodevelopment).
- Computational approaches are considered central to progress in neuroscience. They could similarly benefit the field of psychiatry.

1.5 Further Study

A historical perspective about the field of Psychiatry can be found in Fisher, B.A. (2012). For reviews describing the emerging field of Computational Psychiatry, see Montague, Dolan, Friston and Dayan (2012); Friston, Stephan, Montague, Dolan (2014) and Stephan and Mathys (2014). To read about the NIMH's RDoC initiative, the reader can consult Kozak and Cuthbert (2016).

Chapter 2: Methods of Computational Psychiatry: A brief Survey

Peggy Seriès, University of Edinburgh

"*One thing I have learned in a long life: That all our science, measured against reality, is primitive and childlike — and yet it is the most precious thing we have."*

Albert Einstein. Creator and Rebel, 1972.

The methods that are currently used in computational psychiatry are very diverse, mirroring progress in computational neuroscience and cognitive science, and ranging from early connectionist work to reinforcement learning, probabilistic methods and applied machine learning. This chapter offers a brief survey of these methods, as well as pointers to additional resources for further study.

2.1 Neural Networks & Circuits Approach

The earliest models that aimed at explaining mental computations and disorders are known as "connectionist" models. Donald Hebb introduced the term "connectionism" in the 1940s to describe a set of approaches that models mental or behavioral phenomena as emergent processes in *interconnected networks of simple units* (**Figure 2.1**), a. k. a "Neural Networks".

Those simple units, often called "neurons" by analogy with the brain, are described by their value or "output" which can be binary (1/0) or real-valued. The value of each unit is equal to the sum of its inputs, passed through a non-linear function, called the "activation function". The network connections typically have a "weight" that determines how strongly the units influence each other. The weight can be positive or negative and may change according to a learning procedure. Units may also have a threshold such that only if the sum of the signals it receives crosses that threshold is the unit activated.

For example, McCulloch and Pitts (1943) proposed a binary threshold unit as a computational model for an artificial neuron. This neuron computes a weighted sum of its n input signals x_j: $y = f\left(\sum_{j=1}^{n} w_j x_j - u\right);$ (1)

and generates an output y of 1 if the sum is above a certain threshold u. Otherwise, it outputs 0. Here, f is chosen to be a unit step function, whose value is zero for negative argument and one for positive argument and w_j is the synapse weight associated with the j_{th} input. Positive weights correspond to excitatory synapses, while negative weights model inhibitory ones. McCulloch and Pitts proved that, in principle, this model could be used to implement any Boolean logic function, such as AND, OR, XOR[2] gates (the latter by combining AND, NOT and OR units). These logical gates are the building blocks of the digital logic electronic circuits, which modern digital computers are built from. In principle therefore, such circuits could achieve any type of computation. The McCulloch & Pitts neuron has been generalized in many ways. Different types of activation functions can be used instead of the unit step function, for example sigmoidal functions such as the logistic function: $f(x) = 1/(1 + \exp(-bx))$ where b is a slope parameter.

< Figure 2.1 around here>

2.1.1 Artificial neural network architectures

Typically, artificial neural networks are organized in layers. Neural networks can be feed-forward or recurrent. In feed-forward networks, the information moves in only one direction, forward, from the input nodes, through the intermediate nodes (if any) that are also called "hidden" nodes, and to the output nodes. The so-called multilayer perceptron for e.g. has neurons organised into layers that have unidirectional connections between them. There are no cycles or loops in the network. In recurrent

[2] The XOr, or "exclusive or", is a digital logic gate that gives a true (1) output when the number of true inputs is odd. An XOR gate implements an exclusive or; that is, a true output results if one, and only one, of the inputs to the gate is true. If both inputs are false (0) or both are true, a false output results.

networks on the contrary, units in the same layer are interconnected. While feed-forward networks are static or "memory-less", in the sense that they produce only one set of output values from a given input, recurrent networks can produce sequences of values and rich temporal dynamics and are considered more biologically plausible since neurons in the brain are also heavily interconnected in circuits that present loops and can generate rich dynamics, such as oscillations.

2.1.2 Learning in Feed-Forward networks

Learning in neural networks is viewed as the problem of updating network architecture and connection weights so that the network becomes more efficient at performing a specific task, defined as mapping a desired output to a given input. At a theoretical level, we can distinguish three main learning paradigms: supervised, by reinforcement and unsupervised. In supervised learning, the network is provided with a correct output for every input pattern in a training dataset. Weights are dynamically updated to allow the network to produce answers as close as possible to the known correct answers. This is achieved using learning rules known as error-correction rules. Reinforcement learning is a variant of supervised learning in which the network is provided with only a critique on the correctness of network output (correct/incorrect or reward/punishment), not the correct answers themselves. In contrast, unsupervised learning does not require a correct answer associated with each input pattern in the training dataset. It explores the underlying structure in the data, or correlations between patterns in the data, and organizes patterns into categories from these correlations.

The basic principle of error-correction rules is to use the error between the real output of the network and the desired output ($y_{desired} - y$) to modify the connection weights so as to gradually reduce this error, using a method known as gradient descent. For example, the perceptr**on** learning rule is based on this error-correction principle. A perceptron consists of a single neuron receiving a number of inputs $\mathbf{x}=\{x_1, x_2\ldots, x_n\}$ fed through connections with adjustable weights w_i and threshold u (**Figure 2.1A**). The net input v of the neuron is:

$$v = \sum_{i}^{n} w_i x_i - u$$

and the output is set to +1 if v>0 and 0 otherwise. The perceptron can be used to classify two classes of inputs: one set of inputs will be trained to lead to an output of 1, while the other set will lead to an output of 0. Rosenblatt (1958) showed that this can be achieved by using the following steps:

1) Initialize the weights and threshold to small random numbers;
2) Present an input vector $\mathbf{x_j}=\{x_{j,1}, x_{j,2,..}, x_{j,n}\}$ for pattern *j* and evaluate the output of the neuron y_j;
3) Update the weights according to:

$$w_i(t+1) = w_i(t) + \eta(d_j - y_j(t))x_{j,i} \qquad (3)$$

where d_j is the desired output for pattern *j*, t is the iteration number and η is the learning rate**,** which determines how much we adjust the weights in each trial with respect to the loss gradient (the lower it is, the slower we travel along the downward slope of the gradient).

Like most AI researchers, Rosenblatt was very optimistic about the power of neural networks, predicting enthusiastically that the "perceptron may eventually be able to learn, make decisions, and translate languages." However, Minsky and Papert (1969) showed that, as a general result, single-layer perceptron are very limited in what they can do: they can only separate linearly separable patterns. It fails to implement the XOR function, for example. This was perceived as devastating result concerning what could be achieved with neural networks and played a part to what is known as the "AI winter" in the 1970s and 1980s: a period of reduced funding and interest in artificial intelligence research.

The "AI winter" came to an end in the middle 1980s, when the work of John Hopfield and David Rumelhart revived interest in neural networks. Rumelhart et al (1986) showed that that error correction rules could be adapted to multi-layer networks, and provided a method that made neural networks able to approximate any nonlinear function: the back-propagation algorithm. Such networks (and the variants that followed) can be trained to perform sophisticated classification or optimization tasks, such as character recognition and speech recognition. These new results led to a new growth of the field. Neural networks would become commercially successful in the 1990s. Initially limited by the computational power of early computers, this research was leading the way to the current success of Deep Learning networks, which are based on similar principles.

2.1.3 Recurrent Networks and Attractor Dynamics

Around the same time as the back-propagation algorithm was introduced, physicist John Hopfield was able to prove that another form of neural network, now called a "Hopfield net", could learn and process information in a completely new way. The units in a Hopfield net are binary threshold units, like in the perceptron, so they take only two different values, usually 1 and -1, depending on whether or not the units' summed input exceeds their threshold (**Figure 2.1B**). The connections in a Hopfield net typically have the following restrictions: i) no unit has a connection with itself: $w_{ii} \neq 0$; and ii) the connections are symmetric: $w_{ij}=w_{ji}$.

Updating one unit in the Hopfield network is performed using the following rule:

$$x_i = Sgn(\sum_{j=1}^{n} w_{ij}x_j - b_i) \quad (4)$$

where Sgn(x) is the sign function, whose output is 1 or -1, w_{ij} is the weight of the connection from unit *j* to unit *i*; x_j is the state of unit *j*, b_i is the threshold of unit i.

Updates in the Hopfield network can be performed in two different ways: either asynchronously: Only one unit is updated at a time. This unit can be picked at random, or a pre-defined order can be imposed from the very beginning, or synchronously: All units are updated at the same time. This requires a central clock to the system in order to maintain synchronization. Importantly, Hopfield found that the network could be described by a quantity that he called the energy E (by analogy with the potential energy of spin glass**)**, defined by:

$$E = -\frac{1}{2}\sum_{i,j=1}^{n} w_{ij}x_i x_j - \sum_{i=1}^{n} b_i\, x_i \qquad (5)$$

As the network state evolves according to the network dynamics, E always decreases and eventually reaches a local minimum point, called attractor, where the energy stays constant. Hopfield also showed that those energy minima could be set to correspond to particular n-dimensional patterns $\{ \varepsilon_1, \varepsilon_2 \ldots \varepsilon$ This is done by setting the weight from unit *j* to unit *i* such that it corresponds to the average (over all patterns) product of the i^{th} and j^{th} elements of each pattern:

$$w_{ij} = \frac{1}{n}\sum_{k=1}^{n} \varepsilon_i{}^k \varepsilon_j{}^k \qquad (6)$$

where $\varepsilon^k = \{\varepsilon_1{}^k, \varepsilon_2{}^k, \ldots, \varepsilon_n{}^k\}$ denotes the pattern number k to be encoded. This is called the storage stage. The network can then be used as an associative memory: in the so-called retrieval stage, an input is

given to the network to be used as initial state of the network, and the network will evolve according to its dynamics to finally reach an equilibrium that will correspond to the stored pattern that is most similar to the input. For example, if we train a Hopfield net with five units so that the state (1, -1, 1, -1, 1) is an energy minimum, and we give the network the state (1, -1, -1, -1, 1), it will converge to (1, -1, 1, -1, 1).

2.1.4 Application to Psychiatry

The discovery of the back-propagation algorithm and Hopfield networks triggered a strong revival of interest for neural networks. In cognitive science, connectionism – as a movement that hopes to explain intellectual abilities in terms of neural networks – became further inspired by the appearance of *Parallel Distributed Processing* (PDP) in 1986. This is a two volume collection of papers edited by David E. Rumelhart, Geoff Hinton and psychologist James L. McClelland (D. E. Rumelhart, Hinton, and Williams 1986) that has been particularly influential. Connectionism offered a new theory about cognition, knowledge and learning – and their impairments. In theory, neural networks can be trained to perform any task (pattern classification, categorization, function approximation, prediction, optimization, content addressable memory, control) to the level of human participants. The PDP approach led to the idea that possible impairments in cognitive function, such as those observed in mental illness for example, could be explained by impairments in either the structure or the elements of the underlying neural networks e.g. the destruction of some connections, or an increase in the noise of some units.

For example, connectionist models have been prominently applied to schizophrenia. Patients with schizophrenia or mania can characteristically display hallucinations and delusions as well as rapidly changing, loose associations in their speech. Early work examined how parameters governing the dynamics of Hopfield networks might reproduce this. An increase in noise can lead to less specific memories, mirroring a broadening of associations in schizophrenia, and less stable, constantly altering memories. Similarly, deletions of connections, mimicking excessive pruning, or overload of the network with memories beyond its capacity, produce the emergence of localized, parasitic attractors, reminiscent of hallucinations or delusions (for a review, see Hoffman and McGlashan 2001).

2.1.5 Biological Networks

More recently, such hypotheses have been explored in the context of neural networks that are much more biologically realistic. Those networks are made of so-called "spiking" neurons that mimic what is known of real neurons: biological neurons use short and sudden increases in voltage, known as action potentials or "spikes", to send information (**Figure 2.1C**). The leaky integrate-and-fire neuron (LIF) is probably the simplest example of a spiking neuron model but it is still very popular due to the ease with which it can be analyzed and simulated.

The state of the neuron at time *t* is described by the membrane potential of its soma *v(t)*. The neuron is modeled as a "leaky integrator" of its input I(t):

$$\tau_m \frac{dv(t)}{dt} = -v(t) + RI(t) \qquad (7)$$

Here, τ_m is the membrane time constant and R is the membrane resistance. In electronics terms, this equation describes a simple resistor-capacitor (RC) circuit: the membrane of a neuron can be described as a capacitor because of its ability to store and separate charges. Ion channels allow current to flow in and out of the cell. When more ion channels are open, more ions are able to flow. This represents a decreased resistance, which leads to an increase in conductance.

The dynamics of the spike are not explicitly modelled in the LIF model. Instead, when the membrane potential v(t) reaches a certain threshold v_{th} (spiking threshold), it is instantaneously reset to a lower value v_r (reset potential) and the leaky integration process described by Eq. 7 starts anew with the initial value v_r. To add more realism, it is possible to add an absolute refractory period Δabs immediately after v(t) crosses the threshold v_{th}. During the absolute refractory period, *v(t)* might be clamped to v_r and the leaky integration process is re-initiated following a delay of Δ*abs* after the spike.

The input current can be constant, or dynamic. If the neuron is modelled as part of the network, the input current will reflect the synaptic inputs coming from other neurons. These in turn can be modelled as weighted inputs, where each connection is given a weight (positive for excitatory neurons or negative for inhibitory neurons), or in a more realistic way as a synaptic conductance that model the dynamics of

real synaptic inputs (excitatory post-synaptic potentials, a.k.a. EPSPs – and inhibitory post-synaptic potentials, a.k.a. IPSPs).

More detailed information about modelling individual neurons and networks of biologically realistic neurons can be found for example in Dayan & Abbott (2000).

Chapter 3 describes applications of such spiking neural networks to understand decision-making and working memory deficits in healthy subjects and schizophrenia. Patients with schizophrenia also show impairments in cognitive flexibility and control tasks that require the inhibition of a pre-potent response. **Chapter 4** shows how modelling the circuits involved in those tasks might lead to a better understanding of the cognitive control deficits in mental illness.

2.2 Drift-Diffusion models[3]

Drift-Diffusion models (DDM) belong to another class of models that are also inspired from Physics. Here, the aim is to provide a phenomenological description of a particular psychological process: the performance of animals or humans when they make simple decisions between two choices, without worrying about the underlying possible biological substrate. These models are interesting because although they were initially proposed only as phenomenological description of psychological processes, it is now clear that they also connect to notions of optimal decision theory as well as observed dynamical processes in real biological neurons.

The DDM is applied to relatively fast decisions (commonly less than 2 seconds) and only to decisions that are a single-stage decision process (as opposed to the multiple-stage processes that might be involved in, for example, reasoning tasks). Such tasks include for example perceptual discrimination (are these two objects the same or different?), recognition memory (is this image new or was it presented before?), lexical decision (is this a word or a non-word?) etc. Performance is described in terms of reaction times and accuracy. Such tasks are commonly used in Psychiatry to assess how information is

[3] This subsection is loosely based on White et al (2010).

processed in different groups. For example, whether anxious or depressed participants process threatening or negative information differently from controls when they have to make simple decisions.

Drift decision models aim at dissecting the different elements that are involved in the decision: in particular at separating the quality of evidence entering the decision from decision criteria and from other, non-decision, processes such as stimulus encoding and response execution. In these models, decisions are made by accumulating noisy evidence, until a threshold has been reached, at which point a response is initiated (**Figure 2.2**).

< Figure 2.2 around here>

Several mathematical expressions exist for the DDM. A typical equation will be of the form of a Wiener process (one dimensional Brownian motion). The diffusion process *x(t)* evolves dynamically according to:

$$\frac{dx(t)}{dt} = v + \sigma\eta(t) \quad (8)$$

- Where *v* is called the **drift rate**. It represents the quality of the information evidence from the stimulus. If the stimulus is easily classified, it will have a high rate of drift and approach the correct boundary quickly, leading to fast and accurate responses.
- $\eta(t)$ is a white noise term.
- σ^2 is the variance of the process.

In the model, noisy evidence is accumulated from a starting point, *z*, to one of two boundaries, *a,* or *0*. The two boundaries represent the two possible decisions, such as yes/no, word/non-word, etc. Once the process *x(t)* reaches a boundary, the corresponding response is initiated.

Each component of the model – the boundary separation (*a*), drift rate (*v*), starting point (*z*), and non-decision processing (T_{er}) - has a straightforward psychological interpretation. The position of the starting point, *z*, indexes response bias. If an individual is biased towards a response (e.g., through different frequencies of each option, or payoffs), their starting point will be closer to the corresponding boundary, meaning that less evidence is required to make that response. This will lead to faster and more probable

responses at that boundary compared to the other. The separation *a* between the two boundaries indexes response caution or speed/accuracy settings. A wide boundary separation reflects a cautious response style. In this case, the accumulation process will take longer to reach a boundary, but it is less likely to hit the wrong boundary by mistake, producing slow but accurate responses. One can also add parameters that capture between-trial variability in the starting point, drift rate, and non-decision time. Such variability is necessary for the model to correctly account for the relative speeds of correct and error responses. The model can also be extended to include contaminants, i.e. responses that come from some process other than the diffusion decision process (e.g., lapses in attention) so as to account for aberrant responses or outliers in the data.

This model was shown to be a satisfying description of the choice process as it produces the characteristic right skew of empirical RT distributions. For more mathematical details of the diffusion model, interested readers can consult Ratcliff and Tuerlinckx (2002) or Ratcliff and Smith (2004). There are several advantages of the diffusion model over traditional analyses of RTs and/or accuracy. First, it allows for the decomposition of behavioral data into processing components. This allows researchers to compare values of response caution, response bias, non-decision time, and stimulus evidence. With this approach, researchers can better identify the source(s) of differences between groups of subjects.

For example, White et al (2010b) used the diffusion model to study how processing of threatening information might differ in high-anxious individuals. In their lexical decision experiment, participants were shown strings of letters and had to decide if the strings were words or non-words. Some words were threatening words, while others were neutral. They found a consistent processing advantage for threatening words in high-anxious individuals, even in situations that did not present a competition between different inputs, whereas traditional comparisons showed no significant differences. Specifically, participants with high anxiety had larger drift rates for threatening compared to non-threatening words whereas participants with low anxiety did not.

Another advantage of this model is that, by fitting RTs and accuracy jointly, it can aid with the identification of different types of bias that are notoriously hard to discriminate: in particular disentangling discriminability (a change in the quality of evidence from the stimulus) vs. response bias (a shift of the decision criterion). Finally, because it uses all the data at once, contrary to classical analyses, which look at RTs or percent correct separately, it is potentially more sensitive to detect

differences.

2.2.1 Optimality and Model Extensions

A number of extensions and variants of the DDM have been proposed. The link with optimality theory, on the one hand, and neural studies of decision making, on the other, has led to models in which the decision bounds collapse over time. In this model, less evidence is required to trigger a decision as time passes. Another variant, which has a similar effect, has fixed boundaries, but uses an "urgency signal" added to the accumulated evidence.

Recent work has shown that learning effects can be accounted for by integrating the DDM with reinforcement learning models (Pedersen, Frank, and Biele 2017). In such models, reward expectations are computed and dynamically updated for each of the options using a reinforcement learning scheme, while the DDM is the choice mechanism – with the drift rate being dependent on the difference in reward expectation for the two options.

It has also long been known that the random walk of the DDM can be easily related to the Sequential Probability Ratio test (Bogacz et al. 2006), a procedure that makes statistically optimal decisions when evidence is accumulated in time. The strict mathematical equivalence between the DDM and a Bayesian model has recently been explicitly derived (Bitzer et al. 2014). Other extensions have been proposed to account for longer, more complex decisions between more than two options (Roe, Busemeyer, and Townsend 2001).

For a review of DDM models to investigations of clinical disorders and individual differences, see White et al. (2010a) and White et al (2016).

2.2.2 Accumulation of evidence in biological neurons

Whether the brain uses diffusion-like algorithms is a matter of significant interest and contention. In a pioneering series of studies, Michael Shadlen, Bill Newsome and collaborators observed that neurons in the lateral intra-parietal sulcus (LIP) of macaque monkeys behaved very similarly to what one would expect if they implemented a diffusion process (Shadlen and Newsome 2001). These researchers used a stochastic motion discrimination task where moving stimuli were shown to a monkey and the monkey had to indicate whether the motion was left or right. The experimenter could control the amount of motion (the "evidence") on a single trial. They found that LIP neurons had a

mean spike rate that ramped up for choices that result in an eye movement into their response field (RF) and ramped down for choices out of their RF. The level to which the neuron's activity ramped up before leading to a saccadic response seemed fixed, mirroring the boundary of diffusion process. Moreover, the slope of the ramp was steeper for easier trials, mirroring the drift rate of the model. Since then, other cortical and subcortical regions have also been found to also exhibit possible correlates of a diffusion process. How evidence accumulation is implemented in real neural circuits is still debated, however. This issue has led to great theoretical advances, such as described in **Chapter 3.**

2.3 Reinforcement learning models[4]

In machine learning, Reinforcement Learning concerns the study of learned optimal control, primarily in multi-step (sequential) decision problems (Sutton and Barto 1998). Most classic work on this subject concerns a class of tasks known as Markov decision processes (MDPs). MDPs are formal models of multi-step decision tasks, such as navigating in a maze, or games such as Tetris (**Figure 2.3**). The goal of RL is typically to learn, by trial and error, to make optimal choices.

< Figure 2.3 around here>

Formally, MDPs are expressed in terms of discrete states *s*, actions *a*, and numeric rewards *r*. Informally, states are like situations in a task (e.g., locations in a spatial maze), actions are like behavioral choices (turn left or right), and rewards are a measure of the utility obtained in some state (e.g., a high value for food obtained at some location, if one is hungry, or money).

An MDP consists of a series of discrete time steps, in which the agent observes some state s_t of the environment, receives some reward r_t, and chooses some action a_t. The agent's goal is to choose actions at each step so as to maximize the expected cumulative future rewards. Future rewards are usually penalized by how far in the future they would be received (to account for the intuitive idea that a reward

[4] This section is partly inspired from Gershman and Daw (2017).

obtained in the near future is more attractive than the same reward in the far future). This delay discounting is usually implemented by applying a decay factor γ <1: the expected cumulative future rewards is then defined as the sum $r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots$ of future rewards. Thus, the goal is to maximize not the immediate reward of an action but instead the cumulative reward (a.k.a. the "return"), summed over all future time steps. Each action not only affects the current reward but, by affecting the next state, also sets the stage for subsequent rewards. As a consequence, choosing optimally can be quite complicated. What makes these problems nevertheless tractable is the characteristic property of MDPs, the Markov conditional independence property: At any time-step t, all future states and rewards depend only on the current state and action. This means that conditional on the present state and action, all future events are independent of all preceding events.

To solve such problem, we can compute the "value" of each state. The state value can be written in terms of the sum of future expected rewards and, thanks to the Markov property, has a recursive mathematical expression:

$$V(s_t) = E[r_t + \gamma r_{t+1} + \gamma^2 r_{t+1} + \cdots | s_t] = P(r|s_t) + \gamma \sum_{s_{\{t+1\}}} P(s_{t+1}|s_t) \quad V(s_{t+1}) \quad (9)$$

Equation 9 is a form of the so-called Bellman equation, versions of which underlie most classical RL algorithms. Here, it says that the expected future reward in state s_t is given by the sum of two terms: the current reward and the second term, which stands in for all the remaining rewards $r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots$ The insight is that this sum is itself just the value V of the subsequent state, averaged over possible successor states, according to their probability. If we manage to learn the values of all the states in the environment (see below), we can choose our actions so as to move towards the most promising ones. The agent will use the value function to select which state to choose at each step: taking the step with the highest value. This is called value-based reinforcement learning.

A common alternative, called policy-based reinforcement learning, is to directly compute the value of taking any action a_t in each state s_t. This is called the state-action value function $Q_\pi(s_t, a_t)$ and is the quantity we will want to optimize. This equation has the same form as before:

$$Q_\pi(s_t, a_t) = r_t + \gamma \sum_{s_{\{t+1\}}} P(s_{t+1}|s_t, a_t) Q_\pi(s_{t+1}, \pi(s_{t+1})) \quad (10)$$

The function $\pi(s_{t+1})$ is called the policy and denotes the way by which the agent choses which action to perform in a given state. It takes the current environment state to return an action. It can be either deterministic or probabilistic.

2.3.1 Learning the V or Q values

If we can get a good estimate of the V values, we can choose the best action simply by taking steps that will move to the state with highest value. Similarly, if we have an estimate of the Q value function, we can choose the best action simply by comparing Q values across candidate actions. Many RL algorithms rely on variations on this basic logic.

How do we learn those values, though? There are two main classes of algorithms for RL based on Equation 9. These classes focus on either the left- or right-hand side of the equal sign in that equation (Gershman and Daw 2017).

The first approach (focusing on the right side of equation 9) is known as model-based reinforcement learning due to its reliance on learning the probabilistic internal model, i.e. the one-step reward and state transition distributions $P(r_t|s_t)$ and $P(s_{t+1}|\ s_t,a_t)$. Because these transitions concern only immediate events, i.e. which rewards or states directly follow other states, they can be learnt easily from local experience, essentially by counting. Given these probabilities, it is possible to iteratively expand the right-hand side of Equation 9 to compute the state-action value for any state and possible action. Algorithms for doing this, such as value iteration, essentially work by simulation: by listing the possible sequences of states that can follow a starting state and action, summing the rewards expected along these sequences, and using the learned model to keep track of their probability. The main advantage of model-based learning is in its simplicity. However, this simplicity comes with a cost of computational complexity because producing the state-action values depends on extensive computation over many branching possible paths.

The second class of algorithms is called model-free reinforcement learning. These algorithms avoid learning the internal model (the transition and reward probabilities). Instead, they learn a table of state-action values Q (the left-hand side of Equation 9) directly from experience and sampling the environment.

The discovery of such algorithms, - in particular, the family of temporal-difference (TD) learning algorithms (Sutton 1988) - was a major advance in machine learning and continues to provide the foundation for modern applications.

Briefly, these algorithms use experienced states, actions and rewards to approximate the right-hand side of Equation 9 and average these to update a table of long-run reward predictions. More precisely, many algorithms are based on a quantity called the reward prediction error δ_t. This quantity corresponds to the comparison between the value $V(s_t)$ (the predicted reward) and the actual reward plus the prediction computed one time-step later:

$$\delta_t = r_t + \gamma V(s_{t+1}) - V(s_t) \qquad (11)$$

The expression is similar if we are learning the Q values: $\delta_t = r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$. When the value function is well estimated, this difference should on average be zero. If the values are incorrect, however, there will be a discrepancy between the two sides of the equation. In that case, the stored values are updated iteratively to reduce the discrepancy:

$$V_{t+1}(s_t) = V_t(s_t) + \alpha\delta_t = V_t(s_t) + \alpha(r_t + \gamma V(s_{t+1}) - V(s_t)) \qquad (12)$$

where α is a learning rate between 0 and 1. This is known as the TD algorithm.

Decisions under model-free models are much simpler than using model-based algorithms because the long-run values are pre-computed and need only be compared to find the best action. However, this computational simplicity comes at the cost of inflexibility and less efficient learning.

2.3.2 Reinforcement Learning in the Brain

A most-celebrated success of linking theory and neuroscience was the observation that the firing of dopamine neurons in the midbrain of monkeys resembles the reward prediction error of Equation 11, when the monkey are engaged in a reward learning task. This suggests that the brain uses this signal for reinforcement learning (Montague, Dayan, and Sejnowski 1996). The trial–trial fluctuations in this signal track the model quite precisely and can also be measured in rodents using both physiology and

voltammetry. A similar signal can also be measured in the ventral striatum (an important target of the dopamine neurons) in humans using fMRI. Many researchers believe that dopamine drives learning about actions by modulating plasticity at its targets, for e.g. in the striatum. Elicitation and suppression of dopaminergic responses have been shown to modulate learning in tasks specifically designed to isolate error-driven learning.

The link between dopamine and prediction error has important consequences for understanding mental illness and maladaptive behaviors such as addiction. As we will see in **Chapter 9**, for example, drugs of abuse invariably agonize dopamine neurons. This suggests that some aspects of drug abuse and addiction could be understood in terms of the drugs hijacking reinforcement learning processes by interfering with prediction error signals, giving increasingly higher values to actions leading to the drug.

2.3.3 Evidence for model-based and model-free systems

How can we assess whether or when the brain is using model-based or model-free learning?

Although model-free and model-based algorithms both ultimately converge to the optimal value predictions (under various technical assumptions), they differ in the trial-by-trial dynamics by which they approach the solution. Evidence for one or the other model can be shown in experimental tasks that use staged sequences of experience ordered in such a way so as to defeat a model-free learner. For example, in latent learning or 'sensory preconditioning' tasks, animals are first pre-exposed to an environment that does not have any reward (e.g., by exploring a maze). Later, rewards are introduced at particular locations. For a model-based learner, this experience results in them first learning the transition function $P(s_{t+1}| s_t ,a_t)$, i.e., the map of the maze, and then, subsequently, the reward function $P(r_t|s_t)$ which they will incorporate to their model. However, for a model-free learner, the pre-exposure stage does not teach them anything useful (only that Q values are zero everywhere). They will not learn a representation of the map of the maze (the state transition distribution). Because of this, when rewards are introduced, they must re-learn the navigation task from scratch.

There is some evidence for model-free learning in animal behavior. As the theory predicts, under certain circumstances, animals fail to integrate information about contingencies and rewards if both types of information have been learned separately. For instance, following overtraining on lever pressing for food, rodents will press the lever even after being satiated – despite satiation corresponding to a devaluation in the outcome. However, less thoroughly trained animals can successfully adjust. In

general, experiments looking at how animals adjust their decisions following changes in reward value (e.g., outcome devaluation) or task contingencies show that their behavior cannot be entirely accounted for with model-free RL.

In psychology, these two sorts of behaviors (incapable and capable of integration, respectively) are known as habitual and goal-directed behaviors. The predictions of model-free learning and the prediction error theories of dopamine are well matched to habitual behavior but fail to account for goal-directed behavior and the ability of organisms to integrate experiences. It is thought that model-based learning operates alongside the model-free system and that both systems compete to control behavioral output (Daw, Niv, and Dayan 2005). Little is known about how the brain determines which of these systems controls behavior at one moment in time. Various models have been proposed to govern arbitration between MB and MF values – for instance according to their relative certainties (which vary with the degree of learning and computational inefficiencies; Daw et al., 2005), or the opportunity cost of the time that it takes to perform model-based calculations (Keramati et al. 2011; Pezzulo et al. 2013). Lee et al (2014) for example proposed an arbitration mechanism that allocates the degree of control over behavior by model-based and model-free systems as a function of the reliability of their respective predictions (**Figure 2.4;** see **Section 5.2.2** in this volume).

< insert Figure 2.4 here>

The neural circuits supporting putatively model-based behavior are not well understood. Human neuroimaging suggests that there might be more overlap between neural signals associated with model-based and model-free learning than initially expected.

2.3.4 Implications for Psychiatry

The distinction between model-based learning and model-free learning appear particularly relevant for Psychiatry. It has been proposed in particular that addictive and compulsive disorders might involve a shift from model-based to model-free decision-making, which would explain inflexible behavior in

patients.

Daw et al. (2011) designed a task (**Figure 2.5**) to measure the trade-off between the two types of learning within an individual. This task has since been examined extensively, with some supporting evidence for an association between deficits in goal-directed control and compulsive behavior (Gillan et al. 2016).

< Figure 2.5 around here>

More generally, deficits in learning could be at the core of the issues observed in mental illness. Learning and decision-making are highly intertwined processes. If learning mechanisms are impaired, maladaptive decisions will be taken, which in turn will influence what will be learn.

The idea that patients with mental illnesses operate with a "wrong" internal model of the world is one that is also central to the Bayesian approach, which we discuss next.

2.4 Bayesian Models and Predictive Coding[5]

2.4.1 Uncertainty and the Bayesian Approach.

Bayesian approaches focus on the idea that we live in world of uncertainty. Our environment is often ambiguous or noisy, and our sensory receptors are limited. Often, multiple interpretations are possible. In this context, the best our brain can do is to try to guess what is happening in the world and what best action to take.

This idea of the brain as a 'guessing machine' has been formalized in recent years taking ideas from machine learning and statistics. It is proposed that the brain works by constantly forming hypotheses or 'beliefs' about what is present in the world and the actions to take, and by evaluating those hypotheses based on current evidence and prior knowledge. Those hypotheses can be described mathematically as conditional probabilities, denoted P(hypothesis | data): the probability of the hypothesis given the data, where 'data' represents the signals available to our

[5] The section on Bayesian models is based on Seriès and Sprevak (2014).

senses. Statisticians have shown that the best way to compute those probabilities is to use Bayes' rule, named after Thomas Bayes (1701–1761):

$$p(hypothesis|data) = \frac{p(data|hypothesis)p(hypothesis)}{p(data)} \qquad (13)$$

Bayes' rule is of fundamental importance in statistics. Using Bayes' rule to update beliefs is called Bayesian inference. For example, suppose you are trying to figure out whether it is going to rain today. The data available might be the dark clouds that you can observe by the window. Bayes' rule states that we can update our belief, the probability P(hypothesis | data), which we call the *posterior probability*, by multiplying two other probabilities:

- P(data | hypothesis): our knowledge about the probability of the data given the hypothesis (e.g. 'how probable is it that the clouds look the way they do now, when you actually know it is going to rain?'), which is called the *likelihood*, times:
- P(hypothesis): called the *prior* probability, which represents our knowledge about the hypothesis before we collect any new information, here for example the probability that it is going to rain in a day, independently of the shape of the cloud, a number which would be very different whether you live in Edinburgh or Los Angeles.

The denominator, P(data), ensures the resulting probability is comprised between 0 and 1. This posterior becomes our new prior belief and can be further updated based on new sensory input. In a perceptual context, a hypothesis could be about the presence of a given object, or about the value of a given stimulus, while the data consists in the noisy available inputs.

The critical assumptions about Bayesian inference as a model of how the brain works are:

- The uncertainty of the environment is taken into account and manipulated in the brain by always keeping track of the probabilities of the different possible interpretations;
- The brain has developed (through development and experience) an internal model of the world in the form of prior beliefs and likelihoods that can be consulted to predict and interpret new situations;
- The brain combines new evidence with prior beliefs in a principled way, through the application of Bayes' rule (or an approximation);

- Because currently developed intelligent machines also work in that way — learning from data to make sense of their noisy or ambiguous inputs and updating beliefs — we can get inspiration from machine learning algorithms to understand how this could be implemented in the brain.

2.4.1 Testing Bayesian predictions experimentally

Bayesian inference as a model of cognition makes predictions that can be tested using behavioral experiments. This has been the aim of a lot of research in the last twenty years. The first line of research focused on multisensory integration, i.e. how the brain combines information coming from different sensory modalities, such as vision and sound. Bayesian inference makes clear predictions about how this should be done: the individual information sources to be integrated should be weighted according to their reliabilities. It also predicts that the combined estimate will then be more reliable than any estimate based on a single one of the sensory cues. For example, if the visual information is much clearer than the auditory information, it should have much more influence on your experience. This can lead to sensory illusions, in situation where there is a conflict between the two modalities and one modality is much more reliable than the other (as observed in the ventriloquism illusion, or the McGurk effect). When Bayesian model predictions are compared to experimental data, the general finding is that human behavior is well approximated by Bayesian integration.

The Bayesian model not only predicts how simultaneous signals should optimally be combined, but also how to include prior knowledge. According to Bayes' rule, such knowledge can be represented as a prior probability, which would serve as a summary of all previous experience, and which should be multiplied with the incoming information, the likelihood. An important line of research aims at understanding which priors the brain is using and how such priors impact perception, action and cognition (see e.g., Seriès and Seitz 2013). A good way to discover the brain's expectations or assumptions is to study perception or cognition in situations of strong uncertainty or ambiguity - where the current sensory inputs or the 'evidence' is very limited. Studying such situations reveals that our brains make automatic assumptions all the time. Sensory illusions have proved particularly important in this field. For example, looking at how the brain interprets shaded objects reveals that the brain assumes that light comes 'from above'. This makes sense of course, since light usually comes from the sun, above us. Similarly, we seem to expect objects to be symmetrical, to change smoothly in space and time,

orientations to be more frequently horizontal or vertical and angles to look like perpendicular corners. We also expect objects to bulge outward more than inward (i.e. to be convex shapes, like balloons or pears), that background images are colored in a uniform way, that objects move slowly or not at all, that the gaze of other people is directed towards us, and that faces correspond to convex surfaces. Such assumptions have been successfully described using the framework of Bayesian priors.

Techniques have been developed that allow measurement of individual participants' priors based on experiment measurements and performance biases, by fitting Bayesian models to performance (Stocker and Simoncelli 2006). In the sensory domain, it is commonly found that human participants learn quickly effortlessly and unconsciously the statistics of the perceptual environment and come to expect the perceptual inputs that are most likely. This can lead to biases in the estimation of sensory features, perceiving the world as being more similar to what is expected than it really is, and sometimes even "hallucinating" expected inputs, even when they are absent (Chalk, Seitz, and Seriès 2010). It has also been shown that the brain can update 'long-term' prior beliefs, such as that light comes from above or that objects move slowly, if placed in environments where lights come from below or where objects move quickly (Seriès and Seitz 2013). This shows that the brain constantly revises its assumptions and updates its internal model of the environment.

Many aspects of human cognition, such as language acquisition and processing, action selection, prediction, reasoning and sensory inference have been modeled as optimal readouts of statistical inference processes.

2.4.3 Decision Theory

Bayes rule computes beliefs about the state of the world *s* given noisy or ambiguous sensory input *D*, *P(s|D)*. However, it does not specify how these beliefs are used to generate decisions and actions. Decision theory extends Bayesian inference to deal with the problem of selecting the best decision or action based on our current beliefs and as such, encompasses the methods of reinforcement learning described above. The difference of course is that here we have to infer the states, instead of them being given as commonly assumed in RL formalisms. The essence for making the best decisions is then to minimize the expected loss (or maximizes expected reward/utility) given our beliefs. One simply calculates the expected loss for a given action, that is the loss averaged across the possible states weighted by the degree of belief in the state: $\sum_s L(a,s)P(s|D)$ and then chooses the action that has the

smallest expected loss.

Here, *L(a,s)* denotes the benefit or loss associated with taking action *a* in state *s*, and Σ denotes a summation over all possible states s.

We can also consider domains with temporal dynamics, where s(t) changes over time, with each action evoking its own set of stochastic transitions. Here, it is necessary to determine not an individual optimal action, but rather a sequence of actions in the light of possible transitions.

Bayesian Decision Theory (BDT) therefore requires two sorts of inference: one is to compute the posteriors to estimate the states as well as possible (using Bayesian inference), and the other is to compute the best action. Both are computationally hard; the latter is particularly difficult when it is necessary to optimize over long trajectories of future actions—and becomes much harder in the face of the first problem. When the state s is known, BDT reduces to common descriptions of reinforcement learning (see **Section 2.3**).

2.4.4. Heuristics and approximations, implementation in the brain

Exact Bayesian inference is thought to be intractable for most everyday problems that the brain encounters. An important line of research tries to understand whether human behavior can be described using simple heuristics that might approximate Bayesian inference, without involving complex computations.

While in theory it seems feasible for neural circuits to implement (possibly rough) approximations of Bayesian computations, how such computations are actually implemented in the brain and relate to neural activity is still an open question and an active area of research. Whether the Bayesian approach can actually make testable predictions for neurobiology (for e.g., which parts of the brain would be involved, or how neural activity could represent probabilities) is also debated. It is yet unclear whether the Bayesian approach is only useful at the 'computational' level, to describe the computations performed by the brain overall, or whether it can be also useful at the 'implementation level' to constrain and predict how those algorithms might be implemented in the neural tissue.

2.4.5 Application to Psychiatry

Many researchers believe that the Bayesian approach has promising application for the field of Psychiatry. Bayesian models could potentially help quantifying differences between different groups (e.g. healthy vs. ill) and identifying whether such differences come from using different internal models, for example different prior beliefs, or from different learning or decision strategies.

A common idea in psychiatry is that the internal models used by patients, in particular their prior beliefs, could be different from those of healthy subjects. In the study of schizophrenia, for example, it has been proposed that 'positive symptoms' (hallucination and delusions) could be related to an imbalance between information coming from the senses and prior beliefs or expectations (see **Chapter 6**). In autism, similarly, it has been proposed that the influence of prior expectations might be weaker compared to that of sensory inputs, which could explain that patients feel overwhelmed by a world perceived as being 'too real' (see **Section 11.1**).

More generally, it has been proposed that three different classes of failure modes could be at the root of mental illness. They stem from either: 1) abnormalities in the framing of problems or tasks that the brain is trying to solve (abnormalities in the priors, likelihood or utility), or 2) from the mechanisms of cognition used to solve the tasks, or 3) from the historical data available from the environment, i.e. abnormal experience for e.g. trauma (Huys et al. 2015).

2.4.6 Predictive Coding and Bayesian models used in Psychiatry

Despite the popularity of these ideas, in practice, only a limited number of studies have tried to compare or fit the behavior of participants with quantitative models. We here describe some of the models that can be found in the literature.

a) Predictive Coding models

Predictive coding became popular as a model of visual processing at the turn of the century (Rao and Ballard 1999). The general idea is to view visual processing as a hierarchical system, composed of a number of levels connected by feed-forward and feedback connections. In this system, feedback

projections from one level to the lower level are trying to predict the activity of the neurons they target: feedback connections carry predictions of lower-level neural activities, whereas the feed-forward connections carry the residual errors between the predictions and the actual lower-level activities (**Figure 2.6**). It was shown that this model could explain a number of phenomena observed in real neural activities in the visual cortex. The idea that the brain uses some form of predictive coding has become very widespread. There is however a debate about how such a scheme would be implemented in neural activities, in particular about whether neural activities in the visual cortex should really be interpreted in terms of prediction errors (as proposed by Rao & Ballard 1999), or whether they would be better understood in terms of the predicted input itself – or maybe whether there would be different categories of neurons representing either the prediction error, or the predicted input.

<Figure 2.6 around here>

Predictive coding and Bayesian inference are concepts that are often confounded. Although predictive coding and Bayesian inference do not necessarily imply each other (Aitchison and Lengyel 2017), predictive coding is often proposed as an effective way for Bayesian inference to be implemented in the brain: while Bayesian inference would describe the general computation that the brain is trying to perform, predictive coding would describe the algorithm that is being carried out.

Why the two concepts are related can be understood as follows. As explained above, Bayesian inference entails updating our existing belief (the prior distribution) with new information (the likelihood distribution) to form our new belief (the posterior distribution). If we assume those distributions are Gaussian, they can each be represented by their mean μ and their variance, or precision π (where π refers to the inverse variance). It can be shown that the Bayesian sequential updating of beliefs can be expressed as follows (C. Mathys et al. 2011; Palmer, Lawson, and Hohwy 2017), where x is the new measurement, $\mu_{posterior}$ is the mean of the new posterior and $\pi_{posterior}$ its precision:

$$\mu_{posterior} = \mu_{prior} + \frac{\pi_{likelihood}}{\pi_{posterior}}(x - \mu_{prior}) \qquad (14)$$

Where:

$$\pi_{posterior} = \pi_{prior} + \pi_{likelihood} \quad (15)$$

This last term of Eq. 14, (x- μ_{prior}), can be interpreted as a prediction error: the mean of the prior belief (prior), $\mu_{prior,}$ can be considered a prediction about what the new measurement, *x*, will be. This means that Bayesian inference can be implemented by iteratively updating predictions with the prediction error produced by each new measurement.

The precision of the prior distribution π_{prior} indicates our confidence in our existing prediction, while the precision of the likelihood distribution $\pi_{likelihood}$ represents the ambiguity inherent in the measurement (the noisiness of incoming data). Together, these two parameters give an indication of how reliable or informative prediction errors are expected to be regarding the true (hidden) state of the world. Prediction errors are therefore weighted by the estimated precision of the new information relative to the estimated precision of existing beliefs.

The weighting term ($\pi_{likelihood}$/ $\pi_{posterior}$) plays the role of a learning rate. A high learning rate means that prediction errors will drive inference about the state of the world to a greater extent. Conversely, a low learning rate means that prior information is given more weight in determining what is inferred. The fact that the learning rate depends on the ratio $\pi_{likelihood}$/ (π_{prior+} $\pi_{likelihood)}$ implies that beliefs are more highly sensitive to new measurements when we know little about the environment (π_{prior} is small) but less sensitive when we have already gathered plenty of information (π_{prior} is large).

Importantly this update expression – that links Bayesian inference with predictive coding - is not specific to the univariate Gaussian case, but can be shown to be valid much more generally[6].

b) Hierarchical Gaussian Filter model

Inference in realistic environments is thought to be hierarchical, involving different levels of predictions, described by random variables, which interact with each other. The set of probabilistic steps that can be followed to generate the values of these random variables is known as the generative model. Often, we represent these steps using a graph representation. In such a graphical model, the nodes represent the random variables, and the edges represent condition dependencies (see e.g. **Figure 2.6B**).

[6] to Bayesian updates for all exponential families of likelihood distributions with conjugate prior.

While Bayesian belief updating in such generative models is optimal from the point of view of probability theory, it is difficult to achieve in practice: it requires computing complicated integrals which are not tractable analytically and difficult to evaluate in real time. For this reason, it is thought that the brain can only achieve approximations of Bayesian inference. Different types of approximations are usually considered, inspired from research in Machine Learning. A popular model recently developed is the hierarchical Gaussian filter (HGF).

The Hierarchical Gaussian Filter model (C. Mathys et al. 2011; C. D. Mathys et al. 2014) describes a hierarchical generative model of the environment and its (in)stability. In this model, all states except the lowest level evolve as coupled Gaussian random walks[7], such that each state determines the step size of the evolution of the next lower state.

For example, imagine a task where participants have to perform a binary classification of images as either faces or houses, where the images had high, medium or no noise added (**Figure 2.6A**). A tone preceding each image is highly, weakly or not predictive of a given outcome, and the associations between images and tones change across time. Such a task can be represented by the graphical model depicted in **Figure 2.6B**, where the lowest level variable x_1 describes the uncertainty about outcomes, i.e the presence of a house or face, level 2 (x_2) addresses uncertainty about the cue-outcome contingencies, and level 3 (x_3) addresses uncertainty about environmental change, i.e. the volatility of the cue-outcome contingencies.

Using the so-called "mean-field" approximation, Chris Mathys and collaborators derived analytic update equations for beliefs at each level, whose form resembles RL updates and the equation (14) above, with dynamic learning rates and precision-weighted prediction errors. The update equations make the model well suited for filtering purposes, i.e. they can be used to predict the value of, and the uncertainty about, a hidden and moving quantity based on all information acquired up to a certain point. Importantly, the coupling across levels is controlled by parameters whose values can be fit to each individual participants performing the task.

The HGF model has been used for example to investigate how participants with autism learn about changing environments (Lawson, Mathys, and Rees 2017). They used the task described above where

[7] A Gaussian random walk is a random walk that has a step size that varies according to a normal distribution.

participants performed binary classification of images as either faces or houses. By fitting the HGF model to their data, Lawson et al could show that participants with autism tended to overestimate the volatility of the sensory environment, at the expense of building stable expectations that would lead to be surprised when aberrant outcomes arise.

Chapter 6 will provide another example for the use of the HGF in schizophrenia research.

c) Belief networks and circular inference[8]

An alternative, very general, powerful and efficient algorithm to perform inference in generative models is known as belief propagation**.** Consider for example a hierarchical generative model with 3 nodes: the "leaf" is caused by a "tree", which is caused by a "forest". In belief propagation, sensory information S, e.g. the probability that a leaf is present in the image, climbs the hierarchy in a feedforward way and at the same time, prior information moves downward as feedback. Each node calculates a belief for the underlying variable it represents, equivalent to the posterior e.g. $P(X_{tree}|S)$ and sends local messages (e.g. $M_{tree->leaf} = P(X_{leaf}|X_{tree})$) to all the neighboring nodes. As a result, information, in the form of beliefs, is propagated throughout the system. Assuming binary variables and using the log-ratios of the probabilities, then beliefs and messages can be calculated by the recursive equations of the form:

$$M_{ij}^{t+1} = W_{ij}(B_i^t - M_{ij}^t) \quad (16)$$

$$B_i^{t+1} = \sum_j M_{ji}^{t+1} \quad (17)$$

where M_{ij}^t is the message from node i to node j at time t, B_i^t is the belief of node i at time t and $W_{ij}(B)$ is a sigmoid function of B.

The second equation simply means that each node calculates a belief by summing the messages coming from all its connected neighbors (e.g., the belief about the presence of a tree is equal to the sum of the messages from the forest and the leaf nodes). The first equation, on the other hand, means that the message travelling from node *i* (here, the forest) to node *j* (the tree) is a function of the belief of the sending node *i* after we subtract the effect that the receiving node *j* has on the sending node *i* (e.g., here, the message from tree to forest). This latter correction is crucial. Without it, the algorithm would produce loops, i.e. reverberations of bottom-up and or top-down information. In such "loopy" belief propagation, the consequences are treated as causes and vice versa and the information in the upward

[8] This section is based on Leptourgos , Denève and Jardri 2017.

and the downward stream can be mixed and over-counted. Jardri & Denève have proposed that such "circular inference" could underlie the symptoms of schizophrenia and may also be present to some extent in the general population (Jardri and Denève 2013; Jardri et al. 2017).

2.5 Model Fitting and Model Comparison

Having provided a brief overview over different modeling techniques, we now turn to a tutorial overview of how these techniques can be used to probe behavior and fit to real data. We mostly focus on the computational methods that use a generative framework, namely reinforcement learning models and Bayesian models. The following uses material from Huys (2017).

2.5.1 Choosing a suitable model

Assessing whether a given model is a suitable description of the data at hand can be tricky. In theory, there will always be many other types of models that could be suitable as well. So how do we choose and validate a particular one? As a rule of thumb for good practice, the modeling should contain three general steps. We first need to build the model. Second, this model should be validated with artificial data. Finally, the model is applied to the real data. These points are detailed below:

1. **Clarifying the hypotheses to be tested**. The initial choice of the model is usually motivated by the hypotheses that we wish to test. We will usually be guided by an effort to: i) have a model that is flexible enough to describe the data and relates to previous literature in the field; ii) contains parameters that directly relate to our hypotheses; iii) is as simple as possible given those constraints. There will usually be different possible variants of the model. A reasonable approach is to build a series of models starting from a very simple 'null' hypothesis (a "no–interest" model that does not include the element we wish to show the importance of) and then adding in the various features of interest to examine to what extent they contribute towards explaining the data. A probabilistic component will need to be included, so as to account for the variability intrinsic to each individual's performance. The different variants can be tested against each other

using model comparison (see below).

2. **Validation on artificial data** means using the model to generate artificial data, by setting the parameters by hand and exploring the different behaviors exhibited by the model. First, this is a way to check that the data the model generates is actually comparable to the data obtained in the experiment. Second, this can be used to test the fitting procedure: once the parameters have been chosen by hand, and the artificial data has been generated, we can try to recover the parameters using our fitting procedure (i.e. inverting the model): can we discover which parameters were used for the model just by looking at the model's performance? This step is called parameter recovery. This is an important step prior to interpreting any parameters. This can be used as well to ask whether we can distinguish between the behavior generated by different models, and whether we can recover a particular model reliably. This is called model recovery. It is recommended to attempt to perform these steps prior to collecting the experiment data as they may suggest changes in experimental parameters, such as the number of trials or the number of subjects to run.

3. Finally, the models need to also be validated on the actual data of interest. One possibility is to compare data generated from the model (with fitted parameters) to the real data. For learning experiments, it is for instance often useful to plot learning curves and ask whether the model captures the shape of these curves well. Once the models have been validated in this way, it is meaningful to ask which of the models provides the most parsimonious account of the data. This is the domain of model comparison, where the performances of different models are weighted against their number of free parameters. Model comparison is always relative: even the best amongst a set of models may still be too poor to provide any meaningful information. The interpretation of parameters in the models should only follow at the end, once one model has been chosen as a satisfactory characterization of the data.

2.5.2 A Toy Example

To illustrate this process, we can consider a very simple learning experiment (Huys, 2017). On each trial, participants have to choose one of two squares. The blue square yields small rewards on 80% of trials, and the red square on 20% of trials. Participants have to discover which of the two squares is best,

based on their successive choices. On each trial t, they thus perform a single choice a_t, which yields an immediate reward r_t. This choice does not have any influence on future options.

We can consider two different models. The first model assumes that individuals perform temporal difference learning to compute the values of the two stimuli in this extremely simple scenario. Taking equation 12 and observing that there is no next state, but only immediate rewards, the temporal difference prediction error learning takes the simpler form of Rescorla-Wagner learning (Rescorla and Wagner 1972):

$$V_{t+1}^{TD}(s_t) = V_t^{TD}(s_t) + \alpha(r_t - V_t^{TD}(s_t)) \quad (18)$$

The second model assumes that individuals simply perform averages over the rewards earned for each of the two stimuli. This model is actually the correct inference to perform given how the outcomes are generated.

$$V_{t+1}^{av}(s_t) = \frac{1}{t}\sum_t^1 r_{t'} \quad (19)$$

It can be easily shown that this equation can also be expressed in this recurrent form:

$$V_{t+1}^{av}(s_t) = V_t^{av}(s_t) + \frac{1}{t}(r_t - V_t^{av}(s_t)) \quad (20)$$

Comparing these expressions, we see that while the TD learning rule uses a fixed learning rate α, the average has a decaying term 1/t. The TD rule has one free parameter: α, while the averaging rule has no free parameter. How can we determine which model best accounts for participants' performance?

a) Generating data

We first start by generating artificial data from both models. To do this, we need to determine a model for the function that maps the values V onto probabilities of choosing one action or the other (here, choosing one square or the other). A frequent choice is the use of a softmax function whereby the probability of choosing stimulus s on trial t (e.g. the blue square) is:

$$p(a_t = s|V_t) = \frac{e^{\beta V_t(s)}}{e^{\beta V_t(s)} + e^{\beta V_t(\bar{s})}} \quad (21)$$

where $\bar{s}$ denotes the alternative stimulus (i.e. the green square) and β determines how precisely the choices follow the values, i.e. also how noisy the choice process is. This parameter can also be interpreted as controlling exploration vs exploitation.

b) Fitting models

Once we have built a model and generated artificial data from it, we can proceed to the next step: fit the model to the generated data to assess how well we are able to recover the model's parameters. To find the set of parameters that are most compatible with the data, we can use maximum likelihood (ML). To find the ML parameters, for each subject, we look for the parameters θ (in our example, $\theta = \{\alpha, \beta\}$) that maximize the likelihood of all their T actions $a_{1;} \ldots a_T$:

$$\hat{\theta}_{ML} = argmax_\theta \log p(a_1, a_2 .., a_T|\theta) \quad (22)$$

On first sight, this calculation may appear difficult because the choices a_t depend on previous choices. However, since every choice only depends on the value V_t at the time of the choice t, then the probability of observing a sequence of stimulus choices $a_{1, ..}, a_T$ is simply:

$$\log p(a_1, a_2, .. a_T|\theta) = \log \prod_{t=1}^{T} p(a_t|V_t) = \sum_{t=1}^{T} \log p(a_t|V_t) \quad (23)$$

This is: once we condition on the values the choices become independent of the previous choices.

The values can be updated iteratively prior to computing the likelihood of each choice, leading to an algorithm that takes this general and very simple form:

- Initialize the values V for each stimulus
- f**oreach** trial t **do**:
  compute log likelihood of choice a_t on trial t given parameters: $l_t = \log p(a_t|V_t, \theta)$
  update value V_{t+1} given outcomes on trial t
  end
- compute total log likelihood $l_T = \sum_1^T l_t$

The total likelihood (a function of α and β) can now be passed to any of a number of optimization tools to solve Equation 22.

Parameter recovery using ML is however often very imperfect. This is particularly true in situations where parameters have overlapping effects and therefore can trade off each other. A very simple and often very powerful solution is to impose a soft prior on the parameters and performing maximum a posteriori (MAP) inference rather than ML. This is very simply achieved by replacing equation 22 with

$$\hat{\theta}_{MAP} = argmax_{\theta} \log p(a_1, a_2.., a_T|\theta)\, p(\theta) \quad (24)$$

The computation of the posterior likelihood is thus just the same as before but now we also add the log likelihood of the prior to the total log likelihood of the choices. The choice of the prior p(θ) is not always straightforward. In many situations, it can make sense to infer the prior from the data itself. This is called empirical Bayes. There are a number of techniques available for this, and this is becoming a more common approach. In this toy example, little would be gained over the basic MAP approach, but this would change for larger models (Huys, 2017).

c) Model comparison

Having fitted the model to the data, the next step is to assess how well the model can actually account for the data. Simply look at how closely the model fits the data is not sufficient: a model that is too flexible (has many free parameters) could fit the data perfectly but would lead to poor prediction of new data. This issue is known as over-fitting.

Bayesian model comparison takes into account the trade-off between the flexibility of the model and the fit it provides to the data by using as a measure of fit not the best possible likelihood, but the average likelihood over all possible parameter settings:

$$P(A|M) = \int d\theta P(A|\theta, M)p(\theta) (25)$$

where A denotes the all the behavioral data and M the model.

The Bayes factor, that measures whether model M_1 is more strongly supported by the data under consideration than model M_2, is then defined as:

$$BF = \frac{p(A|M_1)}{p(A|M_2)} \quad (26)$$

and is considered substantial if greater than 3, and conclusive if greater than 5. Unfortunately, the integral in equation 25 is not always straightforward to evaluate, and there exists a number of approximations to it. A commonly used measure is the Bayesian Information Criterion (BIC), defined as:

$$BIC = -2\log p\left(A|\hat{\theta}_{ML}\right) + d\log(n) \quad (27)$$

where *d* is the number of parameters in the model and *n* is the number of data points. Other measures exist, such as the Aikake Information Criterion (AIC), or other related techniques such as using a Laplace approximation for P(A|M) i.e. approximate the function being integrated with a Gaussian, for which the integral can then be computed analytically (see Daw et al 2009).

d) Group studies

The methods so far have considered individual subjects. However, most studies, particularly in clinical settings, deal with group data. Two simple approaches for model fitting exist in this case. First, we can treat all individuals in one group as using the same parameters; this is called a fixed-effects treatment. Alternatively, we can treat them as having entirely separate parameters. This is called a random-effects treatment. A fixed-effects treatment confounds inter- and intra- individual variability and is therefore not recommended. On the other hand, a random-effects treatment can inflate noise depending on how the parameters are estimated. One solution to this is to consider that individuals in a group tend to be similar, and hence should have similar parameters. For instance, parameters of individuals in a group could cluster around a particular value. To implement this idea, we can follow a hierarchical approach and formulate a model about how the parameters vary across the population.

Another question is whether all individuals use the same generative model to do the task, which might not always be the case. Here again, we can either employ a random-effects treatment over models, considerthat some individuals in a group will behave according to model 1, others according to model 2, and yet others according to model 3 etc. This implies that different individuals may differ in terms of the internal processes they invoke to perform a given task. Alternatively, one can nest multiple models in a more complex model. This solution corresponds to assuming that individuals use a mixture of strategies

but that this is true across the entire group. Daw et al (2009) offers a more in-depth treatment of those issues.

2.5 Chapter Summary

A variety of computational tools have been developed that can be applied to psychiatry, either to describe behavior or to try to relate observed behavior to underlying neurobiological differences. The choice of the model will depend on the data to be modeled, the hypothesis that is tested and the questions to be addressed.

- Connectionist models, or neural networks, can be used to explore the relationship between connectivity, dynamics and function. In psychiatry, they have for example been used to explore how attractor dynamics could be impaired in mental illness (see also **Chapter 3**).
- Drift diffusion models can be used to dissect the origin of differences in performance and reaction times between groups, in tasks involving choices between two alternatives.
- Reinforcement learning models are used to model the dynamics of learning of an environment, where discrete states (or objects) are associated with rewards or punishment. Because of the link between prediction error signals and dopamine, reinforcement learning models have shown to be very promising tools to understand impairments in learning and decision making in mental illness (see also **Chapter 5-10**).
- Bayesian models account for learning and decision-making in terms of statistical inference. They can be used to assess how "optimal" a given performance is, and to discover the internal models that participants have learned or use in a particular environment. Because they explicitly model beliefs, they can be used to describe mental illness in terms of maladaptive or broken beliefs and false inference.
- Fitting a particular model to data is usually performed using maximum likelihood or maximum a posteriori. One then needs to verify that the model can account well for the data. Model comparison is used to assess what model describes the data best, taking into account the number of free parameters of each model. Before using them on real data, it is recommended to test the model fitting and comparison techniques on artificial data generated by each model, i.e. to perform parameter and model recovery.

2.6. Further study

Due to space limitations, this chapter could only provide a very quick survey of the methods of computational neuroscience and computational cognitive neuroscience that can be applied to the psychiatry. Each section has been the subject of entire books and review articles. A great reference regarding artificial neural networks is Hertz, Krogh, and Palmer (1991). Dayan and Abbott (2001) is also a recommended reference for further study of biological neurons and of neural networks.

For the Drift decision model, we recommend reviews by Roger Ratcliff, for e.g. Ratcliff et al. (2016). The use of DDM in Psychiatry has been covered for example by White, Curl, and Sloane (2016).

For Reinforcement learning techniques, the classic reference is Sutton and Barto (1998). More recent developments and relation to neuroscience have been covered e.g., by Daw (2009) and Gold and Shadlen (2007).

Bogacz (2017) provides a tutorial on the free-energy framework for modelling perception developed by Friston, which extends the predictive coding model of Rao and Ballard (1999).

Readers particularly interested in modelling fMRI data can also consult Cohen et al. (2017).

Chapter 3: Biophysically Based Neural Circuit Modeling of Working Memory and Decision Making and Related Psychiatric Deficits

Xiao-Jing Wang[1] and John D. Murray[2]

1 Center for Neural Science, New York University, New York, New York 10003, USA

2 Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06510, USA

3.1 Introduction

The brain is not a uniform system made of equal parts. Instead, it is characterized by a modular organization of areas with distinct properties, connection patterns and specialized functions. In the primate cerebral cortex, certain areas like the prefrontal cortex (PFC**)** play a central role in higher cognitive functions, in contrast to early sensory information processing or motor generation. Those areas of the "cognitive type" are the ones commonly implicated in a variety of mental disorders; therefore, understanding such systems is especially relevant to the field of Computational Psychiatry.

A key property of cognitive-type neural circuits is the presence of strong recurrent connections underlying reverberatory network dynamics. The behavior of any nonlinear system endowed with an abundance of feedback connection loops is difficult to predict by intuition alone. To illustrate our point, consider two identical, mutually inhibitory, neurons (Kristan and Katz 2006) (**Figure 3.1A**). Given this "connectome", how would the network behave? It turns out that experiments and theory have uncovered multiple dynamical scenarios. First, both neurons may simply stay silent. Second, when driven by inputs, the system may behave as a "switch", with one neuron active while inhibiting the other neuron, or vice versa, and a brief input can switch the system between the two states (**Figure 3.1B**). Third, if neurons are endowed with a slow adaptation, each of the two neurons could take turn to be active and over time eventually stop firing due to "fatigue" when the other neuron takes over, leading to a "half-center" oscillator which is the core of rhythmic central pattern generators (**Figure 3.1C**). Fourth and

finally, under certain conditions, the two neurons can be perfectly synchronized, spike by spike: the two neurons fire at the same time, leading to mutual inhibition after a brief delay, and when this inhibition decays way they can fire again together (Wang 2010) (**Figure 3.1D**). This simple example illustrates that behavior often cannot be deduced from anatomy in a straightforward fashion; physiology and modeling are important for discovering the dynamical operations of neural circuits.

< Insert Figure 3.1 around here>

In Computational Psychiatry, some researchers are concerned with behavioral performance and its mathematical modeling. For instance, as described later in this book, reinforcement learning models have been applied to addiction, anxiety and depression (see **Chapters 7-9**). Such models can be used to quantify abnormal sensitivity to motivation and reward in affected subjects. However, they are typically relatively abstract and difficult to relate to specific brain circuits in a concrete manner. Another approach, which is the focus of this chapter, strives to develop neural circuit models that are capable of linking a particular behavior to the biological mechanism responsible for generating neural activity patterns that causally underlie an observed behavioral trait.

This is a tall order. One could argue that, at present, we do not yet adequately understand how a neural system generates complex symptom of any psychiatric disorder. But that is precisely why biologically realistic neural circuit modeling should be a priority of our field. Ultimately, a central goal of neuropsychiatric research is to explain how symptoms and cognitive deficits arise from neurobiological pathologies. This demands us to bridge the stark explanatory gaps between levels of analysis: mechanisms underlying a psychiatric disease occur at the level of neurons and synapses, whereas symptoms are manifested and diagnosed at the level of cognition and behavior, which involve collective computations in brain circuits. Linking these levels is vital for gaining mechanistic insight into mental illness, and for the rational development of pharmacological treatments, which act at the molecular level with physiological impact at the synaptic level. Biophysically based neural circuit modeling is a framework particularly well suited to link synaptic-level disruptions to emergent brain dysfunction.

In the following, we will review a set of studies that use biophysically based neural circuit models to understand how synaptic disruptions may induce cognitive deficits, with particular relevance for schizophrenia (Murray et al. 2014; Starc et al. 2017; Lam et al. 2017).

3.2 What is biophysically based neural circuit modeling?

Biophysically based neural circuit modeling incorporates key physiological properties of neurons and synapses, as well as circuit connectivity. Dynamic neural activity is simulated through systems of differential equations governing the biophysical properties of neurons and synapses (see **Section 2.1**). Emergent patterns of activity in the model can be informed by — and tested with — empirical measures of neural activity. In certain circuit models, neural activity can also be mapped onto a behavioral response, thereby generating model predictions that can be tested with behavioral data from corresponding task paradigms.

It is important to emphasize that biological realism does not mean that the more biological details a model incorporates, the better. We typically first formulate a well-defined question, such as "what is the microcircuit mechanism of stimulus-selective persistent activity during working memory?" Then we carefully determine the level of complexity of models for single neurons and synapses as well as network connectivity that are appropriate for investigating that question. For instance, a single neuron can be modeled in detail with morphologically reconstructed dendrite and axon, or using a few compartments so that dendritic compartments are separated from soma, or a single compartment described by the Hodgkin-Huxley model or the integrate-and-fire model (cf. **Section 2.1**). Which one to choose depends on the question under study (e.g. which may or may not require distinct dendritic compartments).

Neurons in a network interact with each other through synaptic connections, which are either excitatory (respectively, inhibitory) if spiking of a (presynaptic) neuron produces an increase (respectively, a decrease) of membrane potential in recipient (postsynaptic) neurons. Synaptic excitation is mediated by AMPA and NMDA receptors that bind with neurotransmitter glutamate, and which have different times constants, NMDA being much slower than AMPA. Synaptic inhibition is mainly mediated by the $GABA_A$ receptor that binds with neurotransmitter GABA. Like single neuron models, synaptic interactions can be described mathematically with varying degrees of complexity. For the sake of simplicity, even to this day, many recurrent neural network models use "kick synapses", namely a presynaptic spike induces an instantaneous jump of postsynaptic potential (which is positive for excitation, negative for inhibition). Therefore, no temporal aspects (latency, rise and decay times, summation) are taken into account. However, the basic dynamical properties of synaptic transmission

can play a crucial role in shaping the collective behavior of a recurrent neural circuit. As described below, for example, slow NMDA receptors at recurrent excitatory synapses have been found to be crucial for the maintenance of mnemonic persistent activity (Wang 1999), a theoretical prediction that years later was supported by a monkey experiment (Wang et al. 2013). Such a discovery suggests that biologically based modeling has a potential to make predictions at the receptor level that bridges with a cognition function via understanding neural circuit dynamics. This may in turn provide opportunities to mechanistically understand how synapse-level disruptions produce aberrant neural activity and deficits in cognition and behavior.

The specific scientific questions under study determine the level of biophysical detail included in a particular model. For instance, questions related to dopaminergic dysregulation (such as found in schizophrenia, see also **Chapter 6 and 10**) can be addressed in a biophysically based model of an individual synapse that includes subcellular signaling pathways (Qi et al. 2010). In contrast, emergent circuit-level dynamics, such as oscillations or persistent activity, can be simulated in thousands of recurrently connected spiking neurons whose individual dynamics are simplified to include only certain channels and receptors (Wang, 2010, 2008). Modeling systems-level disturbances, such as large-scale connectivity alterations in schizophrenia, may entail coarse-grained mean-field models of local nodes organized in large-scale networks. Such models still contain neurophysiologically interpretable parameters and enable study of questions related to Excitatory /Inhibitory (E/I) balance (Yang et al. 2014, 2016a).

An important area of research in clinical neuroscience is the discovery and characterization of predictive neurophysiological biomarkers for psychiatric disorders, i.e. characteristics that can be objectively measured and evaluated as an indicator of pathogenic biological processes.

As one area of modeling progress with relevance to biomarkers, there is a large literature on studying neural oscillations that emerge at the network level in recurrent cortical circuits (Wang 2010). Cortical oscillatory activity is found to be abnormal in a number of neuropsychiatric disorders. In particular, schizophrenia is associated with alterations in oscillatory activity in the gamma (30–80 Hz) range (Gonzalez-Burgos and Lewis 2012; Uhlhaas 2013). Computational models, in conjunction with physiological findings, support the idea that neocortical gamma oscillations arise from a feedback loop in a microcircuit of pyramidal cells reciprocally connected with perisomatic-targeting, parvalbumin-

expressing interneurons (Buzsáki and Wang 2012). These models of gamma oscillations can be used to explore the dynamical effects of putative synaptic perturbations associated with schizophrenia, including reduced production of GABA and parvalbumin[9] in inhibitory interneurons (Vierling-Claassen et al. 2008; Spencer 2009; Volman et al. 2011; Rotaru et al. 2011). In each case, the models provide specific hypotheses for how systems-level dynamics, which can be measured in humans through techniques such as EEG or MEG, may be altered as a result of synaptic- or cellular-level changes.

Below, we focus on how circuit models of cognitive functions can be applied to understand cognitive deficits resulting from synaptic disruptions associated with schizophrenia. For some core cognitive computations, we have knowledge of the neural circuit basis underlying these processes, which typically involve contributions from animal studies. For these cases, detailed circuit models can be developed rigorously to provide the link from synaptic disruptions to behavior (e.g., cognitive deficits discussed below). In other cases, psychiatric symptoms relate to complex cognitive functions for which we lack understanding of the underlying neuronal representations or circuit mechanisms. At present, these circuit models are limited and cannot be applied to complex behavioral tasks, for which we lack understanding of neural circuit correlates. We now turn to the conditions in which circuit models may be best suited to study cognitive deficits in psychiatric disorders.

3.3 Linking propositions for cognitive processes

A major goal in computational psychiatry research is for biophysically based neural circuit models to explain mechanistically how synaptic-level disruptions induce cognitive-level deficits. For this approach to be most effective, the circuit model should be grounded in a well-supported relationship between neuronal activity and a given cognitive process. Such relationships have been formalized by the concept of a linking proposition, which states the nature of a statistical correspondence between a given neural state and a cognitive state. Related to the concept of the linking proposition is that of a bridge locus, which is the set of neurons for which this linking proposition holds (Teller 1984). Convergent evidence supporting a linking proportion comes from a number of experimental methodologies applied to animal

[9] Parvalbumin is a calcium binding protein involved in calcium signaling alterations in the function of parvalbumin-expressing neurons have been implicated in various areas of clinical interest such as Alzheimer's disease,[5] age-related cognitive defects and some forms of cancer.

models, especially to the behaving non-human primate, given the strong homologies of areas in the human and non-human primate brains (Schall 2004). Single-neuron recordings can relate neuronal activity to computations posited in psychological processes. Further evidence can come from perturbative techniques such as micro-stimulation or inactivation.

As an exemplary application of this perspective to a non-sensory function, Schall (2004) considered the neural underpinnings of the preparation of saccadic eye movements. In the case of saccade preparation, a well-supported candidate for the bridge locus is a distributed network of cortical and subcortical areas, including the frontal eye field and superior colliculus. During saccade preparation, so-called "movement" neurons in these areas exhibit a location-selective ramping of their firing rates, and a saccade is initiated when their firing rates reach a threshold level. At the level of mental processes, a leading psychological model for response preparation is accumulation of a signal until reaching a fixed threshold level that triggers the response. In such accumulator models, sequential sampling of a stochastic signal generates variability in the rate of rise to the fixed threshold, which can explain the observed variability in saccade reaction times. The linking proposition between a neural state (movement cell firing rates) and a psychological state (level of an accumulator) provides a framework for detailed hypothesis generation and experimental examination of psychological models.

What linking propositions do we have for core cognitive functions, and specifically for working memory and decision making? The neural correlates of working memory have been studied extensively through single-neuron recordings from monkeys performing tasks in which the identity of a transient sensory stimulus must be maintained in working memory across a seconds-long mnemonic delay to guide a future response. For instance, in one well-studied experimental paradigm, the oculomotor delayed response task, the subject is shown a visual cue appearing in one of 8 possible locations. The cue disappears during a delay period of a couple of seconds and the subject needs to maintain the position in working memory. The subject is then trained to perform a saccadic eye movement to the location of the cue so as to receive a reward (Funahashi et al. 1989). These studies revealed that a key neural correlate of working memory is stimulus-selective persistent activity, i.e., stable elevated firing rates in a subset of neurons, that spans the mnemonic delay (Goldman-Rakic 1995; Wang 2001). These neuronal activity patterns are observed across a distributed network of interconnected brain areas, with prefrontal cortex as a key locus. In the oculomotor delayed response task, for example, during the mnemonic delay, a subset of prefrontal neurons exhibit tuned persistent activity patterns, with single neurons firing at

elevated rates for a preferred spatial location (**Figure 3.2**). These neurophysiological findings have grounded the leading hypothesis that working memory is supported by stable persistent activity patterns in prefrontal cortex that bridge the temporal gap between stimulus and response epochs.

< insert Figure 3.2 around here - Funahashi et al (1989) >

The neural computations underlying decision-making have been most studied in task paradigms in which a categorical choice is based on the accumulation of perceptual evidence over time. In one highly influential task paradigm, the subject must decide the net direction of random-dot motion stimuli, which encourages decision-making based on the temporal integration of momentary perceptual evidence (Roitman and Shadlen 2002). Behavior can be well captured by psychological process models of evidence accumulation to a threshold. This is for example the idea behind the drift-diffusion model described in **Section 2.2**. Single-neuron recordings have found possible correlates of such an evidence accumulation process in association cortex, such as the lateral intraparietal area (LIP): choice-selective ramping of neuronal firing rates reflects accumulated perceptual evidence, with activity crossing a threshold level reflecting the decision commitment (Gold and Shadlen 2007). This is illustrated in **Figure 3.3**. These neural correlates reflect two key computations needed for perceptual decision making: accumulation of evidence and formation of categorical choice.

<inset Figure 3.3 (Gold & Shadlen (2007)) around here >

Conceptually, a neural circuit model can instantiate a linking proposition for a cognitive process and propose circuit mechanisms underlying the computations. If associated with a hypothesized bridge locus, model predictions for these circuit mechanisms can be experimentally tested, such as through single-neuron recordings. For instance, in the case of working memory, experiments have tested how focal antagonism of specific synaptic receptors affects persistent activity, thereby informing the neuronal and synaptic mechanisms supporting the computations (Wang et al. 2013; Rao et al. 2000). The stronger these links are among (i) the synaptic and neuronal processes in circuit mechanisms, (ii) neural activity, and (iii) the cognitive function, the better the model is validated. Once established, the model can then make rigorous predictions for the consequences of alterations in those circuit mechanisms. In this way, circuit models can iteratively contribute to our understanding of these links across levels of analysis and leverage them to study dysfunction in neuropsychiatric disorders.

3.4 Attractor network models for core cognitive computations in recurrent cortical circuits

Biophysically based neural circuit modeling has provided mechanistic hypotheses for how working memory and decision-making computations can be performed in recurrent cortical circuits. As noted above, a key neurophysiological correlate of working memory is stimulus-selective, persistent neuronal activity across the mnemonic delay in association cortical areas. Delays in working memory tasks (a few seconds) are longer than the typical timescales of neuronal or synaptic responses (10-100 ms). Similarly, perceptual decision-making demands categorical selection and benefits from temporal integration of evidence over long timescales (hundreds of milliseconds). Both of these computations therefore implicate circuit mechanisms.

Motivated by experimental observations of stable persistent activity in single neurons, a leading theoretical framework proposes that working memory-related persistent activity states are dynamical attractors, i.e., stable states in network activity. In the mathematical formalism of dynamical systems, an attractor state is an activity pattern that is stable in time, so that following a small transient perturbation away from this state; the network will converge back to the attractor state. A class of neural circuit models called "attractor networks" has been applied to explain the mechanisms that allow a recurrent network of spiking neurons to maintain persistent activity during working memory (Amit 1995; Wang 2001). An attractor network typically possesses multiple attractor states: a low-firing baseline state and multiple memory states in which a stimulus-selective subset of neurons is persistently active. Because the memory state is an attractor state, it is self-reinforcing and resistant to noise or perturbation by distractors, allowing the stimulus-selective memory to be stably maintained over time (Brunel and Wang 2001; Compte et al. 2000).

In a typical attractor network, subpopulations of excitatory neurons are selective to different stimuli. Recurrent excitatory synaptic connectivity exhibits a 'Hebbian' pattern such that neurons of similar selectivity have stronger connections between them (**Figure 3.4A**). When the strength of recurrent excitatory connections is strong enough, the circuit can support stimulus-selective attractor states that can subserve working memory (**Figure 3.4B**). Strong recurrent excitation thereby provides the positive feedback that sustains persistent activity. Wang (1999) found that incorporating physiologically realistic synaptic dynamics pose constraints on the synaptic mechanisms supporting this positive feedback. Strong positive feedback is prone to generate large-amplitude oscillations that can destabilize persistent

states and can drive firing rates beyond physiologically plausible ranges. It was found that both of these problems could be solved if recurrent excitation is primarily mediated by slow NMDA receptors.

<Insert Figure 3.4 around here>

Critically, recurrent excitation must be balanced by strong feedback[10] inhibition mediated by GABAergic interneurons. Feedback inhibition stabilizes the low-activity baseline state (Amit and Brunel 1997; Wang 1999). In a persistent activity memory state, recurrent inhibition also enforces selectivity of the working memory representation, preventing the spread of excitation to the entire neuronal population (Murray et al. 2014). Attractor dynamics supporting working memory are thereby supported by recurrent excitation and inhibition that are strong and balanced.

These circuit models make predictions for the relationship between synaptic mechanisms and working memory activity. These predictions have been confirmed through experiments with simultaneous single neuron recording from and pharmacological manipulation of prefrontal cortex: locally blocking excitation mediated by NMDA receptors attenuates persistent activity for the preferred stimulus (Wang et al. 2013). Similarly, locally blocking inhibition mediated by $GABA_A$ receptors reduces stimulus selectivity of delay activity by elevating responses to non-preferred stimuli (Rao et al. 2000).

In addition to working memory computations, strong recurrent excitatory and inhibitory connections in cortical attractor networks provide a circuit mechanism for decision-making, supporting temporal integration of evidence and categorical choice (Wang 2002; Wong and Wang 2006; Wang 2008). In this model, choice-selective neuronal populations receive external inputs corresponding to sensory information (**Figure 3.4C**). Reverberating excitation enables temporal accumulation of evidence through slow ramping of neural activity over time (**Figure 3.4D**). This property highlights that attractor networks not only support multiple stable states (representing categorical choices), but also support slow transient dynamics that can instantiate computations such as temporal integration. In these models, temporal integration via recurrent excitation benefits from the slow biophysical timescale of NMDA receptors (Wang 2002). Feedback and lateral inhibition mediated by GABAergic interneurons mediates competition among neuronal populations underlying the formation of a categorical choice. Irregular

[10] The words “feedback”, “lateral” and “recurrent” are here used interchangeably.

neuronal firing, a ubiquitous feature of cortex, contributes to stochastic choice behavior across trials, even when presented with identical stimulus inputs.

These computational modeling studies demonstrate that an association cortical microcircuit model can support working memory and decision-making computations through attractor dynamics. This therefore suggests a shared "cognitive-type" circuit mechanism for these functions, which may furthermore provide components upon which more complex cognitive processes may be built (Wang 2013). Because these functions rely on strong recurrent excitation and inhibition, they are particularly well suited to study how cognitive deficits may arise from alterations in synaptic function, which are implicated in neuropsychiatric disorders.

3.5 Altered excitation-inhibition balance as a model of cognitive deficits

Cortical attractor network models of working memory and decision-making function can be applied to characterize the impact of altered E/I balance in association cortex. Alteration of cortical E/I balance is implicated in multiple neuropsychiatric disorders, including schizophrenia, autism spectrum disorder, and major depression. A key strength of these circuit models is that they make explicit predictions not just for neural activity but also for behavior, which can be tested experimentally in clinical populations or after causal perturbation.

In schizophrenia, cortical microcircuit alterations are complex, with observed dysfunction in both glutamatergic excitation and GABAergic inhibition. Postmortem investigations of prefrontal cortex in schizophrenia find reductions in spines on layer-3 pyramidal cells, which potentially reflect reduced recurrent excitation. Such studies also have revealed multiple impairments in inhibitory interneurons, which potentially reflect reduced feedback inhibition. Pharmacological models of schizophrenia provide complementary evidence. One such approach is to use NMDA receptor antagonists such as ketamine, which transiently, safely, and reversibly induce cardinal symptoms of schizophrenia in healthy subjects (Krystal et al. 2003). A leading hypothesis regarding ketamine's effects on neural function is that the drug leads to a state of cortical disinhibition, potentially via preferential blockade of NMDA receptors on GABAergic interneurons (Greene 2001; Homayoun and Moghaddam 2007; Kotermanski and Johnson 2009). However, many questions remain regarding the neural effects of ketamine, such as

which NMDA receptor subunits and neuronal cell types may be the preferential sites of action (Khlestova et al. 2016; Zorumski et al. 2016).

Mechanistic links between altered E/I ratio and cognitive impairment remain tenuous, however. To address this issue, the aim of the modeling studies described below is to formulate dissociable behavioral predictions for distinct sites of synaptic perturbation. In particular, they have looked at perturbations of the E/I ratio via hypo-function of NMDA receptors at two recurrent synaptic sites: i) on inhibitory interneurons, which elevates E/I ratio via disinhibition; or ii) on excitatory pyramidal neurons, which on the contrary lowers E/I ratio (**Figure 3.5A**).

< insert Figure 3.5 around here >

3.5.1 Working memory models

Working memory function is a promising candidate in clinical neuroscience as an endophenotype, i.e. as a quantitatively measurable core trait that is intermediate between genetic risk factors and a psychiatric disorder (Insel and Cuthbert 2009). Working memory function involves different component processes: encoding of the memory, maintenance, robustness to distraction, precision, and capacity. Ongoing work in clinical cognitive neuroscience aims at resolving how these processes are impaired. Many studies have found a deficit in working memory encoding in patients with schizophrenia, that is a deficit that is observed even when the delay is set to zero seconds (Lee and Park 2005). For visuospatial working memory, patients with schizophrenia exhibit deficits not only in encoding but also in maintenance, which results in a graded loss of precision (Badcock et al. 2008; Starc et al. 2017). Other visual paradigms find reduced capacity of working memory but not necessarily precision (Gold et al. 2010).

Murray et al. (2014) examined the effects of altered E/I balance in a cortical circuit model of visuospatial working memory. Disinhibition, with results in an elevated E/I ratio, was implemented through antagonism of NMDA receptors preferentially onto interneurons. In this model, disinhibition leads to a broadening in the neural-activity patterns in the mnemonic attractor states (**Figure 3.5B**). This neural change induced specific cognitive deficits. During maintenance, the mnemonic activity pattern undergoes random drift, which leads to decreased precision of responses. Disinhibition increased the rate of this drift, thereby inducing a specific deficit in mnemonic precision during working memory maintenance.

Additionally, Murray et al. (2014) found that broadened neural representations make working memory more vulnerable to intervening distractors. In the model, distractors correspond to additional distracting inputs, modeled identically to the initial cues, with the same intensity and duration, but with a different stimulus location. A distractor is more likely to "attract" the working memory activity towards it if the two representations overlap. Distractibility therefore depends on the similarity between the representations of the memory target and the intervening distractor. Consistent with this model behavior, it has been found empirically that in visuospatial working memory, a distractor can attract the memory report toward its location, but only if the distractor appears within a "distractibility (spatial) window" around the target location (Herwig et al. 2010). Because disinhibition broadens the mnemonic activity patterns, this model predicts an increased range of distractors that can disrupt working memory for patients with schizophrenia.

To test the model prediction of broadened working memory representations under disinhibition, Murray et al. (2014) analyzed behavior from healthy humans administered ketamine during a spatial delayed match-to-sample task (Anticevic et al. 2012). In this task, subjects must retain the position of a cue in working memory. Subjects are then presented with a probe stimulus corresponding to different locations, and they must indicate if these probes are a "match" to the initial cue, or not. The model predicted a pattern of errors that is dependent on whether the probe is similar to the target held in working memory (aka the "memorandum"). Analysis of the behavioral data under ketamine versus control conditions revealed a specific pattern of errors which was similar to that predicted by the computational model. Consistent with model predictions, ketamine increased the rate of errors specifically for probes that would overlap with a broadened mnemonic representation. A similar pattern of errors has been observed in schizophrenia, with a selective increase in false alarms for "near" non-target probes but not for "far" non-target probes (Mayer and Park 2012). In contrast to the model predictions arising from disinhibition, insufficient recurrent excitation in the model leads to a collapse of persistent activity which would induce an error pattern of misses and spatially random errors.

To more directly test model predictions for patients with schizophrenia, Starc et al. (2017) designed a working memory task to be explicitly aligned with the model and with the primate electrophysiology task paradigms for which the model was developed. Such an alignment between the clinical study, basic neurophysiology findings and computational modelling allows stronger inferences and testing of hypotheses. In the working memory task of Starc et al. (2017), the memorandum is a single visuospatial

location and the response is a direct report of the remembered location, which provides a continuous measure of mnemonic coding. To test the model prediction of increased drift during working memory maintenance, the duration of the mnemonic delay is varied. To test the model prediction of increased distractibility dependent on target-distractor similarity, a set of trials included a distractor during the delay with a variable distance from the target. Starc et al. (2017) found that the experimental results largely followed model predictions, whereby patients exhibited increased variance and less working memory precision relative to healthy controls as the delay period increased. Schizophrenia patients also exhibited increased working memory distractibility, with reports biased toward distractors at specific spatial locations. This study illustrates a productive computational psychiatry approach in which predictions from biophysically based neural circuit models of cognition can be translated into experiments in clinical populations.

3.5.2 Decision making models

Broadly, decision-making function is impaired in multiple psychiatric disorders (Lee 2013). To study dysfunction in neural circuit models, we focus on perceptual decision-making in task paradigms similar to those studied via electrophysiology in non-human primates. As reviewed above, cortical attractor network models have been developed to capture behavior and neuronal activity from association cortex during random-dot motion paradigms (Wang 2002; Furman and Wang 2008). In these two-alternative forced choice tasks, a random-dot motion stimulus is presented, and the subject must report the net direction of motion (e.g., left vs. right). The coherence of the random-dot pattern can be parametrically varied to control the strength of perceptual evidence and thereby task difficulty. The psychometric function, giving the percent correct as a function of coherence, defines the discrimination threshold as the coherence eliciting a certain level of accuracy.

Random-dot motion paradigms have been applied to clinical populations and have revealed impaired perceptual discrimination in schizophrenia, as measured by a higher discrimination threshold (Chen et al. 2003, 2004, 2005). Similar impairments in the discrimination threshold have also been observed in patients with autism spectrum disorder (Milne et al. 2002; Koldewyn et al. 2010). These impairments are typically interpreted as evidence of neural dysfunction in sensory representations (Butler et al. 2008). However, it is possible that such impairments may also result from dysfunction in evidence accumulation downstream from early sensory areas, within association cortical circuits.

To explore this issue, Lam et al. (2017) studied the effects of altered E/I balance in the association cortical circuit model of decision-making developed by Wang (2002). The E/I ratio was perturbed bi-directionally, to compare the impact of elevated vs. lowered E/I ratio, via NMDA receptor hypofunction on inhibitory vs. excitatory neurons, respectively. Interestingly, Lam et al. (2017) found that the disruption of E/I balance in either direction can similarly impair decision-making as assessed by psychometric performance, i.e. the dependence of performance on the E/I ration is U-shaped, being degraded for both decreased or increased E/I ratio (**Figure 3.5D**). Therefore, the standard psychophysical measurements from clinical populations cannot dissociate among distinct circuit-level alterations: elevated E/I ratio, lowered E/I ratio, or an upstream sensory coding deficit.

Nonetheless, Lam et al. (2017) found that these regimes make dissociable predictions for the time-course of evidence accumulation. The random-dot motion task promotes a strategy of evidence accumulation across the duration of the stimulus presentation. In these settings, it is generally assumed that subjects continuously accumulate information during the stimulus presentation and only commit to a choice at the end of the stimulus stream. Contrary to these assumptions, however, it can be shown that how information is integrated is not uniform in time, and sometimes the decision is actually made long before the stimulus presentation ends. Multiple task paradigms have been developed to characterize the time-course of evidence accumulation. For instance, in the "pulse" task paradigm (Huk and Shadlen 2005; Wong et al. 2007), a brief pulse of additional coherence is inserted at a variable onset time during the otherwise constant-coherence stimulus (**Figure 3.5E**). This pulse induces a shift of the psychometric function according to pulse coherence. The dependence of this shift on pulse onset time reflects the weight of that time point on choice.

The pulse paradigm, as well as other paradigms, was able to dissociate distinct decision-making impairments under altered E/I ratio (**Figure 3.5E**). Under elevated E/I ratio, decision is impulsive: perceptual evidence presented early in time is weighted much more than late evidence. In contrast, under lowered E/I ratio, decision-making is indecisive: evidence integration and winner-take-all competition between options are weakened. These effects can qualitatively be captured using a modification of the drift-diffusion model, which is a widely used abstract model for decision making from mathematical psychology, described in **Section 2.3**. The standard drift diffusion model assumes perfect integration with an infinite time-constant for memory. Lowered E/I ratio in the circuit model can be captured by "leaky" integration with finite time-constant for memory. In contrast, elevated E/I ratio can be captured

by “unstable” integration, which has an intrinsic tendency to diverge toward the decision threshold. This study demonstrates the potential to link synaptic-level perturbations in neural circuit models to measurable cognitive behavior, as well as to more abstract models from mathematical psychology.

3.5.3 State diagram for the role of E/I balance in cognitive function

As described in the above section, neural circuit models of cognitive functions can generate dissociable predictions for how distinct synaptic perturbations impact behavior under various task paradigms. Biophysically based models can also suggest what aspects of neural activity or behavior may be differentially sensitive or robust to particular manipulations by pathology, compensation, or treatment. Changes in certain network parameters, or the combinations of parameters, may have much stronger impact on model behavior than changes in other parameter combinations. A “sloppy” axis in parameter space is one along which the model response is relatively insensitive to perturbations in that parameter combination, whereas a “stiff” axis is one in which the model response is highly sensitive to perturbations (Gutenkunst et al. 2007).

Murray et al. (2014) and Lam et al. (2017) characterized function in theses neural circuit models under parametric variation in E/I ratio. Specifically, they explored the parameter space of reductions of NMDA receptor conductances onto both inhibitory interneurons (elevating E/I ratio) and onto excitatory pyramidal neurons (reducing E/I ratio) (**Figure 3.6**). For the working memory model, circuit function is determined by the width of the mnemonic persistent activity pattern. For the decision-making model, circuit function can be measured through the discrimination sensitivity (inverse of the discrimination threshold). In both circuit models, E/I ratio was found to be a key parameter for optimal network function. Following relatively small perturbations, circuit function is robust as long as E/I balance is preserved. Preserved E/I ratio therefore corresponds to a “sloppy” axis in this parameter space. In contrast, even subtle changes to E/I ratio (along a “stiff” axis) have a strong impact on model function.

< insert Figure 3.6 around here>

If the imbalance is substantial, either elevated or lowered, the circuit can lose multi-stability. If disinhibition is too strong (via elevated E/I ratio), then the spontaneous state is no longer stable. Conversely, if recurrent excitation is too weak (via lowered E/I ratio), then the circuit cannot support persistent activity. Collectively, these analyses reveal that E/I balance is vital for optimal cognitive

performance in these cortical circuit models. This suggests that despite the complexity of synaptic alterations in a disorder such as schizophrenia, the impact on cognitive function in neural circuits may be understandable in terms of their "net effect" on effective parameters, such as E/I ratio, to which the circuit is preferentially sensitive.

3.6 Future Directions

In this chapter, we have primarily reviewed two studies leveraging biophysically based neural circuit models to explore the effects of altered E/I balance on the core cognitive functions of working memory and decision making. These studies revealed that E/I ratio is a critical property for proper cognitive function in cortical circuits. Furthermore, they provide a test bed for computational psychiatry demonstrating that neural circuit models can play a translational role between basic neurophysiology and clinical applications. Here we turn to some areas to be addressed in future modeling studies.

3.6.1 Integrating cognitive function with neurophysiological biomarkers

As noted above, biophysically based circuit models are well positioned to explore the mechanisms through which synaptic-level perturbations may be associated with neurodynamical biomarkers. In the context of schizophrenia, for example, circuit models have been applied to studying mechanisms of disrupted gamma-band oscillations (Vierling-Claassen et al. 2008; Spencer 2009; Volman et al. 2011; Rotaru et al. 2011), which can be related to EEG/MEG data from patients (Uhlhaas and Singer 2010). At very different spatiotemporal scales, circuit models of large-scale dysconnectivity can be related to resting-state BOLD data (Yang et al. 2014, 2016a). At the moment, such biomarker-related models do not directly relate to cognitive function or behavior. Future modeling work is needed in the integration of cognitive function with neurophysiological biomarkers across multiple scales of analysis.

3.6.2 Incorporating further neurobiological detail

To address increasingly complex and detailed questions about neural circuit dysfunction, future models will need to incorporate further elements of known neurobiology, which can be constrained and tested with experiments. One notable limitation in the current models is that they usually contain a single type of inhibitory interneuron, and therefore they are not able to speak to important questions regarding preferential disruptions in specific interneuron cell types. There are key differences between parvalbumin-expressing and somatostatin-expressing interneurons, which have differences in their

synaptic connectivity and functional responses (Gonzalez-Burgos and Lewis 2008). Microcircuit models that propose a division of labor among interneuron classes (Wang et al. 2004; Yang et al. 2016b) have the potential to make dissociable predictions for dysfunction in distinct cell types. Another avenue for model extension is to take into account the laminar structure observed in cortex (Mejias et al. 2016), which may relate to mechanistic hypotheses of impaired predictive coding (Bastos et al. 2012). Beyond the level of local microcircuitry, further work is needed on distributed cognitive computations across brain areas (Chaudhuri et al. 2015; Murray et al. 2017), and their integration in models of alterations in large-scale network dynamics in psychiatric disorders (Yang et al. 2014, 2016a).

3.6.3 Informing task designs

These modeling studies offer important considerations for the design of cognitive tasks applied to computational psychiatry. In each model, multiple distinct cortical disruptions (e.g., elevated vs. lowered E/I ratio vs. upstream sensory coding deficit) can impair performance. Standard performance analyses in common task paradigms (e.g., error rates in working memory, or psychometric threshold in decision making) may be insufficient to resolve dissociable predictions. Fine-grained analyses of task behavior should distinguish different types of errors or deficits, rather than simply measuring overall impaired performance, which could be due to deficits in distinct cognitive sub-processes (e.g., encoding vs. maintenance for working memory) or opposing deficits in a single sub-process (e.g., leaky vs. unstable integration in decision making). Circuit modeling can provide insight into the variety of potential "failure modes" in a cognitive function, and into which task designs can reveal them. In turn, alignment of a task design with a circuit model allows for generation of mechanistic neurophysiological hypotheses from behavioral measurements.

3.6.4 Studying compensations and treatments

Finally, of utmost relevance to psychiatry, biophysically based circuit modeling has the potential to provide a method for simulating possible effects of treatments that act at level of ion channels and receptors. As a proof of principle example of this, Murray et al. (2014) examined in the working memory circuit model how E/I balance can be restored through compensations acting on multiple parameters; for instance, elevated E/I ratio due to disinhibition can be compensated for by a treatment that strengthens inhibition or by one that attenuates excitation. In turn, restoration of E/I balance ameliorated the associated deficits in working memory behavior. However, further development and

refinement of biophysically based models is needed to go beyond proof of principle. Future development in this area will benefit from the other directions noted above. Incorporation of more detailed micro-circuitry and receptors will be needed to better capture pharmacological effects. Integration of biomarkers and behavior in the models will allow refinement through more direct testing with empirical data from pharmacological manipulations in animal models and humans.

3.6.5 Distributed cognitive process in a large-scale brain system

In this chapter, we focused on local circuit modeling, but cognitive processes involve multiple cortical and subcortical regions. Researchers have begun to consider abnormalities of the global brain connectivity and dynamics in mental illness (Rubinov and Bullmore 2013; Yang et al. 2016a), but this area of research is still in its infancy. In particular, persistent activity during working memory has been observed in a number of brain areas (Christophel et al. 2017). What are the general principles for such distributed persistent activity patterns? What determine whether a given brain area does or does not display persistent activity? For an area engaged in persistent activity, what does it store in working memory and what role does it play in contributing to, or controlling (i.e. as a network hub), the global persistent activity pattern? Time is ripe to seriously tackle these questions. Human brain connectomics and functional imaging are rapidly developing. At the same time, in animal research it is becoming possible to physiologically record from neurons in multiple brain areas of animals during a working memory task, and biologically-based large-scale modeling now can be built on quantitative mescoscopic connectivity data (Chaudhuri et al. 2015; Mejias et al. 2016; Wang and Kennedy 2016).

Advances in all these directions hold exciting promise of rationally guiding treatment development in psychiatry, grounded in basic neuroscience.

3.7 Chapter Summary

Research based on biophysically based neural circuit modeling has led to insights into neural circuits involved in cognitive processes in areas such as the prefrontal cortex. Working-memory is thought to be dependent on persistent neural activity during the delay period in areas such as the prefrontal cortex. Such persistent activity is usually modeled using attractor models. Decision-making is thought to correspond to ramping neural activity during the decision period, corresponding to an “accumulation of evidence”. In simulations, excitatory reverberation and maintenance of sustained activity during

working memory and accumulation of evidence during decision-making was found to require slow synapses, and particularly NMDA receptors. This theoretical prediction was verified experimentally and is meaningful for research in psychiatry: NMDA impairments have indeed been observed in schizophrenia and other psychiatric illness. Computational models can explain how changes in the balance of synaptic excitation and inhibition (through NMDA impairments) give rise to specific impairments in working memory and decision-making. Such local circuit models could provide basic building blocks in the development of more sophisticated models and large-scale brain circuit models in the emerging field of Computational Psychiatry.

3.8 Further Study

Shall (2004) discuss linking propositions and correspondence between mental states and brain states. An approach to relating attractor models and schizophrenia, that is related to the one described here, can be found in the work of Rolls & Deco (2011).

Durstewitz et al (2019) provides a more general dynamical systems perspective on psychiatric symptoms and disease, and discusses its potential implications for diagnosis, prognosis, and treatment.

On the physiological side, a recent review of the neural mechanisms that may underlie memory-associated persistent activity can be found in Zylberberg and Strozbridge (2019).

O Connell et al (2019) offers a recent description of neural and computational viewpoints on perceptual decision-making.

3.9 Acknowledgments

XJW was partly supported by the NIH grant R01 MH062349, STCSM grant 15JC1400104. A version of this article appeared in *Computational Psychiatry: Mathematical Modeling of Mental Illness*, 2017, edited by A. Anticevic and J. Murray, Elsevier Press.

Chapter 4: Computational Models of Cognitive Control: Past and Current Approaches

Debbie M. Yee[1] and Todd S. Braver[1]

[1]*Washington University in St. Louis, St. Louis, MO, USA*

4.1. Introduction

A core challenge of cognitive, computational, and systems neuroscience research is to provide a satisfying answer to the following question: how does cognition arise from neural systems? Although researchers have spent decades using variety of tools (e.g., magnetic resonance imaging, electroencephalography, single-unit recordings) to investigate this question, we have only begun to scratch its surface, in terms of understanding how neural substrates work together in synchrony to give rise to complex cognitive processes.

To provide an analogy, imagine listening to a concerto performed by a symphony orchestra. Perhaps you are interested in understanding how the orchestra can blend together so many different sounds from vastly different instruments to give rise to this beautiful masterpiece. In the initial hearing, the piece sounds clearly melodic, lyrical, and filled with multiple complex musical layers that sound cohesive when in concert. However, upon closer examination, it becomes evident that even such complex musical layers can be deconstructed into the contributions from different instruments within the entire ensemble. One approach for understanding the concerto may be simply to listen one instrument or one section (e.g., attending to a violin solo or the entire violin section when playing the same melody); however, that would only provide a small window into how that specific instrument contributes to the entire piece. Another approach would be to parse out all of the sounds in the piece by instrument, which provides a structural division of the different sounds that comprise the concerto but neglects the temporal ordering of when the instruments are played, an important aspect of the composition. Perhaps the most insidious problem is that even if we are able to understand the structural and temporal aspects of how each instrument contributes to this specific concerto, the same instruments in this symphony

orchestra can also perform a wide variety of other compositions (e.g., other concertos, sonatas, ballads) at other periods in time! Thus, the characterization of the violin's contribution to the current concerto may not be applicable when considering other musical performances, which makes this type of analysis effort not quite as generalizable as one might have hoped.

4.1.1 The Homunculus Problem of Cognitive Control

The challenge of this problem and the "orchestra concerto" metaphor becomes particularly salient when considering one of the most compelling mysteries of human cognition: how the brain enables human beings to plan, implement, and accomplish the types of controlled, complex, and temporally extended goal-directed behaviors that make-up much of modern daily life (e.g., preparing a multi-course meal, constructing IKEA furniture from an instruction manual, writing a computer program, solving a Sudoku puzzle, or figuring out how to successfully complete an MD or PhD). In the orchestra metaphor, it would be akin to understanding how the conductor guides the ensemble to put together a beautifully sounding and cohesive concerto performance. This mystery has often been posed as the "homunculus problem", which presents the following conundrum: if control over thoughts and action emerges from brain function, then are there special neurobiological and computational properties that differentiate the components that should be labeled as "controller" from the components that are "controlled"? Does the controller/controlled distinction even make sense? And if not, how are we ever going to understand the emergence of intelligent, goal-directed behavior in neurobiological terms?

Within psychology and neuroscience, researchers have often taken a primarily localizationist approach, studying individual brain regions in terms of their associated cognitive functions (Poldrack 2007). At the other extreme is the integrationist perspective, which focuses on the entire brain, parsing it into networks that may be structurally or functionally related (Eliasmith et al. 2012). However, neither of these approaches has yet provided a fully satisfying answer to the fundamental problem of cognitive control. Indeed, as this discussion hopefully makes clear, properly addressing the seemingly intractable homunculus problem likely requires a computational modeling approach. Computational approaches can be utilized in both a reductionist and emergentist manner: deconstructing the mysterious intelligence of the homunculus into hopefully more understandable "dumb" neural subcomponents, while at the same time making clear how complex control functions can emerge from the dynamic interactions among these multiple simpler subcomponents of cognitive control.

Computational modeling approaches to cognitive control are uniquely powerful, relative to other neuroscience techniques, in that they provide the researcher with a means of generating specific and concrete hypotheses, along with explicit experimental predictions regarding generative and causally efficacious control mechanisms and their influence on brain activity and behavior (Botvinick and Cohen 2014; O'Reilly 2006; O'Reilly, Herd, and Pauli 2010). More broadly, within the cognitive sciences, the utility of modeling approaches has long been established and appreciated (Newell and Simon 1961). Over thirty years ago, and as described in **Chapter 1** and **Figure 1.4**, David Marr attempted to formalize these approaches by articulating an influential proposal for decomposing and investigating complex cognitive systems across three levels of analysis: the *computational*, the *algorithmic*, and the *implementational* (Marr 1982; Bechtel 1994). These levels of analyses were initially introduced to tackle computational questions in vision, and have been criticized by various researchers as potentially being too rigid to be universally applicable (Dayan 2001). Yet the Marr framework can be fruitfully applied when considering complementary questions about the neural and computational mechanisms that underlie more complex temporally extended goal-directed behavior, such as: What computational goal is accomplished by a putative control function? What is the algorithm that encodes this function? Can we identify the neural systems and mechanisms that implement the algorithm? Consequently, we will make use of the Marr framework in this chapter, in order to provide a general intuition for how various computational models attempt to address specific questions about cognitive control function.

4.1.2 Why Cognitive Control?

The current chapter highlights past and current computational models of cognitive control, and the purpose is two-fold. First, cognitive control is a well-known psychological construct, with a long history of researchers using computational modeling approaches to attempt to explain its underlying cognitive mechanisms (Newell and Simon 1972; Rumelhart et al. 1986; Cohen, Dunbar, and McClelland 1990; Braver and Cohen 2000; Anderson et al. 2008). Second, cognitive control ability is disrupted across a wide range of mental disorders, with a vast body of literature now supporting the hypothesis that cognitive control impairments are prominent in many such disorders, including schizophrenia, depression, obsessive-compulsive disorder, ADHD, addiction, Alzheimer's Disease and Parkinson's Disease (Lesh et al. 2011; Fales et al. 2008; Halari et al. 2009; Greisberg and McKay 2003; van Meel et

al. 2007; Vaidya et al. 2005; Belleville, Chertkow, and Gauthier 2007; R. G. Brown and Marsden 1990; Wylie et al. 2010; H. R. Snyder, Miyake, and Hankin 2015). Indeed, it may not be an exaggeration to argue that an impairment of cognitive control, in one form or another, is the defining feature of many forms of mental illness. Thus, understanding the mechanisms that underlie cognitive control function can provide a crucial window into psychopathology.

Cognitive control is operationalized as the ability to perform task-relevant processing in the face of distractions or in the absence of environmental support, specifically by active maintenance and flexible updating of task representations over time, in order to pursue task-relevant objectives and behavioral goals (Engle and Kane 2004; Braver 2012; O'Reilly, Braver, and Cohen 1999). A core tenet of cognitive control is the distinction between controlled and automatic processing (Posner and Snyder 1975; Shiffrin and Schneider 1977; Norman and Shallice 1986). It is now generally appreciated that a fundamental tradeoff exists between recruiting and directing cognitive resources to deliberately perform a demanding task versus carrying out less effortful and habitual responses that may require fewer attentional resources, but which also may be less flexible. Typically, the allocation of control depends on the amount of cognitive effort or mental demand required. In other words, the control of behavior arises from the cognitive demands required to successfully perform a task, and effort allocation arises from the dynamic recruitment of available cognitive processes that can appropriately meet these demands during task performance (Botvinick and Cohen 2014). Some have proposed various computational models and frameworks to understand this tradeoff between effort and automaticity in controlled behavior (Cohen, Dunbar, and McClelland 1990; Schneider and Chein 2003), whereas others have hypothesized that humans perform a cost-benefit analyses between expected payoff and cognitive effort to determine the optimal allocation of cognitive control (Shenhav et al. 2017; Dixon and Christoff 2012; Kool and Botvinick 2014; Westbrook and Braver 2015). All in all, there still remain many unanswered questions regarding the computational and neural mechanisms that underlie cognitive control; which we argue can be more adequately addressed with computational modeling approaches.

As a brief aside, we wish to acknowledge that such computational modeling approaches have been prevalent and successful in advancing understanding for other related, but potentially more specialized cognitive processes, such as attention (Gershman, Cohen, and Niv 2010), learning (Tenenbaum, Griffiths, and Kemp 2006), semantic knowledge (Rogers and McClelland 2004), and memory (Polyn, Norman, and Kahana 2009). Thus, while this chapter will focus primarily on cognitive

control, we hope that the reader may extrapolate these principles to obtain a broader perspective for how computational models can be used to study other cognitive systems.

4.1.3 Roadmap to this Chapter

This chapter contains two main sections. First, we will provide a brief review of several key computational models that have been influential in advancing understanding of cognitive control mechanisms. This review of such models is not meant to be comprehensive but will hopefully provide a useful primer for readers to become familiar with classical and current computational models of cognitive control, with the understanding that the principles behind these models can be extended to other related models. Next, we discuss key features of computational models that make them particularly useful and generative in guiding further research efforts (i.e., what "tests" can we run to determine whether a computational model can make accurate and generalized predictions about controlled behavior?). The chapter concludes with a concrete example of how such modeling frameworks can be used to make predictions in mental illness, with some speculation about how cognitive control function breaks down in schizophrenia, a psychiatric disorder hypothesized to be strongly associated with cognitive control impairment.

4.2. Past and current models of cognitive control

A broad range of computational models have played a prominent role in the development and understanding of cognitive control theory and its underlying mechanisms, including those that have primarily arisen from symbolic modeling traditions, such as those involving production system architectures (ACT-R, Anderson 1996; EPIC, Kieras and Meyer 1997). At the other end of the spectrum are models arising from the computational neuroscience tradition (Wang 2013), similar to those covered in **Chapter 3**. Here, we focus on four contemporary models that address challenging and unique computational problems integral to cognitive control function, and which have also played an influential role in advancing research within this domain:

1. How do we determine when to actively maintain versus rapidly update contextual information in working memory?
2. How is the demand for cognitive control evaluated and what is the computational role of the anterior cingulate cortex?
3. How do contextual representations guide action selection during hierarchically organized task goals and what is the computational role of the prefrontal cortex?
4. How are task-sets learned and organized during behavioral performance, and when do they generalize to novel contexts?

4.2.1 How do we determine when to actively maintain versus rapidly update contextual information in working memory*?*

How does the brain determine what information is relevant to be maintained (i.e., in working memory) during the pursuit of task goals, and when should this information be updated with newer task-relevant information? A potentially useful analogy for visualizing this issue is the concept of a 'mental blackboard,' which describes the dilemma of deciding between when learned information in working memory should be kept, or instead erased and overwritten (Baddeley 1986). Early computational models attempted to use attractor models to understand the mechanisms that underlie robust active maintenance of working memory against irrelevant distractors (Changeux and Dehaene 1989; Zipser et al. 1993; Cohen, Braver, and O'Reilly 1996; Compte et al. 2000; Durstewitz, Seamans, and Sejnowski 2000; Deco and Rolls 2003). However, a major limitation of these models is their lack of a mechanism for precisely updating working memory when newer, task-relevant information is introduced. This tension between these two working memory functions is difficult to reconcile, as they inherently contradict each other – active maintenance increases resistance to distractors, whereas flexible updating makes the system more vulnerable to distraction. Thus, the computational challenge lies in building a model which can explain how a system regulates the fundamental trade-off between learning when to actively maintain context representations (i.e., task-relevant information that is internally represented) to achieve controlled processing versus rapidly updating new information into working memory, a core problem of cognitive control (O'Reilly, Braver, and Cohen 1999; Braver and Cohen 2000).

One approach towards understanding the computational mechanisms that underlie this trade-off comes from the "parallel-distributed-processing" approach (also dubbed "connectionist" or "neural

network" models in the literature, see **Section 2.1**). These models view control as arising from the interaction of multiple relatively simple elements (e.g., neurons or neural assemblies that perform local processes within a single brain system or unit). Thus, the models emphasize how cognitive control functions emerge from a network of brain regions activated interactively and in parallel, rather than the more historical modular approach of localizing cognitive function to a single brain region (Hinton 1984; O'Reilly 2006).

A well-established model from within the connectionist tradition is the prefrontal cortex and basal ganglia (PBWM) model developed by Frank, O'Reilly, and their colleagues (Frank, Loughry, and O'Reilly 2001; O'Reilly and Frank 2006; Hazy, Frank, and O'Reilly 2007). In the PBWM model, the prefrontal cortex (PFC) and basal ganglia (BG) interact to solve the maintenance vs. updating problem by implementing a flexible working memory system with an adaptive gating mechanism. This represents an elegant algorithmic solution for resolving this computational question, as it provides two separate modes of working memory that optimize active maintenance and flexible updating, respectively (**Figure 4.1a**). Specifically, working memory is insulated from distractor signals (i.e., irrelevant sensory input) when the gating mechanism is closed, but is receptive to utilizing information from such sensory signals when gating mechanisms are open. However, the introduction of this gating mechanism then begs the following question: how does the brain know when to open or close the gate? In other words, who or what controls the gate?

At the biological (i.e., implementational) level, the PBWM model proposes that the PFC facilitates the active maintenance mechanisms for sustaining task-relevant information, whereas the BG provides the selective gating mechanism, which independently switches between updating versus maintenance of information in PFC. Specifically, the key component of PBWM is that the BG performs this selective dynamic gating via disinhibition, and moreover, that this dynamic gating functionality depends upon the dopaminergic system (DA, **Figure 4.1b**). In this framework, dopaminergic "Go" neurons in dorsal striatum fire to disinhibit PFC to enable updating of working memory representations in PFC, while "NoGo" neurons counteract this effect to support robust maintenance of PFC working memory representations and resistance to distractions.

- Figure 4.1 –

Notably, other computational models have proposed similar gating mechanisms that regulate flexible updating and maintenance of task-relevant representations during working memory, but driven primarily by direct DA projections to PFC (Braver and Cohen 1999, 2000). However, a criticism of the global DA firing hypothesis is that this mechanism would not fully explain more complex cognitive tasks in which individuals would need to maintain and update different task representations simultaneously, such as when there is a hierarchical structure to working memory (e.g., remembering to press a button for a specific stimulus only during on context A, but not context B).

Taken together, the PBWM leverages the gating mechanism as an algorithmic solution to the computational problem of switching between active maintenance and flexible updating within working memory mechanisms. This model suggests that the PFC implements active maintenance of task-relevant information, whereas the BG contains selective gating mechanisms which switch between "robust maintenance" and "selective updating" of information held in PFC during working memory. Midbrain DA release is hypothesized to modulate this gating mechanism. However, exactly how, when, and where DA firing drives these working memory functions (e.g., only in the BG or also directly in PFC), is a question that remains to be fully explored.

4.2.2 How is the demand for cognitive control evaluated and what is the computational role of the anterior cingulate cortex?

Another core computational challenge within the domain of cognitive control is the following: how is the current demand for control evaluated, and in what form is this evaluative signal transmitted? In other words, how does the brain determine which situations or task conditions require more mental resources (than are currently available) to successfully pursue task goals, and what is the necessary relevant information that underlies this evaluation? This type of question is difficult to address from a purely theoretical perspective, as 'cognitive demand' is an elusive construct that appears to arise under a wide variety of mentally challenging tasks. Thus, a prerequisite for building a computational solution is understanding which experimental conditions demand and elicit greater cognitive control, and identifying relevant behavioral measures as empirical evidence for increased cognitive effort (note that in the literature, the terms cognitive effort and mental effort are used interchangeably). .

A plethora of work has identified tasks with behavioral measures that demonstrate selective recruitment of cognitive control (Botvinick, Cohen, and Carter 2004; Ridderinkhof et al. 2004; Braver and Ruge 2006). For example, in the Stroop task, cognitive control is required to override the prepotent response to read a word, in order to perform the correct task of reading the color ink of the word. In the N-back, cognitive control is required to respond selectively to N-back matches (e.g., in a 2-back task, a target response should be given only if the current stimulus matches the one presented 2 slides ago) rather than based on simple familiarity. In the stop-signal (or change signal) task, cognitive control is required to cancel an already initiated behavioral response if a stop signal (or change cue) is presented. In the Erikson flanker task, cognitive control is required to respond selectively to a centrally presented stimulus and ignore the flanker stimuli, particularly when these are distracting and incongruent with the central stimulus. Critically, all of these tasks contain experimental conditions that reliably increase cognitive control demands in a transient, trial-by-trial manner (i.e., the cognitive system monitors ongoing responses and adjusts to the level of cognitive control needed on the current trial). Likewise, they are indexed by specific behavioral measures that reflect this enhanced cognitive control demand (e.g., Stroop interference effect, stop-signal reaction time).

A well-established finding is that canonical control tasks, such as the ones listed above, consistently co-activate the dorsolateral prefrontal cortex (dlPFC) and the dorsomedial PFC (Egner 2009; Duverne and Koechlin 2017), a brain region that spans the dorsal anterior cingulate cortex (ACC) and pre-supplementary motor area (pre-SMA) (Duncan and Owen 2000; Duncan 2010). The dlPFC is thought to play a primary role in actively maintaining representations of task goals and the associated actions (or behavioral rules) needed to achieve them. In contrast, the ACC is thought to be involved in signaling when more control should be implemented by the dlPFC to accomplish these goals. It is generally accepted that the interaction between these two brain regions is important for dynamically adjusting cognitive control. Many have argued for the ACC as an important locus of cognitive control (Holroyd et al. 2004; Kerns 2004), although there remains much controversy over what actual information is represented by the ACC and signaled to the dlPFC to indicate that cognitive control is needed during tasks.

Several prominent theoretical accounts of ACC's computational role in cognitive control have arisen in recent years, including the detection of error signals (Gehring et al. 1993; Holroyd et al. 2005), reinforcement learning (Holroyd and Coles 2002), conflict monitoring (Botvinick et al. 2001; Botvinick,

Cohen, and Carter 2004), error likelihood (Carter et al. 1998; J. W. Brown and Braver 2005), cost-benefit analyses of implementing control (Shenhav, Botvinick, and Cohen 2013), and even uncertainty in the environment (Behrens et al. 2007). An account that was developed to reconcile and unify these divergent perspectives, the prediction response-outcome (PRO) model (**Figure 4.2**; Alexander and Brown 2011, 2014). The PRO model contains two components. One component of the model learns to predict multiple likely outcomes of various chosen actions, regardless of whether these outcomes or good or bad (i.e., response-outcome learning). A second component of the model detects discrepancies between actual and predicted outcomes and uses this prediction error signal (i.e., actual outcomes – expected outcomes) to update and refine subsequent predictions. Moreover, a key aspect of the prediction error signal is that it also indicates "negative surprise", when an expected outcome does not occur. This form of negative surprise signal can indicate not only when an unexpected error occurs, but also when the response is slower than expected or when the correct action is more ambiguous (which is likely to happen on trials associated with high response conflict).

- Figure 4.2 --

At the implementational level, the PRO model postulates that separate neural signals within ACC represent outcome prediction and prediction error (negative surprise), respectively. Specifically, the model suggests that the prediction signal should reliably increase immediately prior to when the mostly likely outcome will occur (i.e., a pre-response anticipatory signal). The negative surprise signal, on the other hand, will reliably activate after the action that produces an unpredicted outcome has occurred (i.e., a post-response evaluative signal). Critically, these hypotheses have been tested empirically across multiple tasks (e.g., change signal task, Erikson-flanker), as well as across different types of neural data (e.g., fMRI BOLD activity, ERP, monkey single unit neurophysiology). This validation of the PRO model across such a wide range of neural data demonstrates that it provides a useful generalizable computational algorithm by which the ACC can signal an increased need for cognitive control. Recent efforts have attempted to expand this account to include hierarchical representation within ACC and dlPFC (Alexander and Brown 2015), a topic relevant to the next section. Other recent efforts have attempted to link ACC signals with more affective/motivational quantities (Vassena, Holroyd, and Alexander 2017). These include the Expected Value of Control (EVC; Shenhav, Botvinick, and Cohen

2013) and related accounts (Holroyd and McClure 2015; Westbrook and Braver 2016), which postulate that ACC regulates the allocation and persistence of cognitive effort based on signals indicating the current subjective motivational (and/or hedonic) value of task and goal outcomes.

4.2.3. How do contextual representations guide action selection towards hierarchically organized task goals and what is computational role of the prefrontal cortex?

A third computational question of control relates to the issue of abstraction. How can a 'high-level' goal constrain and implement a 'lower-level' goal? As an example, imagine the following scenario: you hear the phone ring, and you have an instinctive impulse to lift it up from the receiver to answer it. However, context plays an important role in your action plan, so while you might automatically answer the phone in your own home, you would inhibit this tendency to answer a ringing phone at your friend's home. Yet, you might switch your action plan if your preoccupied friend asks you to answer the ringing phone on their behalf (e.g., when they are preoccupied with a task). This example articulates a fundamental computational challenge of implementing task goals – specifically, how do humans utilize contextual representations and higher-level goals to guide action selection during pursuit of lower-level goals, and how does the brain implement this type of hierarchical control?

One promising algorithmic solution for this perplexing question is the concept of hierarchical organization of task goal representations. The notion of applying hierarchical structure to parse complex systems into subordinate and interrelated subsystems has long been established, with subsystems being further subdivided into 'elementary' units (Simon 1962). Similarly, some theorists have argued that control signals used to guide behavioral actions, based on internal plans and goals, can also be subdivided into sensorimotor, contextual, and episodic levels of control (Koechlin, Ody, and Kouneiher 2003; Koechlin and Summerfield 2007; **Figure 4.3**). Critically, this information-theoretic model (i.e., based on principles from information theory; Shannon 1948), which has also been termed the "cascade model", postulates that the hierarchical division occurs according to a temporal dimension; that is, when in time control is implemented. Specifically, according to the model, actions selected based on temporally proximal stimulus would be lower on the hierarchy, whereas actions selected based on past information that is actively maintained in conjunction with the recent stimulus would be higher on the hierarchy. According to this framework, greater demand for cognitive control can also be formalized as

the amount of information required to be actively maintained over longer time periods to enable successful behavioral action selection. As a brief aside, it is worth noting earlier models also utilized hierarchical frameworks to understand temporal abstraction in behavior (Cooper and Shallice 2006), but the primary thrust of the cascade model and related variants been to use reinforcement learning to subdivide temporally abstract complex action plans (i.e., 'options') into simpler behaviors, an adaptive and efficient encoding strategy relevant for understanding structured abstract action representations (Botvinick 2008; Botvinick, Niv, and Barto 2009; Solway et al. 2014; Holroyd and Yeung 2011).

-- Figure 4.3 --

At the neural level, the cascade model implements hierarchical cognitive control along the anterior-posterior (i.e., rostral-caudal) axis of lateral PFC, with control signals higher up in the hierarchy represented in more anterior prefrontal regions (Koechlin, Ody, and Kouneiher 2003; Badre 2008; Badre and D'Esposito 2009). Although it is well accepted that PFC subserves high-level cognitive function and cognitive control, researchers have only recently attempted to build a parcellation scheme of this large brain region according to a functional organizing principle (Fuster 2001). Evidence from human neuroimaging studies supports the hypothesis of hierarchical representation, with more anterior regions of lateral PFC being activated when cognitive control is implemented for past information, and posterior regions being activated during action selection from more immediate information (Velanova et al. 2003; Braver and Bongiolatti 2002; Braver, Reynolds, and Donaldson 2003; Badre and D'Esposito 2007; Nee and Brown 2013). Additionally, single-unit studies in non-human primates are supportive of the idea that PFC is functionally organized according to the rostral-caudal axis: whereas caudal regions are involved in direct sensorimotor mappings, more rostral regions are involved in higher order control processes that regulate action selection among multiple competing responses and stimuli (Petrides 2005; Shima et al. 2007). Thus, the hierarchical organization of PFC appears to be central to performing the neural computations underlying task goal abstraction and action selection. Active research efforts focus on understanding how these divisions in the hierarchy are initially learned (Reynolds and O'Reilly 2009; Frank and Badre 2012), and whether the hierarchical structure is primarily anatomic or dynamic (Reynolds et al. 2012; Nee and D'Esposito 2016).

4.2.4 How are task-sets learned during behavioral performance, and when are they applied to novel contexts?

The fourth and final computational question in this chapter relates to the interaction of cognitive control and learning. In daily life, humans are faced with the challenge of learning a set of actions, sometimes simple or complex, in order to complete a specific task (i.e., a task-set). A related challenge is discerning between knowing when task-set rules that are learned in one context can be applied to a novel context (i.e., they generalize), or instead when a new task-set needs to be constructed. For example, when searching for the restroom at a shopping mall, one may learn a rule to look for signs that contain the text "Bathroom" with arrows pointing to a particular location. However, while this task-set rule may be pertinent when navigating malls in the United States, the same strategy may not be effective when searching for a restroom in other countries (e.g., United Kingdom), since the signs may read "W.C." instead of "Bathroom." Broadly speaking, creating a set of behavioral tools not tied to the context in which they were learned is useful, as this strategy enables flexible and efficient learning of task-set rules that can be generalized to novel contexts. However, the neural computations that underlie how cognitive control is deployed to learn task-sets are less well understood. Thus, the main motivating computational question is the following: in a new context requiring representation of tasks and task-set rules, is it more effective and efficient to generalize from an existing task-set representation (presumably stably encoded in long-term memory), or to instead build a new representation that is more optimized for the current context?

In the last decade, many accounts of cognitive control looked to algorithms and approaches from the reinforcement learning literature for inspiration in how task-set and goal representations might be acquired (Botvinick, Niv, and Barto 2009; Dayan 2012). A recent model that directly targeted this learning question is the context-task-set (C-TS) model, which aims to approximate how humans create, build, and cluster task-set structures (Collins and Frank 2013; **Figure 4.4**). The model's algorithm harnesses the power of both reinforcement learning and Bayesian generative processes that can infer the presence of latent states. Specifically, the model is designed to accomplish three goals: 1) create representations of task-sets and their parameters, 2) infer at each trial or time point which task-set is relevant in order to guide action selection, and 3) discover hidden task-set rules not already in its repertoire. A key element that drives the learning process is context - here defined as a higher-order

factor associated with a lower-level stimulus - which influences which action/motor plan would be selected. When the model is exposed to a novel context, the likelihood of selecting an existing task-set is based on the popularity of that task-set, i.e., its relevance across multiple other contexts. Conversely, the probability of creating a new task-set is set to be inversely proportional to a parameter indicating conservativeness, i.e. the prior probability that the stimulus-action relationship would be governed by an existing rule rather than a new one. Further, if a new task-set is created, the model must learn predicted reward outcomes following action selection in response to the current stimulus, as well as determine if the task-set is valid for the given context. If a selected action leads to a rewarding outcome, the model then updates the parameters to strengthen the association between a context and a specific task-set. Thus, the C-TS model provides a computationally tractable algorithm for task-set learning and clustering that not only feasibly links multiple contexts to the same task-set, but also discerns when to build a new task-set to accommodate a novel context. This process has been since dubbed 'structure learning.'

This structure learning process also has an implementational solution, simulated in a biologically plausible neural network model (in the same PDP tradition as the PBWM model), which provides a specific hypothesis about how structure learning occurs in the brain. In particular, the model formalizes how higher and lower level task-set structures and stimulus-action relationships are learned analogously within a distributed brain network involving interactions between PFC and BG. The key functional components of the model are two corticostriatal circuits arranged hierarchically with independent gating mechanisms. The higher-order loop involves anterior regions of PFC and striatum, which learn to gate an abstract task-set and cluster contexts associated with the same task set. The lower-order loop between posterior PFC and striatum also projects to the subthalamic nucleus, which provides the capability of gating motor responses based on the selected task-set and perceptual stimulus. Thus, the execution of viable motor responses is constrained by task-set selection, and conflict that occurs at the level of task-set selection delays the motor response, thus preventing premature action selection until a valid task-set is verified.

Both the algorithmic C-TS and the neural network model lead to similar predictions in human behavior. The convergence between these modeling approaches makes clear their joint utility as explanatory tools for understanding the processes that underlie structure learning. Specifically, together these models make an important claim: that humans have a bias towards structure learning, even when it

is costly, because such learning enables longer-term benefits in generalization and overall flexibility in novel situations (Collins 2017).

- Figure 4.4 –

From a broader perspective, a unique strength of using multiple computational modeling approaches is the ability to provide complementary insight into the cognitive and neural processes that result from the interaction of cognitive control and learning functions. These two variants of the C-TS model provide an admirable exemplar for how to integrate computational, algorithmic, and implemental analysis levels, and thus formalize a theoretical account that can approximate human implementation of cognitive control processing and structure learning. Thus, while the C-TS specifically targets understanding key mechanisms of cognitive control, the multi-level approach adopted to investigate these mechanisms provides excellent scaffolding for future computational investigation in other cognitive research domains.

4.3. Discussion: Evaluating Models of Cognitive Control

Next, we address two relevant issues in evaluating computational models of cognitive control: 1) what are good metrics for determining whether a model provides a useful contribution to our understanding of cognitive control mechanisms? and 2) how can models in this domain be successfully applied to understand the nature of cognitive control deficits in psychiatric disorders?

4.3.1 Model Evaluation: Determining Whether A Computational Model is Useful

A famous adage by the British statistician George E. P. Box states the following – “all models are bad; some models are useful.” It is generally accepted that most computational models are limited in their ability to account for all observed behavior, and at best typically encompass the critical data

variability within a certain limited cognitive domain (e.g., cognitive control phenomena related to standard experimental response conflict tasks), but do not generalize well beyond this limited domain, such as to novel tasks or contexts. Another common critique of algorithmic approaches in particular, is that these computations may not necessarily accurately reflect how cognitive processes are implemented on the biological level. For example, while a model may provide a sufficient hypothesis of cognitive control function and account for the key behavioral variance in a task, it is possible that the brain-behavior relationship may arise from a completely different computational or neural process altogether in the brain. Thus, an important step in this approach is model evaluation, i.e., deciding whether a model has utility. In other words, what makes a model useful for advancing cognitive research? Here we describe two complementary metrics for determining the utility of computational models – specifically, examining whether they are descriptive or predictive.

A computational model is *descriptive* if it provides a detailed explanation that accounts for significant variability of observed data (i.e., how well the model fits the data). Since models provide hypotheses about the data generating process, a descriptive computational model should provide insight into the mechanisms that give rise to the observed behavioral or neural responses in a given task. For example, an indisputable strength of Alexander and Brown's (2011, 2014) PRO model is its ability to account for a diverse range of empirical results, related to evaluation of demands for cognitive control, that span across both human and primate studies. Since the PRO model successfully models diverse neural and behavioral data from multiple cognitive control studies, it consequently provides compelling evidence for the hypothesis that predictive neural computation relating actions to outcomes implemented in the ACC and associated medial frontal regions may be a useful signal linked to the engagement of cognitive control. However, although the PRO model formalizes one potential algorithmic explanation for the generative process underlying extant data, it may neither reflect the actual neural computations that occur in the brain, nor necessarily accurately predict data outcomes in future studies. Thus, a limitation of this evaluation metric is that while a model with high explanatory power may explain prior data, the proposed mechanism may not be able to explain new data.

Conversely, a computational model is *predictive* if it describes a generative process that accurately forecasts and extrapolates to novel tasks or contexts. A predictive model contains a specific hypothesis about the neural computations that generate relevant data from one task or context and incorporates theory to reliably estimate behavioral and neural outcomes in a novel task/context. Collins and Frank's

(2013) convergent C-TS and neural network models provide excellent examples of predictive modeling, as both models make accurate predictions of behavioral outcomes in novel tasks/contexts. Critically, a theoretical assumption guiding development of these models is that humans spontaneously build task-set structure in learning problems. This structure learning assumption was tested in empirical studies, validating that the model could generalize to task contexts not previously learned. To summarize the key distinction put forth here, both 'descriptive' and 'predictive' computational models provide process mechanisms for how data are generated, but the former describes how well the model may fit extant data, whereas the latter describes how well the model generalizes to unseen data.

More broadly and generally, a computational model can serve a very useful function if it is explicitly specified to the degree that it can provide a focal point to drive and rejuvenate new research efforts. For example, while there is much controversy over ACC function, computational models have helped to elucidate potentially relevant cognitive mechanisms by providing specific testable hypotheses for empirical study (Botvinick and Cohen 2014; Vassena, Holroyd, and Alexander 2017). Moreover, although models may not always be accurate, they can highlight limitations of existing theory (e.g. what can and cannot be predicted by the model) and provide insight into how the theory should be revised in future iterations. The computational models described in this chapter are theory-driven approaches that attempt to describe how the brain implements cognitive control in an explicit way, in contrast to more vague descriptions by conceptual or verbal models. Thus, by attempting to spell out the exact mechanism for how cognitive control systems can be realized, the models described here provide explicit answers to the mysterious 'homunculus' problem of cognitive control. Furthermore, our hope is that such models will eventually be directly useful for elucidating how and why abnormal psychological and neurological processes arise in mental illness.

4.3.2 Cognitive Control Impairments in Schizophrenia

As an example of the point made above, we conclude this chapter with an example in which computational models of cognitive control have already been directly applied to a psychiatric disorder: specifically, to investigate the etiology of cognitive impairments in schizophrenia. A large literature on cognitive function in schizophrenia has reliably established that patients with this illness demonstrate impairments in attention, working memory, episodic memory, and executive functions (Snitz,

MacDonald, and Carter 2006). More specifically, an influential hypothesis is that schizophrenia is characterized by disrupted cognitive control, specifically a disturbance in the ability to internally represent and maintain contextual or task goal information in the service of exerting control over one's actions or thoughts (Cohen and Servan-Schreiber 1992; Barch and Ceaser 2012; Lesh et al. 2011; Barch, Culbreth, and Sheffield 2018). A key feature of the account is that such disruptions in cognitive control and context representation are directly linked to dysfunction of the DA neuromodulation in PFC, which has long been suggested to be a primary mechanism of pathophysiology in schizophrenia (Meltzer and Stahl 1976; S. H. Snyder 1976; Seeman 1987; Toda and Abi-Dargham 2007; Rolls et al. 2008). In particular, a common view is that at least some of cognitive impairments observed in schizophrenia putatively are related to reduced dysfunctional DA signaling in striatum and PFC, as well as increased 'noise' potentially resulting from increased tonic DA activity or aberrant phasic DA activity (Braver, Barch, and Cohen 1999; Rolls and Grabenhorst 2008; Maia and Frank 2017).

As a direct test for this hypothesis of dysregulated cognitive control and its relationship to DA and PFC, Braver and colleagues modified an extant computational model of prefrontal cortex function and context processing. Specifically, the goal was to make explicit predictions about behavioral and brain activity patterns that would be observed in schizophrenia patients performing the AX-CPT, an experimental paradigm designed to distill key aspects of cognitive control and context / goal maintenance (Braver and Cohen 1999; Braver, Barch, and Cohen 1999; Braver, Cohen, and Barch 2002). A key feature of this connectionist model, similar to the PBWM model discussed earlier by Frank and colleagues, is that contextual / goal representations are actively maintained in dorsolateral PFC, via mechanisms of recurrent connectivity and lateral inhibition. Most importantly, in this model, DA serves a joint neuromodulatory function within PFC, both gating representations into active maintenance (via phasic signals) and also regulating the persistence of maintenance (via tonic signals) (Braver, Barch, and Cohen 1999; Cohen, Braver, and Brown 2002). Model simulations with this DA neuromodulatory mechanism in PFC bolstered this hypothesis, providing evidence that context-dependent task performance, a key deficit in schizophrenia, is impaired with a noisy DA system (for more specific details, see Braver, Barch, and Cohen 1999). In particular, the model predicted very particular patterns of behavioral deficit in the AX-CPT task in participants with schizophrenia, as well as disruptions in the temporal dynamics of dorsolateral PFC activity, which were later confirmed experimentally (Barch et al. 2001; Braver, Cohen, and Barch 2002). Nevertheless, it has been difficult to demonstrate direct evidence that such deficits are specifically linked to DA neuromodulatory mechanisms, though recent

advances in fMRI techniques have allowed researchers to more precisely measure dopaminergic phasic signals within the brainstem (D'Ardenne et al. 2012).

Evidence for a related account of contextual / goal representation deficits in schizophrenia was shown by Chambon et al. (2008). Here, the goal was to test Koechlin's (2003) cascade model of hierarchical cognitive control in PFC to see whether it could account for particular patterns of behavioral impairment in individuals with schizophrenia. Interestingly, they observed that sensory and episodic dimensions of cognitive control were preserved in schizophrenic patients, whereas contextual control was impaired compared to matched healthy controls. In the study, patients generated significantly greater errors in tasks that required the ability to maintain context representations, and these impairments were highly correlated with disorganization score (e.g., a measure of disordered thought and behavior). Thus, the evidence is so far consistent with the hypothesis that in schizophrenia the ability to represent and actively maintain contextual or task goal information is disrupted. In future investigations it will be important to more directly test the claims of the cascade model that these deficits map appropriately along the rostro-caudal axis of PFC among individuals with schizophrenia.

4.4. Chapter Summary

This chapter highlighted several computational models that have played a seminal role in guiding theoretical accounts of cognitive control. Critically, we have selected these models because they provide promising testable hypotheses that have already stimulated a great deal of current experimental research, and which are likely to guide future investigations seeking to further elucidate the core neurocomputational mechanisms that underlie cognitive control. Furthermore, we hope that these models can be a useful primer for understanding computational approaches to cognitive processes more broadly, and how these processes may be disrupted in mental illness. Although computational modeling approaches have played a central role in understanding normative cognitive function (e.g., memory, attention), many of these models have not yet been explicitly tested in psychiatric populations. Thus, we argue that developing accurate mechanistic models of normative cognitive functions can, in principle and in practice, facilitate greater insight into the etiology of psychopathology.

4.5 Further Study

Rumelhart et al (1987) and O'Reilly & Munakata (2000) are seminal textbooks, which both provide an in-depth introduction into connectionist computational models. The second book incorporates more biologically realistic algorithms and architectures, and explicitly accounts for extant cognitive neuroscience data.

For a review of the main scientific questions of cognitive control, and computational approaches that have been proposed to address these questions, see also O'Reilly et al (2010) and Botvinick & Cohen (2014). An example of how different modeling levels can be utilized to provide converging evidence for cognitive control mechanisms can be found in Collins & Frank (2013). An example of how a computational model of cognitive control can be applied to make predictions about psychiatric disorder, specifically schizophrenia is offered by Braver et al (1999).

Chapter 5: The Value of Almost Everything: Models of the Positive and Negative Valence Systems and their relevance to Psychiatry

Peter Dayan

Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

5.1 Introduction

Humans and other animals are sufficiently competent at making choices capable of increasing their long-run expected rewards and decreasing their long run expected punishments that they can survive in a complex, threat-prone, and changing environment. Bodies of formal theory in economics, statistics, operations research, computer science and control engineering provide a foundation for understanding how systems of any sort can learn to perform well under such circumstances.

As is richly apparent in the present book, this understanding has been progressively rendered into the modern discipline of reinforcement learning (RL; Sutton and Barto, 1998), which provides close links to the ethology, psychology and neuroscience of adaptive behaviour. Further, it is perhaps the central leitmotif of computational psychiatry (CP) that dysfunctional behavior can be understood in terms of flaws, inefficiencies or miscalibration of RL mechanisms (Huys et al. 2015b, 2016; Maia and Frank 2011; Montague et al. 2012).

In this chapter, we consider one substantial aspect of these treatments, namely the notion of value – in both its definition and use. Conventionally, in RL, immediate utilities quantify the rewards or punishments (such as a pellet of food) provided upon doing a particular action (such as pressing a lever) when the world is in a given state (for instance, if a particular light is shining). These are weighted and summed or averaged over the long run to give rise to values. Since the long-run rewards depend on the long-run policy (i.e., the systematic choice of actions at states), various different sorts of values are often considered depending on aspects of these policies.

We here ask about the source and nature of actual and imaginary utilities, the calculations leading to value, and the influences of utilities and values on action.

In particular, we discuss topics phrased at Marr's computational level (as defined in **Section 1.2**) and

relevant to CP that arise over both the shorter- and longer-term. Shorter-term issues include risk-sensitivity (when utilities may be combined in a sub- or super-additive manner), and, in social contexts, other-regarding preferences (Fehr and Schmidt 1999), which arise when the utility that one agent derives depends on the returns that other agents achieve.

In the longer-term, it turns out to be possible to formalize the drive for exploration as a form of optimism in the face of uncertainty. This amounts to a fictitious or virtual reinforcement (Gittins 1979) (sometimes known as an exploration bonus) for actions about which there remains ignorance. Separately, we can quantify the opportunity cost of passage of time as arising from rewards that are foregone (Niv et al. 2007).

Finally, we also note the algorithmic problems posed by calculations of long-run expected value. These are addressed by the use of multiple mechanisms, each of which works well in a different regime of the amount of learning, and the time available for making a choice (Daw et al. 2005; Dolan and Dayan 2013; Doya 1999; Keramati et al. 2011; Pezzulo et al. 2013).

These ideas provide the lens through which we then view psychological and neural aspects of utility and value. We discuss three substantive ways in which this is not transparent. The first concerns the very definition of utility and its ties to emotion and affect (Bach and Dayan 2017; Buck 2014; Lindquist and Barrett 2012; Russell and Barrett 1999). The second is to consider asymmetries in the representation and effect of positive and negative values, along with associated notions of opponency (Boureau and Dayan 2010; Daw et al. 2002; Dayan and Huys 2009; Deakin and Graeff, 1991). Third, we view various apparent flaws in the way that we make choices in terms of heuristics that are tied to values (i.e., Pavlovian programming; Breland and Breland 1961; Dayan et al. 2006). Finally, we suggest some outline links to psychiatric dysfunction.

It is regrettably impossible to provide a complete account of such a rich and complex topic as value in a short and didactic chapter. The references should be consulted for a fuller picture.

5.2. Utility and Value in Decision Theory

5.2.1 Utility

As we have seen in **Section 2.3**, one of the core concepts in RL is the positive or negative immediate scalar utility, u_t associated with states or states and actions. This quantity describes that the experience

of the state or result of the action can be more or less profitable for the agent. For instance, an agent moving in a maze might be penalized ($u_t < 0$) for the cost of every move; and also suffer particular costs for being at a location in the maze that is very muddy or wet. Utilities are typically treated as being part of the description of the problem (or perhaps the environment) – they are an input to the algorithms that we will consider.[11]

Although utilities are basic, they are not necessarily simple. Two complexities that are important from the perspective of CP are their dependence on the state of the self or of others.

In terms of the self: utilities depend critically on one's current circumstances; failing to take this into account can readily lead to apparently dysfunctional decision-making. For instance, it seems obvious that food should have greater immediate utility when food-deprived; disruptions in this could have an obvious association with eating disorders. Similarly, there is some evidence that the marginal utility of leisure time also decreases as total leisure time increases (Niyogi et al. 2014a); breakdown in this can lead to over- or under-activity. Unfortunately, the relevant calculations for all these are not necessarily straightforward – and, as we will see, this problem is exacerbated for the case of distal rewards. The latter is particularly important for quantities such as money, whose immediate utility is questionable. Again, decreasing marginal utility with increasing wealth has been suggested as underpinning phenomena such as risk aversion, i.e. the reluctance to accept risky monetary prospects even when they involve an expected gain (but see Rabin and Thaler 2001 for a discussion about the plausibility of this explanation).

Utilities can also depend on other, more psychological, issues such as counterfactual (i.e. imaginary or potential) reinforcement (Breiter et al. 2001; Camille et al. 2004; Lohrenz et al. 2007). These can generate emotions such as regret (Bell 1982; Loomes and Sugden, 1982).

In terms of the other: most conventional applications of RL pit a single subject against the vicissitudes of an uncaring stochastic environment. However, it is often the case that psychiatric contexts involve the cooperation and competition of multiple intentional agents. Characterizing this requires taking

[11] There is a prominent exception in much of classical economics to the central role played by utility. There, preference between actions is considered to be primary (e.g., that the agent prefers reward A to reward B, without any necessary associated utility). Nevertheless under certain regularity conditions, such as transitivity (such that if the agent prefers going left to right, and right to up, then it will also prefer going left to up), it is known to be possible to derive a possibly non-unique set of scalar utilities that are consistent with the preferences (Houthakker 1950; Samuelson 1938, 1948). Of course, human (and animal) choices rarely satisfy these conditions, even stochastically, leaving the link to be the subject of rich study. The primacy of preference also arises in policy-based RL algorithms such as policy gradient (Baxter and Bartlett 2001).

considerations from Game theory, i.e. mathematical models of strategic interaction between rational decision-makers, into account (Camerer 2003). Such contexts typically imply a second dimension of complexity to immediate utilities – namely components that depend on the relative outcomes of the various players. For instance, consider two players (A and B) sharing a fixed amount of money. The utility that player A derives from a split might be decreased if she wins too much more than player B (a form of guilt) or not enough more than player B (a form of envy) (Fehr and Schmidt 1999). Such other-regarding utilities can then underpin strategies that seem beneficent or malign to other players (who then may engage in recursive modeling of each other's utility function in order to optimize their own utilities; Camerer et al. 2004; Costa-Gomes et al. 2001).

5.2.2 Value

With immediate utility as a basic concept in RL, the essential computational step for all decision-making algorithms is to compute either the value of a state ($V(s_t)$), or that of a state-action pair ($Q(s_t; a_t)$) under either a given policy or the optimal policy. These values are defined as long run summed or averaged utilities expected to accumulate over whole trajectories of interaction between a subject and the environment (potentially including other subjects). The expectations are taken over states and actions (i.e., over the transitions that govern state occupancy) and outcomes or utilities, if these are stochastic. There are different ways of quantifying the present value of multiple future utilities – the two most common are to use a form of exponential or hyperbolic temporal discounting (Ainslie 1992, 2001; Kable and Glimcher 2010; Myerson and Green 1995; Samuelson 1937; **Figure 9.2** this volume), or to consider the long run average rate of the delivery of utility (Kacelnik 1997; Mahadevan 1996; Stephens and Krebs 1986). Oddities of the calibration of this discounting are an obvious route to impulsivity in conditions such as attention deficit hyperactivity disorder, for instance (Williams and Dayan 2005). The more general observation that hyperbolic discounting leads to temporally inconsistent preferences has been of great importance in understanding a variety of behavioural anomalies (Ainslie 1992, 2001).

Temporal discounting also animates a different aspect of choice – namely when, or perhaps how vigorously, to perform a selected action. In some cases, the environment itself mandates an appropriate speed (when, for instance, it is necessary to perform a deed by a certain time to avoid a punishment; Dayan 2012). In other cases, the passage of time is penalized because the next, and indeed all

subsequent, contributions to the long-run utility are postponed when acting slowly. In other words, waiting to act will result in not receiving potential, or indeed certain, future rewards sooner. This leads to a form of opportunity cost (Niv et al. 2007). If it is also expensive to be quick – for instance, because of the energetic cost of a fast movement (Shadmehr et al. 2010), then the optimal choice of speed of action arises as a balance between these two costs (Niv et al. 2007; Niyogi et al. 2014b).

One prominent symptom of neurological (Parkinson's; Mazzoni et al. 2007) and psychiatric (depression; Huys et al. 2015a) diseases is a sloth or reluctance to act; for depression, this could arise because the opportunity cost of time is incorrectly perceived to be low (for one of a variety of psychological and neural reasons).

Estimating long-run values is a substantial challenge as trajectories become extended. As explained in **Section 2.2**, two classes of method have considerable currency in RL. Model-based (MB) calculations start from a characterization or cognitive map (Tolman 1948) of the environment and calculate forward to estimate the value. These calculations could involve building and exploring a tree, as in Monte Carlo tree search (MCTS; Kocsis and Szepesvári 2006). Alternatively, they could involve something closer to the methods of dynamic programming such as value or policy iteration (Bellman 1952; Puterman 2005). MB methods turn out to be statistically efficient, since cognitive maps are straightforward to learn, but computationally challenging, since long-run estimates are necessary and require long-run or recursive calculations.

One of the most important aspects of MB methods is that they allow the calculation of scalar utilities by combining predictions about what outcomes are imminent with information about the current motivational state (Dickinson 1985; Dickinson and Balleine 2002). The resulting sensitivity to occurrent changes in motivation (for instance refusing to do work to attain an outcome, such as water, that is not currently valuable, such as when not thirsty) is an important form of flexibility.

Note, though, that making present choices that are sensitive to the motivational state that will pertain in the future, i.e., anticipating how we will feel when the outcome will actually arrive – is apparently hard for us (Loewenstein 2000, though not necessarily for all organisms; Raby et al. 2007).

The second class of methods is model-free (MF). These involve *caching* the results of observed (or imagined; Sutton, 1990) transitions, typically by the bootstrapping method of enforcing sequential consistency in successive value estimates along trajectories (Samuel 1959; Sutton 1988; Watkins 1989). Q-learning MF methods (Watkins 1989) choose actions based on these values: $Q(s_t; a_t)$ reports the

benefit of performing action a_t. By contrast, actor-critic methods (Barto et al. 1983) exploit learned values $V(s_t)$ to criticize, and thereby occasion improvement to, a policy, which is the systematic specification of what to do at each state. The dependence on value in actor-critic methods is thus subtly different. Indeed, one should remember that, as in classical economics, it is the policy that ultimately matters; the values are a means to the end of defining an appropriate policy. Multiple different values might even be consistent with a given policy.

MF methods have the opposite characteristics to MB – they are statistically inefficient, since enforcing consistency allows inaccuracies to persist for long periods. On the other hand, they are computationally efficient (since one only need retrieve the value from the cache). However, this efficiency depends on their foundation on scalar utilities. This means that any sensitivity to motivational state has to be explicitly learned based on experiencing the utility of an outcome in that state (perhaps as a consequence of a relevant action), rather than being inferred, as for MB values. There is evidence for an excess influence of model-free decision-making in diseases involving inflexible compulsions (Gillan et al. 2016; Voon et al. 2015).

Various rationales have been suggested to govern arbitration between MB and MF values (see e.g. **Figure 2.4** in this volume) – for instance according to their relative certainties (which vary with the degree of learning and computational inefficiencies; Daw et al., 2005), or the opportunity cost of the time that it takes to perform model-based calculations (Keramati et al. 2011; Pezzulo et al. 2013). There are also possible ways in which the MB and MF systems could interact – for instance MB generation of imagined or simulated samples could train MF values or policies (Sutton,1990; Mattar & Daw 2018), MF values could be incorporated into calculations primarily involving MB reasoning (Keramati et al. 2016; Pezzulo et al. 2013), and MB methods could influence the way that MF systems assign credit for long-run rewards to actions or states (Moran et al. 2019).

One particular facet of optimization over trajectories is the influence of ignorance or uncertainty about the environment (or about one's compatriots). This ignorance increases if the environment undergoes change; however, it decreases with observations. Since taking optimal advantage of environments requires knowing enough about them, there is a form of trade-off between exploration and exploitation. Exploration is necessary to be able to be able to exploit; however, if every choice counts, then the possibility that the exploratory choice is bad instead of good (a possibility that must exist for exploration to be worthwhile in the first place), implies that there is a conflict between the two. One approach to the

exploration/exploitation tradeoff that has both formal and informal underpinnings (Dayan 2013; Gittins 1979; Ng et al. 1999; Sutton 1990; Szita and Lörincz 2008) is the exploration bonus. This is an internally awarded addition to the current expected utilities associated with only partly known actions and states. It is justified on the grounds that if, when explored, actions appear to be good, then they can be employed repeatedly in the future.

Exploration is particularly complicated in game-theoretic interactions with other intentional agents. An important formalism suggested by Harsanyi (1967) involves thinking of agents as having types, which determine their utility functions. One example would be their degree of guilt or envy (Fehr and Schmidt 1999). Agents know their *own* type; but can only engage in (Bayesian) inference about the unknown type of their partner. The fact that the players can model each other, and indeed model the other player's model of them, etc., leads to a structure known as a cognitive hierarchy (Camerer et al. 2004; Nagel 1995). In combination with the uncertainty about their partner's type, this can be represented as what is called an interactive Partially Observable Markov Decision Process (I-POMDP; Gmytrasiewicz and Doshi, 2005). This framework generalizes POMDPs to the presence of other, incompletely known, intentional agents. In I-POMDPs, it is necessary to worry that the probing, exploratory, actions that one does to gain information about a partially known environment risk convincing other players that one's own utility function is different from its actual form (e.g., that one is more envious and less guilty than is true). One's inferences about the types of other players based on their actions, have similarly to be tempered (Hula et al. 2015).

Other, arguably more heuristic, additions to utilities associated with curiosity and intrinsic motivation are also popular (Oudeyer et al. 2007; Schmidhuber 2010; Singh et al. 2004). These offer rewards for such things as reducing uncertainty or improving one's model of the environment.

However, some forms of ignorance and uncertainty appear to generate negative rather than positive value or utility. This is true in the case of ambiguity (or second-order probability) aversion (Ellsberg 1961; Fox and Tversky 1995), when subjects apparently unreasonably devalue gambles whose outcomes are imperfectly known. It is also seen in the fact that subjects are willing to incur costs to resolve uncertainty early (Bromberg-Martin and Hikosaka 2009; Dinsmoor 1983; Gottlieb and Balan 2010; Kreps and Porteus 1978) – think of how much it would be worth knowing your exam results early, even if you can't change them, and so not experience extended dread (Loewenstein 1987).

5.3. Utility and Value in Behaviour and the Brain

In the previous section, we discussed the abstractions of utility and valence that underpin the formal theory of optimizing control. We also saw some of the relevant complexities of both constructs in circumstances relevant to CP. In this section, we consider further aspects of their realization in psychological and neuroscience terms.

5.3.1 Utility

Our first concern is the provenance of utility – i.e., what determines the degree to which a given outcome or circumstance is rewarding or punishing. This is less straightforward than it may seem, for at least three reasons. First, evolution operates on the macroscopic timescale of reproductive success rather than the microscopic one of, for instance, slaking one's thirst or sating one's hunger. Our ability to procreate will ultimately depend on our survival and ability to maintain internal states, such as body temperature and blood sugar levels, within narrowly defined ranges, despite being subject to constantly changing external forces. Thus, there are attempts to define utility in terms of change in state relative to a point of homeostatic grace (Keramati and Gutkin 2014). Homeostasis involves maintaining a constant internal environment (e.g., suitable hydration) in the face of external challenges. Deviations from such a set point can be dangerous; so reducing such deviations is rewarding. This therefore provides a conventional reward for drinking whilst thirst, for instance. Nevertheless, the difference in timescale mentioned above suggests that these many apparently obvious components of utility are secondary to a primary aim, rather in the way that money, which clearly affords no direct benefit of its own, has become a central secondary target for modern humans.

One consequence of this is that it becomes compelling to see utility as a pawn in a (micro-economic) game between competing or cooperating systems, and so detached from hedonic notions such as liking (Berridge and Robinson 2003). That is, if the architecture of control is such that behaviour becomes optimized in order to increase notional utility, then these utility functions become surrogate means for how systems attempt to achieve ends. In other words, one system can seize control over behavioural output indirectly, by manipulating utility, rather than directly, by determining motor output. One possible illustration of the sort of behavioural anomalies that can result is anhedonia in chronic mild stress (CMS; Willner 2017). Anhedonia is generally defined as the inability to feel pleasure in normally pleasurable activities (and is one of the core symptoms of certain types of depression). CMS is a

protocol typically for rodents that involves multiple different and unpredictable irritations such as changing or wetting their bedding, tilting their home cages, exposing them to white noise, reversing the light/dark cycle, etc. Rodents subject to this regime exhibit reduced preference for sweetened, over neutral, fluids. From a conventional utility perspective, this might seem puzzling, since although CMS might reasonably lead the animals to conclude that the environment contains threats and even lacks opportunities, it seems implausible that sugar should be less immediately valuable. However, if one sees the utility associated with the sweetness as motivating vigorous behaviour (as it might for a control animal), then decreasing it might be appropriate in CMS, as it will prevent the animals embarking on potentially dangerous quests in a poor-quality environment. One could imagine that anhedonia in depression might arise in a similar manner (Huys et al. 2015a, 2013).

A second, and related, complexity attached to the provenance of utility is the way it might arise as part of a mechanism of arbitration between separate emotional systems (Bach and Dayan, 2017). By treating utility as a primitive, we have tacitly adopted a dimensional or, with some caveats, constructionist view of emotions (Lindquist and Barrett 2012; Russell and Barrett 1999). These view subjectively experienced emotions as constructed representations of more-basic psychological components such as valence (positive or negative affectivity – the component of particular relevance to utility) and arousal **(**how calming or exciting the information is). However, there are also many adherents of an alternative view, according to which there are multiple separate emotional systems that are individually optimized to respond to particular challenges or opportunities in the environment (Buck, 2014). Given that more than one such system might be active simultaneously, as for instance in conflicts between approach and avoidance, for instance, the brain has to have some form of arbitration. What Bach and Dayan (2017) discussed as virtual or "as-if" utilities can arise as an intrinsic part of the mechanism of arbitration - as studied in design economics (Roth, 2002). This is another utilitarian notion of utility, detached from any strong bond to 'liking' (Berridge and Robinson, 2003).

One might have hoped for a neural resolution to the nature of primary utility. For instance, dopamine neurons are deeply involvement in reinforcement learning (as we discuss below; Montague et al. 1996; Schultz et al. 1997). Amongst other things, their activity reports a signal that includes information about at least positive utility. Thus, one might hope that an analysis of the activity of their inputs might provide unambiguous information about these utilities. Further, a common assumption had been that at least the positive aspects of utility could be associated with nuclei in the lateral hypothalamus (LH), which are known to be involved in functions such as reward seeking in the context of food (see, e.g.,

Berridge, 1996; Kelley et al., 2005). There is also some direct electrophysiological evidence for reward-sensitivity in neurons in this structure (Nakamura and Ono, 1986). Unfortunately, investigations of the activity of those LH neurons that specifically project to dopamine neurons significantly complicate this view (Tian et al., 2016).

A third, psychological, complexity associated with utility is the influence of counterfactual outcomes. Counterfactual reasoning captures the process in which humans think about potential or imaginary events and consequences that are alternatives to what actually occurred. Regret is the emotion experienced upon discovering that an option not chosen would have been more valuable than the one that was (rejoicing being the positive alternative). This has an important impact in economic choice (Bell 1982; Loomes and Sugden, 1982), including as part of algorithms for game-theoretic performance (Camerer and Ho, 1999), and has also been frequently examined in psychological and neuroimaging studies (e.g., Coricelli et al. 2007; Kishida et al. 2016; Lohrenz et al. 2007). The prospect of future regret plays an important role in certain choice environments, i.e., there can be a substantial contribution to the present value of an option from the future disappointment to which it might lead.

5.3.2 Value

We have argued that it is mostly values – of both states and actions at states – that determine behaviour. Indeed, 'true' utility is only assessable after the fact – in some cases long after. One famous example of this involves a comparison between nutritive and non-nutritive sweeteners such as glucose and saccharine respectively. An initial preference can develop for both; but then reverse for the latter, presumably as subjects discover that they are nugatory (e.g., Warwick and Weingarten, 1994, see also the discussion in McCutcheon 2015). This implies that any immediate report on instant palatability (i.e., the sweet taste) might be best thought of as a prediction that something of actual biological relevance (sugar) has been delivered. More generally, it is the structure of predictions of future outcomes, and their net worth, that determines many aspects of behaviour.

5.3.3 Evaluation

There is by now a huge wealth of information about the construction and competition of MB and MF predictions, at least in the appetitive case (Adams and Dickinson 1981; Balleine, 2005; Daw et al. 2005; Daw and Dayan 2014; Daw et al. 2011; Dayan and Berridge 2014; Dickinson, 1985; Dickinson and Balleine 2002; Dolan and Dayan 2013; Doya 1999; Hikosaka et al. 1999; Killcross and Coutureau 2003; Lee et al. 2014; Montague et al. 1996; Schultz et al. 1997). As noted above, model-free predictions typically arise by measuring the inconsistency between successive estimates of long-run value in the form of a temporal difference prediction error (Sutton, 1988), and using this to update predictions.

There is evidence that this prediction error is broadcast via the phasic activity of dopamine neurons (Cohen et al. 2012; Eshel et al. 2013; Montague et al. 1996; Schultz et al., 1997) to key target structures, notably the striatum (Hart et al., 2014; Kishida et al., 2016), the amygdala, and beyond. Much is known about sources of this prediction error (e.g., Matsumoto and Hikosaka, 2007; Tian et al. 2016), although the loci where state or state-action values, or even the actor portion of the actor-critic, are stored is less clear. One prominent idea is that successive 'twists' of a helically spiralling connection between the dorsolateral striatum and the dopamine system (Haber et al. 2000; Haruno and Kawato 2006; Joel and Weiner 2000) are implicated in forms of MF control (Balleine, 2005) that go from being related to state-action values (Samejima et al. 2005) towards being simpler and actor-based (Li and Daw 2011).

This arrangement has various implications. For instance, one of many routes to drug addiction involves substances seizing control of this prediction error, allowing them to masquerade as having substantial value (Redish et al. 2008; see also **Chapter 9**). This can then dramatically retune the behavioural direction of the subject towards increased acquisition and consumption of these substances. Subsequent neural changes, such as adaptation, can then cement the malign assessments, making them hard to change.

MB values are constructed on the fly via a process of planning, for instance through a form of constraint satisfaction (Friedrich and Lengyel 2016; Solway and Botvinick 2012) or Monte-Carlo tree search (Kocsis and Szepesvári 2006). One account of the latter is episodic future thinking (Schacter et al. 2012), i.e. using memory for specific happenings in one's personal past to imagine the future, an operation that involves the hippocampus. Other structures implicated in MB evaluation include regions of the prefrontal cortex and the dorsomedial striatum (Balleine 2005); and there is evidence that parietal regions are involved in the construction and maintenance of the model (Gläscher et al. 2010). Evidence for MB decision-making can be found in devaluation paradigms (Dickinson and Balleine 2002). In such

experiments, the outcome values are changed suddenly. The change could be internal, e.g. through selective satiation of the animal, or external e.g. if the previously pleasurable reward is then poisoned. Immediate choice adaptation to this change (before any learning can occur) is evidence for MB (or goal-directed) control, since the MF (habitual) system would require learning about reward experience before it can alter behavior accordingly).

Other sorts of MB sensitivity have also been reported. For instance, aspects of the expected values associated with potential future food rewards appear to be reported in the vmPFC region of prefrontal cortex in human subjects (Hare et al. 2008). These representations are modulated by apparent top-down goals (potentially via connections with other regions of the PFC) such as healthy eating (Hare et al. 2009, 2011). Something similar is apparently true for other forms of top-down modulation, as in charitable giving (Hare et al. 2010) or even remunerated sadism (Crockett et al. 2017). As with the observations above about the malleability of primary utilities, this shows how value may also be a pawn in battles over choice.

When salient outcomes are modestly distant in time, their expectation appears also to have direct consequences for value, by generating what are known as anticipatory utilities (Loewenstein 1987). Appetitive prospects generate savouring, which grows as the time of acquisition nears (albeit also then lasting, and so accumulating, for a shorter period of time); aversive prospects generate dread. It has been argued (Iigaya et al. 2016) that such anticipation accounts for the value contributions associated with observing, generating the preference for the early resolution of uncertainty mentioned above (Bromberg-Martin and Hikosaka 2009; Dinsmoor 1983).

5.3.4 Aversive values and opponency

We have so far mostly considered appetitive values. However, aversion, punishment and even the cost of effort are also critical – as is the integration between all these factors. One possibility is that utility and value are signaled by a single system whose neurons enjoy an elevated baseline firing rate, so that positive and negative values could be equally represented by above and below baseline activity respectively (or vice versa). There is some evidence for this (Hart et al. 2014); and indeed, the dopaminergic architecture of the striatum has been argued to be exquisitely tailored to this job, with direct and indirect pathways associated with choosing and suppressing actions and associated with

different dopamine receptors, and chiefly sensitive to increases and decreases in dopamine (Collins and Frank, 2014; Frank and Claus 2006; Frank et al. 2004). However, there is evidence for asymmetric signaling in dopamine activity (Niv et al. 2005), and also for heterogeneity, with particular dopamine neurons responding in aversive circumstances (Brischoux et al. 2009; de Jong et al 2019; Lammel et al. 2014, 2012; Matsumoto and Hikosaka 2009; Mirenowicz and Schultz 1996, but see Fiorillo 2013). There are findings that dopamine concentrations barely reflect effort at all (Gan et al., 2010; Hollon et al., 2014) (along with known associations between this neuromodulator and vigour; Beierholm et al. 2013; Hamid et al. 2016; Niv et al. 2007; Salamone et al. 2016). There are, instead, substantial, albeit controversial, suggestions for opponent representations of reward and punishment (Boureau and Dayan 2010; Daw et al. 2002; Deakin and Graeff 1991; Deakin 1983), involving two, interacting, systems. It has been argued that these are consistent with what is known as a two-factor account of aversion (Johnson et al. 2002; Maia 2010; Moutoussis et al. 2008; Mowrer 1947), in which actions that cancel predictions of potential negative outcomes (for instance by leading to signals for safety; Fernando et al. 2014) are themselves reinforced. To put it another way, a unitary mode of reinforcement of choices comes from outcomes being better than expected, rather than good (Dayan 2012; Lloyd and Dayan 2016).

We noted above that one could justify exploration in the face of ignorance by the benefit that would accrue if one thereby discovers facets of the environment that can be exploited (Dayan 2013). This requires a suitable degree of controllability, such that, for instance, actions have reliable consequences (Huys and Dayan 2009). Dual to this beneficial effect of uncertainty are aversive assessments which amount to forms of predictive anxiety: ignorance can be dangerous if bad outcomes are legion (which might perhaps underpin ambiguity aversion; Ellsberg 1961); similarly, change can be expensive, if hard-won knowledge about how to exploit the environment effectively expires. One interpretation of the neuromodulator norepinephrine (NE) is that it reports on forms of unexpected uncertainty - induced by unpredictable change (Devauges and Sara 1990; Yu and Dayan 2005); there is indeed evidence of a close association between NE, stress and anxiety (Itoi and Sugimoto 2010).

5.3.5 Instrumental and Pavlovian use of values

Given some of the various ways that state and state-action values may be determined and learned, we

next consider their effect. It is here that the differences between Pavlovian and instrumental behavior become critical (Mackintosh, 1983). State-action values (such as $Q(s_t; a_t)$), which estimate the long-run future value that is expected to accrue starting from state s_t, choosing action a_t, and then following a conventional policy thereafter, are part of an instrumental control structure. These values are learned or inferred based on the contingency between actions and outcomes; thus, choice is similarly contingent. The same is true when values just of states $V(s_t)$ are used to train a policy (i.e., in an actor-critic method; Barto et al., 1983), based on the changes in these values contingent on the actions. However, animals are also equipped with forms of preparatory Pavlovian control (Mackintosh 1983). In this, stimuli (signifying states) associated with appetitive or aversive values, i.e., predictions of (net) future gain or less respectively, elicit actions without regard to the actual contingent consequences of those actions. Appetitive predictions lead to active, vigorous engagement and approach. By contrast, aversive predictions lead to withdrawal, inhibition, suppression and freezing. Thus, for instance, pigeons will peck at lights that have been turned on just before food is delivered, even if this pecking has no contingent consequence at all. These behaviors are presumably evolutionarily appropriate and have the benefit of not needing to be learnt. However, the lack of contingency implies that the actions are elicited even if they paradoxically actually make less likely the outcomes that support the underlying predictions (Breland and Breland 1961; Dayan et al. 2006; Guitart-Masip et al. 2014). For instance, pecking can still be observed in omission schedules, i.e., when the pigeons do not actually receive food on any trial in which they peck at an illuminated light (Williams and Williams 1969),

Pavlovian influences interact with instrumental behavior in at least two further ways. One is by modulating the vigour of ongoing instrumentally directed responses, in the form of what is known as Pavlovian to instrumental transfer (PIT; Cartoni et al. 2016; Estes 1943; Murschall and Hauber 2006; Rescorla and Solomon 1967). Pavlovian-instrumental transfer is defined as the phenomenon that occurs when a conditioned stimulus (CS) that has been associated with rewarding or aversive stimuli via Pavlovian/ classical conditioning alters motivational salience and operant behavior. PIT comes in two flavours: specific and general. Specific PIT happens when a CS associated with a reward enhances an instrumental response directed to the same reward. For example, a rat is trained to associate a sound (CS) with the delivery of a particular food. Later, the rat undergoes an instrumental training where it learns to press a lever to get that particular food (without the sound being present). Finally, the rat is presented again with the opportunity to press the lever, this time both in the presence and absence of the sound. The results show that the rat will press the lever more in the presence of the sound than without,

even if the sound has not been previously paired with lever pressing. The Pavlovian sound-food association learned in the first phase has somehow transferred to the instrumental situation, hence the name 'Pavlovian-instrumental transfer. Under general PIT, instead, the CS enhances a response directed to a different reward (e.g. water). The difference between these flavors is analogous to that between the MF and MB predictions that may underpin them.

The anatomical basis of preparatory appetitive and aversive Pavlovian actions and PIT is not completely clear, although there is evidence for the involvement of various regions. One is the ventral striatum (Reynolds and Berridge 2001, 2002, 2008). Another is dopaminergic neuromodulation (Faure et al. 2008; Murschall and Hauber 2006), which is important for active responses in appetitive and aversive domains, and serotonergic neuromodulation, which plays a particular part in aversive contexts (Faulkner and Deakin 2014), underlying its part role as the putative opponent to dopamine (Boureau and Dayan 2010; Daw et al. 2002; Deakin and Graeff 1991; Deakin1983). Specific and general PIT depend on distinct circuits linking central and basal nuclei of the amygdala to the core and shell compartments of the ventral striatum respectively (Balleine 2005; Corbit and Balleine 2011; Corbit et al. 2016).

A second interaction between Pavlovian and instrumental behaviour is more restricted. As we noted, one process underlying model-based evaluation of states or actions at states is thought to be building and traversing a tree of prospective future states – i.e., chains of episodic future thinking (Schacter et al. 2012). In planning series of actions in this way, it is usually infeasible to consider all potential future sequences; instead, one must cut the expanding decision tree down to a computationally manageable size. There is evidence that Pavlovian predictions can be involved in this process of pruning, being reflexively evoked by large losses and persisting even when disadvantageous (Huys et al. 2012). This is an internal analogue of aversion-induced behavioral inhibition, i.e. the tendency to withdraw from unfamiliar situations, people, or environments in the face of expected aversive outcomes. For a more concrete example, imagine planning chess moves by considering future board positions. A variation in which a queen was lost might be pruned away in this manner, even if this variation would ultimately have led to an advantageous checkmate.

Both instrumental and Pavlovian predictions can themselves be MB or MF. We already pointed to this in the instrumental case – apparent, for instance, in the wealth of devaluation paradigms (Adams and Dickinson 1981; Dickinson 1985; Dickinson and Balleine 2002). There has perhaps been less focus on this in Pavlovian circumstances, although we noted that evidence of both specific and general PIT can

be interpreted in this manner; and there are also some direct observations about preparatory Pavlovian actions along with modulation of instrumental ones (Dayan and Berridge 2014; Robinson and Berridge 2013). Furthermore, the form of the preparatory Pavlovian response can be influenced by the nature of the outcome as well as that of the predictor (Davey et al. 1989). For, instance pigeons exhibit distinct food- and water-directed pecks; and apply them specifically to lit keys that predict food and water respectively (Jenkins and Moore 1973). There are also clear individual differences. For instance, during Pavlovian conditioning, individuals vary widely in their propensity to engage with CSs (called sign tracking) or the sites of eventual reward (goal tracking) in circumstances under which these differ. Sign- and goal-tracking subjects appear to rely more on MF and MB systems respectively (Robinson and Flagel 2009).

Along with preparatory Pavlovian responses are consummatory ones that are typically elicited by the presence of the biologically significant outcomes that inspire consumption or defense (rather than by values). There is a particularly rich and complex set of defensive responses that are specific to the species concerned (Bolles 1970), and sensitive to subtle aspects of the relationship between the subject and the threat (McNaughton and Corr 2004). This is apparently controlled in rodents (Blanchard and Blanchard 1988; Keay and Bandler, 2001) and humans (Mobbs et al. 2007) via a structure called the periaqueductal gray.

5.4. Discussion

In this chapter, we have discussed issues of utility and value, which are the engines underlying choice. We saw some of the many complexities of the definition and determinants of utility, and then the computational issues that arise with either learning (in a model-free manner) or inferring online (in a model-based manner) the long-run predicted values of states or states along with actions. We noted additional factors such as information, risk, ambiguity and motivational state that can change or influence utility and value, along with the behavioral impact of values in terms of mandatory preparatory Pavlovian behaviors such as approach and withdrawal. We also noted that values should optimally influence the alacrity or vigor of action.

We observed that many different neural systems are involved in the assessment and effects of both

appetitive and aversive utility and value. Evidence is unfortunately currently somewhat patchy as to how they all fit together, and indeed the many opportunities that each, and their combinations, afford for supporting benign and malign individual differences. Some foundational questions remain to be answered, such as whether there is opponency between systems associated with each valence.

One theoretical approach to computational psychiatry (CP) starts from some of the different sources of dysfunctional decision-making: it can result from the brain trying to solve the wrong computational problem, solving the correct problem incorrectly, and solving the correct problem, correctly, but calibrated to an incorrect environment (Huys et al. 2015b). Although utility and value influence all of these in various ways, in terms of the current chapter, it is most straightforward to consider incorrect utilities ('solving the wrong problem') and inefficient or ineffective calculations (the correct problem 'solved incorrectly').

It seems obvious that utilities actually define optimal choices – however, we noted that utilities are not primary and impenetrable, but rather are contextually determined and what one might call meta-adaptive. This affords attractive flexibility; but it is also a clear point of vulnerability: utilities might be influenced by early insult, or incorrect or outdated priors. For instance, following seemingly random aversive events, a person could develop a prior that the world is not very controllable, with actions having highly stochastic consequences. This would come with the implication that there is little point exerting effort trying to explore it, since the information gained would not be expected to be exploitable (Huys and Dayan 2009). As in our description of chronic mild stress, one way for the brain to inhibit exploration would be to dial down the subjective utility of outcomes. This would be pernicious, since failing to explore may entail failing to find out that the prior no longer pertains. We have argued that various such failings of the prior can lead to forms of depression (Huys et al. 2015a), but they can readily extend to addiction and beyond.

Incorrect calculations, leading to incorrect acquisition or calculation of long run values, are another substantial source of problems. Some particular cases have been studied. One concerns the automaticity of the Pavlovian pruning of the internal search tree used to calculate expected future values; this has been considered a point of vulnerability (Huys et al. 2012). Another case concerns the under-weighting of MB over MF choice (Voon et al. 2015). This leads to behavior that is inflexible in the face of change and fails to reflect information that the subject can be shown to possess. On the surface, many psychiatric conditions share this characteristic; a deeper investigation using comparisons with factor

analytical summaries of answers to structured questionnaires showed that it is actually most closely associated with measures of compulsivity (Gillan et al., 2016). A second set of issues with calculation arises from Pavlovian influences over actions. For instance, we see people as being impulsive (Evenden, 1999) when they are apparently chosen immediate, short term, positive outcomes. One of many possible sources of this is a form of Pavlovian misbehaviour (Dayan et al., 2006) – approach in the face of predictions of future positive valence irrespective of the contingent consequences.

5.5 Chapter Summary

In sum, although it might seem that nothing could be simpler than learning to favor actions that lead to positive outcomes, there are actually many richly complicating factors. These factors can achieve important ends – including tailoring behavior to long run goals; adapting those goals in the light of particular contexts; accommodating prior expectations over the brutishness and brevity afforded by evolutionary contexts by using hard-wiring and heuristics to avoid as much of the cost and danger of learning as possible. These factors leave an architecture of choice replete with readily exposable flaws and vulnerability to the sort of psychiatric disorder on which this book concentrates.

5.6 Further Study

Bach, D. R. and Dayan, P. (2017) offers a computational perspective on emotion that analyses its relationship with various aspects of appetitive and aversive utility.

Berridge, K. C. and Robinson, T. E. (2003) is part of a long series of arguments that there is an important separation between 'liking' (the hedonic components of reward) and 'wanting' (the motivational and learning force associated with reward that influences choice).

Hare, T. A., et al (2008) is an early paper using fMRI in humans to dissociate various different signals related to value.

Keramati, M. and Gutkin, B. (2014) argues that movements of the internal state relative to homeostatic optimality generate internal rewards.

5.7 Acknowledgements

I am grateful to the Gatsby Charitable Foundation for support and to Peggy Seriès for comments.

Chapter 6: Psychosis and Schizophrenia from a Computational Perspective

Rick A Adams, University College London

6.1 Introduction

Schizophrenia is a psychiatric disorder that affects around 0.5% of the population worldwide. Whilst it is less common than anxiety and depression, it can have more devastating effects: from its onset usually around 18-30 years, it can transform a person from being a university student to someone chronically unwell and dependent on social support for the rest of his/her life. It also carries the same risk of suicide as major depression. It is a 'psychotic' disorder, meaning that its sufferers' experience of reality departs from others' experience of reality in important and characteristic ways. Its diagnostic symptoms form three broad clusters, known as 'positive', 'negative' and 'disorganized'.

Positive symptoms include delusions and hallucinations; in schizophrenia, the former are commonly beliefs about being persecuted or surveilled, or beliefs that people or events refer to you or communicate messages to you in some way, or beliefs that one is controlling or controlled by other people or events, although there can be numerous other themes. Hallucinations can occur in any modality but the commonest in psychosis are auditory and verbal, i.e. voices. Whilst voice-hearing is not uncommon in the general population, voices referring to the subject in the third (rather than second) person, e.g. commenting on them or discussing them, especially in unpleasant ways, are more characteristic of schizophrenia. Symptoms of 'thought interference' are a group of experiences part way between hallucinations (i.e. abnormal sensory experiences) and delusions: e.g. that others are inserting thoughts into or extracting them from one's mind. Such symptoms are often accompanied by a loss of 'insight', i.e. a denial that these experiences might stem from an abnormal state of mind.

Negative symptoms refer to losses of normal function, including poverty of speech, reduced emotional expression, and, above all, a loss of motivation. They are distinct from depression, in which affect is very negative (i.e. the person feels very low and cries very easily, etc), in that affect is apparently reduced or absent. Disorganized symptoms are also known as 'thought disorder' and refer to abnormal structure in a person's speech or writing (as thoughts cannot be directly assessed). These may

be relatively mild, in the form of altered or new words (neologisms) or rather circumstantial answers to straightforward questions, or more substantial, e.g. sudden tangents or breaks in one's train of thought, or statements that are connected by bizarre or irrelevant associations, or severe, in which it is difficult to discern any meaningful content from an utterance.

Alongside positive, negative and disorganized symptoms, perhaps the commonest symptom of schizophrenia is of a generalized cognitive impairment: a loss of IQ of around 10-20 points (Meier et al. 2014). This decrement in cognitive function is hard to detect in a clinical interview, and so it does not form part of the diagnostic criteria (which were designed to maximize inter-rater reliability), but it seems fundamental to schizophrenia itself and poses a major public health problem, as returning to meaningful employment is often the biggest challenge for those diagnosed with the condition. Worse still, whilst antipsychotic drugs are reasonably effective for positive and disorganized symptoms, neither the cognitive impairment nor the negative symptoms have any effective medical therapy at present (although psychological interventions for both have been devised).

It should be stressed that the unitary diagnosis of 'schizophrenia' is unlikely to stand the test of time, although it seems equally unlikely to be replaced in the near future. Psychiatry is gradually moving away from categorical diagnostic systems and towards more dimensional approaches, as it becomes clear that other psychotic disorders such as bipolar affective disorder and schizoaffective disorder share not just some symptoms but also some genetic risk variance (and presumably neurobiological mechanisms, and psychosocial risk factors) with schizophrenia itself (Cross-Disorder Group of the Psychiatric Genomics Consortium 2013). Population surveys have also revealed that many psychotic symptom dimensions (e.g. positive, negative, cognitive and mood symptoms) are also continuous with the general population (Linscott and van Os 2010).

What part can Computational Psychiatry play in the future of schizophrenia research? It should make a major contribution to understanding how the different clusters of psychotic symptoms come about, by linking the biological, psychological and social risk factors for the disorder to the brain's function as a model of its physical and social environment. Such understanding would benefit our categorization of, diagnosis of and design of therapies for psychotic disorders.

6.2 Past and Current Computational Approaches

In describing the computational approaches to schizophrenia below, the negative, positive (and disorganized) and cognitive clusters will be considered in turn, as the models used to describe each are often quite different.

6.2.1 Negative symptoms

Negative symptoms can be grouped into two domains: those involving the loss of emotional expression (in affect and in speech) and those involving the loss of motivation for behavior; crucially, the latter predict functional outcome and quality of life. The fundamental question they pose can be stated as: "Why do these subjects not pursue policies that would result in outcomes that most people would find rewarding?"

One can easily see the relevance of reinforcement learning (RL) models (described in **Section 2.3** and **Chapter 5**) to answering this question (much more detailed accounts of reinforcement learning in schizophrenia include (Strauss, Waltz, and Gold 2014; Deserno et al. 2013) – what follows is a précis of this highly recommended work).

There are numerous potential RL-based explanations of why those with negative symptoms might not act to obtain rewards:

i) Because they underestimate the value of rewards. Interestingly, although 'anhedonia' (i.e. the loss of experience of pleasure) is listed as a negative symptom, subjects with schizophrenia actually show normal subjective and hedonic responses to rewards, so an explanation of negative symptoms is unlikely to be this straightforward (Strauss and Gold 2012).

ii) Because they learn more from negative feedback than positive feedback. If striatal dopamine release is disordered in schizophrenia, then one might expect greater difficulty in encoding positive reward prediction errors (RPEs) – via increased phasic dopamine release – than negative RPEs, via pauses in dopamine neuron firing. The consequence of this would be an asymmetry in RL, in which relevant stimuli tend not to be associated with rewards but can still be associated with punishments or loss of reward, perhaps causing a loss of motivation

for most actions over time. Such an asymmetry has indeed been demonstrated in subjects with high negative symptoms (Gold et al. 2012).

iii) Because they have difficulty building more accurate but complex models of the values of given actions. There are different ways of learning which action to take in a given situation: a simple way is using an actor-critic model to learn which actions are better- or worse-than-average, or a more complex way is to use Q-learning to learn the expected values of specific action-stimulus pairs. The latter is computationally more costly but can differentiate between stimuli that are rewarding and those that merely avoid loss. Gold et al. (2012) showed that subjects with high negative symptoms learned optimal actions like the simpler actor-critic model – which may indicate pathology in orbitofrontal cortex, where representation of expected values is thought to occur – whereas controls and schizophrenic subjects without negative symptoms were fit best by a Q-learning model.

iv) Because they have difficulty comparing the values of different stimuli or the costs and benefits of a given action. Subjects with schizophrenia describe inconsistent preferences when judging between two stimuli even outside a cognitively demanding learning-based task, implying that their representation of expected values and its use to make decisions is corrupted in the disorder (Strauss et al. 2011). Similarly, subjects with high negative symptoms are less likely to select high-cost high-reward actions (Gold et al. 2013), although whether this is due to problems in the valuation of reward or effort or the comparison of the two is unknown.

Of note, model-based fMRI studies of subjects with schizophrenia have also shown blunted ventral striatal activations to reward anticipation and RPEs, which in some cases correlate with the degree of negative symptoms (e.g. Juckel et al. 2006; meta-analyzed by Radua et al. 2015).

Motivational problems in schizophrenia show some similarities to anhedonia in major depression, in that in both disorders, basic reward experience and learning mechanisms seem largely preserved (e.g. hedonic responses to primary rewards and actor-critic reward learning), whereas inferences about – and hence affective responses to – more complex rewards are impaired.

6.2.2 Positive symptoms

Given that delusions seem *a priori* to relate to abnormal learning, and the strong association between presynaptic striatal dopamine availability (measured using positron emission tomography, PET) and positive symptoms (Howes and Kapur 2009), one might be optimistic that delusions could also be explained in terms of aberrant RL mechanisms. The first attempt to link dopamine 'hyperactivity', behavioural neuroscience and positive symptoms, however, was based not on RL but on the related field of motivational salience.

In the aberrant salience hypothesis, Kapur (2003) drew on Berridge and Robinson (1998)'s observations that some striatal dopamine innervation is crucial not for learning the values of stimuli, but for motivating responses to stimuli *whose values have already been learned*, a property they termed 'incentive salience'. Kapur proposed that in early psychosis there is an increased release of dopamine, including at inappropriate times (i.e. to stimuli with no expected value). This would generate a state of 'aberrant [incentive] salience' in the subject, in which various percepts, ideas or memories have great (but unwarranted) importance. As a consequence, delusions could arise as (rational) attempts by the subject to explain these bizarre experiences. Hallucinations could also be a direct consequence of percepts (e.g. inner speech) or memories being imbued with too much salience.

A Salience Attribution paradigm was devised to test Kapur's theory, and correlations between aberrant salience measures (speeding of reaction times to and/or incorrect beliefs about non-rewarding stimulus dimensions) and positive symptoms have been found. However, the most consistent finding in subjects with schizophrenia (and those with delusions in particular) is that of reduced explicit aberrant salience (i.e. altered belief updating), although implicit aberrant salience (i.e. altered motivational signalling) has also been found (J. P. Roiser et al. 2009; Jonathan P. Roiser et al. 2013; Smieskova et al. 2015; Abboud et al. 2016; Katthagen et al. 2018).

From a computational perspective, modelling aberrant salience is not straightforward, because the term 'salience' is now used to describe many different things: not just motivation signals (or, in more computational terms, average reward rate) but also unsigned RPEs (some dopamine neurons respond equally to rewarding and aversive PEs) and also either surprising or informative (Barto, Mirolli, and Baldassarre 2013) sensory states (unrelated to reward, but to which some dopamine neurons also respond). Interestingly, more evidence is emerging that dopamine neurons respond to changes in beliefs

independent of any reward prediction error (Corlett et al. 2007; Schwartenbeck, FitzGerald, and Dolan 2016; Nour et al. 2018), and indeed their role may be causal in this regard (Sharpe et al. 2017).

Maia and Frank (2016) produced a computationally simpler but still comprehensive account of how schizophrenia symptoms could arise from abnormal dopaminergic RPE signaling alone. They propose that negative symptoms could result from attenuated dopamine RPEs while positive symptoms could result from increased 'spontaneous' dopamine RPEs. Crucially, they observe that value and incentive salience depend mostly on dopaminergic signals in the limbic (ventral) striatum. However, in schizophrenia it is the associative striatum – where combined representations of states and actions may be learned – that is more consistently found to have increased dopamine synthesis and release (Howes and Kapur 2009). They therefore propose that delusions, hallucinations and otherwise bizarre and disordered thoughts could come about through abnormal gating of random percepts, thoughts or actions through the 'Go' pathway in associative striatum**.**

This account is admirable for its clarity and for its explanation of abnormal cognition alongside abnormal value learning, which the aberrant salience hypothesis struggles to account for. Additional hypotheses may be required to account for some other findings in schizophrenia, however. This includes in particular the considerable genetic and neuropathological evidence for *N*-methyl-D-aspartate receptor (NMDAR) hypofunction in the disorder, and various empirical findings which seem to relate more to NMDAR dysfunction than to dopaminergic abnormalities. One such finding (also see below) is that in schizophrenia that is resistant to (anti-dopaminergic) treatment, PET imaging has not found evidence of increased striatal dopamine synthesis, but magnetic resonance spectroscopy (MRS) has found evidence of cortical glutamatergic abnormalities (Demjaha et al. 2014). Nevertheless, there are complex interactions between NMDARs and the dopamine system (Klaas E Stephan, Friston, and Frith 2009) that careful empirical work is required to explore.

Hierarchical Bayesian predictive coding accounts of schizophrenia share Maia and Frank (2016)'s notion that PE signaling is aberrant in schizophrenia, but propose that it is not just RPE signaling in the striatum that is affected, but prediction error signaling throughout the cortex (Sterzer et al. 2018; Adams et al. 2013). This account is based on the idea that the brain uses (or approximates) Bayesian inference on its sensory inputs to infer their hidden causes in the environment (see **Section 2.4**). To do so it must use a hierarchical generative model of its sensations that encodes the sufficient statistics of the distributions over their causes – i.e. both their means and their precisions (inverse variances). The most

popular scheme for performing inference in such a model is predictive coding – in which higher levels pass predictions of activity down to lower levels, that return only errors to the higher levels, which correct their predictions, etc. – although other message passing schemes can be used (see below).

The key pathology in the predictive coding account of schizophrenia is proposed to be the encoding of precision of the signals related to incoming information (the likelihood) and prior beliefs (in the cortex and elsewhere). Given precision is used to weight one distribution over another in Bayesian inference, its neural substrate is likely to be synaptic gain (the factor by which an input to a neuron is multiplied to generate its output), which could likewise alter the influence (but not the content) of neural messages. Many neurobiological risk factors for schizophrenia affect synaptic gain, including neuromodulators such as dopamine and NMDARs, especially those on inhibitory interneurons which affect the oscillatory properties of networks and hence their ease of communication with other brain areas.

In schizophrenia it seems there is a hierarchical imbalance in synaptic gain, as primary sensory areas have been shown to be 'hyperconnected' (i.e. show increased correlation with other brain areas compared with controls) whereas higher regions (e.g. prefrontal and medial temporal cortex) are 'hypo-connected' (Anticevic et al. 2014). If this corresponds to a similar imbalance in the encoding of precision in a hierarchical model, then its effect would be to reduce the effect of priors on inference and cause larger belief updates in response to unexpected sensory evidence.

A loss of precision of prior beliefs could account for numerous phenomena in schizophrenia, including a resistance to visual illusions (which exploit prior beliefs to create their effects), impairments in smooth oculomotor pursuit of visual targets, abnormal electrophysiological responses to both predictable and oddball stimuli (e.g. the mismatch negativity) and a loss of attenuation of self-generated sensations (reviewed in Adams et al. 2013). Likewise, perceiving relatively inconsequential events as being imbued with significance (i.e. according too much precision to lower level PEs) and updating one's beliefs as a result fits comfortably with this framework. Indeed, it may be that these kinds of higher-level updates, encouraged by the loss of precision encoding in those areas, are the source (or the consequence) of the apparently spontaneous dopamine transients in the striatum.

It is unlikely that there is a uniform loss of precision of prior beliefs in schizophrenia, however: two recent studies have shown that some prior beliefs (about visual stimuli) have a *greater* influence over sensory data in subjects with schizophrenia or schizotypal traits compared with controls (Teufel et al.

2015; Schmack et al. 2013), although this is not always the case (Valton et al. 2019). In the auditory domain, there is evidence that prior beliefs about sounds learned during a task are more strongly weighted in hallucinators, with or without psychosis (Powers, Mathys, and Corlett 2017) and that this increased weighting may relate to striatal dopamine (Cassidy et al. 2018).

How these apparently opposite imbalances between prior beliefs and sensory evidence might co-exist in schizophrenia is an open question: there are numerous possible explanations that can only be resolved by empirical studies. For example, loss of precision in the middle (e.g. cognitive) levels of a hierarchy might allow both sensory evidence and higher level (e.g. affective) beliefs to dominate that level, causing sensory hypersensitivity and delusional ideas respectively. Or it may be that there is a loss of ability to optimally adjust synaptic gain according to context, rather than a persistent over- or under-estimation in any given area.

An alternative message passing scheme that could perform Bayesian inference (on discrete, not continuous, states – unlike predictive coding) is belief propagation, in which ascending and descending messages are not PEs and predictions but likelihoods and prior expectations respectively. Jardri et al. (2017) propose that a loss of inhibitory interneuron function in schizophrenia could allow ascending or descending messages to be passed back down or up the hierarchy, thus leading to 'overcounting' of either sensory evidence or prior beliefs. They called their model the Circular Inference model, in reference to the loopy amplifications caused by the impaired inhibitory interneurons in the hierarchy. They demonstrate evidence for both overcounting of sensory evidence and prior beliefs using a task in which subjects with schizophrenia had to update some preliminary knowledge in the light of new data, and, interestingly, find that on the group level, these subjects showed more evidence for ascending loops (i.e. overcounting sensory evidence), but individual subjects showed evidence for both ascending and descending loops which correlated with positive and negative symptom severity respectively (both correlated with disorganization).

The overcounting (or increased precision) of sensory evidence may contribute to a well described phenomenon in probabilistic belief updating in schizophrenia: the 'jumping to conclusions' (JTC) bias, which is also associated with the presence of delusions (Dudley et al. 2016). This bias is usually assessed with the urn or beads task (Garety, Hemsley, and Wessely 1991), in which subjects are shown two jars containing red and green beads in ratios of 80:20 and 20:80. The jars are hidden and a sequence of beads is drawn (with replacement) from one jar; the subject either has to stop the sequence when they

are sure of the jar's identity (the 'draws to decision' version) or rate the probability of the jar after seeing each bead (the 'probability estimates' version).

The best-replicated finding in this literature (Dudley et al. 2016) is that many more subjects with schizophrenia than controls decide on the jar in the 'draws to decision' task after seeing only one or two beads (the 'jumping to conclusions' bias). There are many other computational parameters, aside from sensory overcounting or precision, which could account for this effect, however: a lower decision threshold, an inability to inhibit a prepotent response, more stochastic decision-making (i.e. higher decision 'temperature'), a lower perceived cost of making a wrong decision, or a higher perceived cost of sampling. Unfortunately, most 'draws to decision' paradigms have not controlled or manipulated these parameters, so it is not possible to distinguish conclusively between them.

Moutoussis et al. (2011) explored whether the last three parameters listed above – i.e. decision temperature τ, cost of wrong decision C_W or cost of sampling C_S – could explain the jumping to conclusions bias in schizophrenic subjects with or without active psychosis. The authors found that acutely psychotic subjects had much higher τ but no differences in C_W and C_S. In a subsequent study, first episode psychosis subjects were found to have a higher C_S and only a borderline increase in τ (Ermakova et al. 2019). Note that a higher τ may mean that decisions are truly more stochastic, or it may mean that the source of variability in decision-making has not been captured by the model.

One weakness of this model is that its τ, C_W and C_S parameters don't allow for individual differences in belief updating: it assumes that all subjects update their beliefs in a Bayes-optimal fashion. However, there is evidence that subjects with schizophrenia update their beliefs differently to controls (neither of whom are Bayes-optimal). A scrupulous and well-controlled study of the beads task recently showed that whilst the main effect of schizophrenia diagnosis was more liberal (i.e. larger) belief updates, delusions were correlated with more conservative (i.e. *smaller*) belief updates – contrary to most interpretations of the 'jumping to conclusions' bias (Baker et al. 2019).

In a similar vein, Averbeck et al. (2010) asked subjects with schizophrenia and controls to perform a sequence learning task with probabilistic feedback. This allowed them to estimate how much subjects learned from positive and negative feedback. The schizophrenic subjects learned *less* from positive feedback than controls, mirroring their reward-learning deficits described in the preceding section. Moreover, the less they learned, the *more* likely they were to show the jumping to conclusions bias. This apparently paradoxical finding is explored in greater detail in the next sections.

Although a significant amount of work has been done on belief-updating and value-learning in schizophrenia, there has been relatively little exploration of how language could be spontaneously created (as in auditory verbal hallucinations) or become disorganised, both in terms of its form (e.g. derailment – one subject changing into another without an obvious connection) or its content (e.g. attributing events to bizarre agents, such as famous people). In some pioneering studies, Hoffman and McGlashan (2006) showed that excessive pruning of connections and hypodopaminergia (i.e. disinhibition) in the hidden layer of a sentence recognition network could reproduce speech detection performance in human hallucinators, and that this excessive pruning could also lead to hallucinations (although these hallucinations were only of a single word appended to an existing sentence).

Hoffman et al. (2011) trained a more complex model comprised of multiple connected modules, each containing recurrent networks, to learn 28 narratives of varying emotional intensity and about different agents. They showed that of many possible perturbations, only enhancing PE-learning (i.e. increasing backpropagation learning rates) during memory encoding matched errors made in schizophrenic subjects' memories for narratives. These included exchanging the identities of agents (especially of similar social status) between autobiographical and other stories and derailments from one story to another, particularly between those of similar type or emotional valence.

6.2.3 Cognitive symptoms

One implication of Hoffman's work is that abnormalities of working memory (WM) and memory encoding processes may not just manifest in those processes, but also contribute to positive symptoms. In a landmark study, Collins et al. (2014) demonstrated that WM deficits could also contribute to apparent RL impairments in schizophrenia. Their subjects had to learn stimulus-response associations for reward, and the stimuli were presented in sets of size two to six. Under these conditions, smaller sets could possibly be learned through WM processes, but larger sets – exceeding WM capacity – would be more reliant on incremental (i.e. RL) mechanisms. Fitting an RL model with a WM component to individuals' data, they found that the subjects with schizophrenia had lower WM capacity and greater WM decay rate, but their RL and decision stochasticity parameters were no different to controls'. This demonstrates that unless WM is explicitly modelled, inferences about RL parameters in schizophrenia

must be treated with caution. How WM relates to symptoms is unclear, though, as none of the model parameters or their principal components correlated with positive or negative symptoms.

In neurobiological terms, this implies that pathology in prefrontal cortex (PFC) and hippocampus might make a greater contribution than the striatum to abnormal inference and learning in schizophrenia. But what kind of pathology? A highly influential spiking network model of pyramidal cell and interneuron function in PFC during a spatial WM task contains excitatory pyramidal cells with bidirectional connections to a single inhibitory interneuron and recurrent excitatory connections to themselves (see also **Chapter 3**). Increased activity of one pyramidal cell is therefore a) self-sustaining, through the E-E connection, and b) laterally inhibiting, through the E-I connection and subsequent inhibition of its neighbouring pyramidal cells. These dynamics can be pictured as energy landscapes containing 'basins' of attraction, the stability of which is determined by their depth and the level of 'noise' in the network (Rolls et al. 2008). NMDAR hypofunction on pyramidal cells would reduce E-E strength and also self-sustaining activity. On the other hand, NMDAR hypofunction on interneurons would reduce E-I strength and increase the spread of excitation through the network. A model in which E-I strength is reduced more than E-E (i.e. an increased E/I ratio) captures the behaviour of subjects with schizophrenia best: it increases 'false alarms' to near (but not far) distractors during a spatial WM task (due to lateral spread of excitation) but not the rate of 'misses' (Murray et al. 2014) (See also **Chapter 3**).

Other models of PFC function have also incorporated the recently-demonstrated cortical dopamine hypo-function in schizophrenia (Slifstein et al. 2015). Dopamine hypo-function reduces activity in both pyramidal cells (via D1 receptors) and interneurons (via unique excitatory D2 receptors) in adult rats (O'Donnell 2012). Modelling studies suggest that this should exacerbate any NMDAR hypo-function and make PFC networks even more vulnerable to distraction (Durstewitz and Seamans 2008). An early connectionist model of dopamine's effects on gating inputs (e.g. sensory cues) into PFC proposed that greater variability of dopamine firing makes cues' effects on PFC less reliable (Braver, Barch, and Cohen 1999). There is clearly much still to be learned about cortical-dopaminergic interactions.

In the previous section we encountered the puzzling finding that in subjects with schizophrenia, the jumping to conclusions bias correlates with a *reduced* tendency to learn from positive feedback (Averbeck et al. 2010). In addition, in the 'probability estimates' versions of the beads task, numerous groups have demonstrated a 'disconfirmatory bias' in schizophrenia, i.e. a tendency to update more than

controls on receipt of evidence *against* one's current hypothesis (Garety, Hemsley, and Wessely 1991; Peters and Garety 2006; Fear and Healy 1997; Young and Bentall 1997). However, when observing patients' behaviour in this task, it appears that they update more to both a 'disconfirmatory' bead *and* to the following bead, e.g. R-R-R-G-R (Langdon, Ward, and Coltheart 2010; Peters and Garety 2006). Yet like Averbeck and colleagues, others have observed *decreased* updating in patients to more consistent sequences both in this task (Baker et al. 2019) and in stimulus-reward learning tasks, especially in patients with more negative symptoms (Gold et al. 2012). Likewise, healthy volunteers given ketamine (an NMDAR antagonist used to model psychosis in humans) show a decrement in updating to consistent stimulus associations (Vinckier et al. 2016).

To summarize, it appears that compared with controls, subjects with schizophrenia may show greater belief updating in more uncertain contexts, but (sometimes) lower belief updating in less uncertain contexts, rather than a straightforward 'disconfirmatory bias'. These effects make sense in the light of attractor models of cortical function, in that NMDAR hypofunction on both pyramidal cells and inhibitory interneurons (to a greater extent) could both reduce recurrent excitation but also increase the E/I ratio. An increase in E/I ratio has been shown to cause more rapid updating and impulsive decision making in a perceptual task model (Lam et al. 2017). On the other hand, a reduction in recurrent excitation could reduce attractor stability (Rolls et al. 2008) and hence make it hard to reach maximum confidence in any one decision. This attractor hypothesis motivated the recent study described below.

6.3 Case Study Example: Attractor-like dynamics in belief updating in schizophrenia

Adams et al. (2018) tested Bayesian belief updating models on 'probability estimates' beads task data obtained from both healthy volunteers, subjects with schizophrenia, and psychiatric controls. Dataset 1 was published previously (Peters and Garety 2006) and comprised 23 patients with delusions (18 diagnosed with schizophrenia), 22 patients with non-psychotic mood disorders, and 36 non-clinical controls, 53 of whom were also tested again once the clinical groups were no longer acutely unwell. Dataset 2 was newly acquired and comprised 56 subjects with a diagnosis of schizophrenia and 111 controls. Subjects in dataset 1 performed the 'probability estimates' beads task with two urns with ratios of 85:15 and 15:85 green and red beads respectively (**Figure 6.1**, upper panel). They had to view a

single sequence of ten beads. After each bead, they had to mark an analogue scale (from 1 to 100) denoting the probability that the urn was the 85% red. Subjects in dataset 2 performed the same task with two urns with ratios of 80:20 and 20:80 red and blue beads respectively (**Figure 6.2**, lower panel). Each subject viewed four separate sequences of ten beads (an A and a B sequence, and A and B again but with the bead colours inverted). After each bead, they had to mark a Likert scale (from 1 to 7) denoting the probability that the urn was the 80% blue one. Two sequences contained an apparent change of jar.

The behavioral differences between groups are detailed in Adams et al. (2018). In brief, subjects with schizophrenia showed increases in 'disconfirmatory' updating in both datasets, although this effect was diminished at follow-up in Dataset 1.

< insert Figure 6.1 around here>

The Bayesian belief updating model was the Hierarchical Gaussian Filter or HGF (Mathys et al. 2011; see **Section 2.4.6**). The HGF contains numerous parameters that can vary between subjects, thus explaining individual differences in inferences whilst preserving the Bayes-optimality of inferences, given these parameters. The HGF has been used to demonstrate numerous interesting parameter differences between controls and groups with psychiatric diagnoses, e.g. ADHD (Hauser et al. 2014), autism (Lawson, Mathys, and Rees 2017) and schizophrenia (Powers, Mathys, and Corlett 2017). The models employed here are described in detail in Adams et al. (2018). In brief, the model's inputs are the bead shown $u^{(k)}$ and the subject's response $y^{(k)}$ on trials k = 1-10. From these inputs and its prior beliefs, the model infers the subject's beliefs about the jar $\mu_1^{(k)}$ (a logistic sigmoid function of the 'tendency' of the jar $\mu_2^{(k)}$) on each trial, and the model parameters (β, ω, φ or κ_1 and $\sigma_2^{(0)}$ – see below, **Table 6.1**, and **Figure 6.1**). The response model generates the subject's response $y^{(k)}$ (i.e. where on the sliding scale they place the arrow on trial k), which is determined by $\mu_1^{(k)}$ and the precision of their response β (similar to inverse temperature, i.e. $1/\tau$) which affects how much $y^{(k)}$ can deviate from $\mu_1^{(k)}$ – i.e. the (inverse) stochasticity of their responding, given their beliefs.

Using the HGF and the two datasets, the following questions were addressed: can differences in belief updating in schizophrenia compared with controls be explained by: i) group differences in general

learning rate ω; ii) differences in response stochasticity β, or by additional parameters encoding: iii) the variance of beliefs about the jars at the start of the sequence $\sigma_2^{(0)}$; or iv) a propensity φ to overweight disconfirmatory evidence specifically, or v) a parameter κ_1 that simulates unstable attractor states, making it easier to shift from believing in one jar to the other (**Figure 6.1**, lower panel)?

Furthermore, are these findings consistent between different groups of schizophrenia, or within schizophrenia tested at different illness phases, and are they unique to schizophrenia or also present in other non-psychotic mood disorders?

< insert Table 6.1 around here>

Six models were tested, all containing ω and β, and either φ, or κ_1, or neither (each with or without σ_2) – see **Table 6.1**. Full details of the models, statistical and behavioural results are given elsewhere (Peters and Garety 2006; Adams et al. 2018). Bayesian model selection for dataset 1 at both baseline and follow-up and dataset 2 produced identical results: Model 6 won in each case. In studies of schizophrenia, it is often the case that many patients are fit best by a different model to controls; usually a much simpler one, e.g. (Moutoussis et al. 2011; Schlagenhauf et al. 2013). Performing model selection within each group separately, however, still found that Model 6 best accounted for the data in all groups (**Figure 6.3**).

In dataset 1 at baseline, there were large group differences in the attractor instability κ_1 and response stochasticity β but not in the initial variance $\sigma_2^{(0)}$ or the learning rate ω (**Figure 6.4**, upper row): κ_1 was significantly larger in the non-clinical controls and the psychotic group than in the clinical control group, and β was smaller in these groups .

In dataset 1 at follow-up (**Figure 6.4**, middle row), the attractor instability κ_1 remained larger and response stochasticity β smaller in the psychotic group than the non-clinical control group but now the clinical and non-clinical control groups were no longer significantly different. Similarly, in dataset 2, κ_1 was significantly higher and β was lower in schizophrenia than in controls. There were no significant group differences in ω or $\sigma_2^{(0)}$ (**Figure 6.4**, lower row). The model fits for two example subjects are shown in **Figure 6.5**.

Neither κ_1 or β in dataset 1 at baseline were predicted by any particular subgroup of (positive, negative or affective) symptoms. In dataset 2, there was only a weak relationship between β and negative symptoms.

We tested for correlations between the Model 6 parameters: κ_1 and β were negatively correlated both at baseline and at follow in dataset 1, and in dataset 2. Note that if parameters are highly correlated, then it can be impossible to estimate them reliably. 200 datasets were therefore simulated using the HGF and the modal parameter values for the control and schizophrenia groups in dataset 2, and then the parameters from these simulated datasets were re-estimated in order to check we could estimate them reliably (parameter recovery, see **Section 2.5**). With the exception of $\sigma_2^{(0)}$ in the simulated schizophrenia dataset, the estimated parameter values closely matched their original values.

In summary, this study showed that in computational models of two independent datasets, all subjects – including subjects with schizophrenia – are best fit by a model simulating the effects of attractor state dynamics on belief updating (Model 6) rather than a model biased towards disconfirmatory updating alone (Model 4). Medium-to-large differences were found between subjects with schizophrenia and controls in both datasets in both the attractor instability parameter (κ_1 was greater in schizophrenia, i.e. more unstable) and the stochasticity of responding (β was smaller, i.e. noisier, in schizophrenia), and κ_1 correlated with β in both datasets. Furthermore, ν correlated with κ_1 but not with ω or $\sigma_2^{(0)}$ in all three experiments, supporting the idea that β is measuring a stochasticity that is related to the attractor unstability κ_1 by an underlying neurobiological process, rather than an effect that just isn't described by the model.

These findings are important because they connect numerous reasoning biases previously found in schizophrenia – e.g. a disconfirmatory bias, increased initial certainty (Peters and Garety 2006), and decreased final certainty (Baker et al. 2019) – with model parameters that describe how non-linear belief updating in cortex could be caused by unstable and noisy attractor states. (In this context, 'non-linear' refers to updating that isn't uniformly increased or decreased relative to controls, e.g. updating more to surprising evidence but less to unsurprising evidence).

Indeed, two recent studies of similar tasks in populations with schizophrenia have also demonstrated evidence of similar belief updating. Jardri et al. (2017) showed that the patients with schizophrenia on average "overcount" the likelihood in a single belief update. Jardri et al attribute this effect to disinhibited cortical message-passing, but it could equally be attributed to attractor network

instability. Stuke et al. (2017) showed in a very similar task that all subjects showed evidence of non-linear updating, but the group with schizophrenia updated more than controls to "irrelevant information" (i.e. disconfirmatory evidence).

NMDAR hypofunction could contribute to an increased tendency to switch between beliefs and increased stochasticity in responding in several ways (Rolls et al. 2008): i) by reducing inhibitory interneuron activity, such that other attractor states are less suppressed when one is active, ii) by reducing pyramidal cell activity, such that attractor states are harder to sustain, and iii) by reducing the NMDAR time constant, making states more vulnerable to random fluctuations in neural activity.

Another important aspect of dataset 1 is the finding that κ_1 and β were also significantly different between the mood disorder clinical group and non-clinical control groups when the former were unwell, but not at follow-up, whereas the differences between the schizophrenia and non-clinical controls remained. This is interesting in light of past work indicating that neuromodulatory activity can have similar impacts on prefrontal network dynamics to NMDARs (Durstewitz and Seamans 2008). One might speculate that both the group with schizophrenia and clinical controls are affected by neuromodulatory changes when unwell, but only the former has an underlying NMDAR hypofunction that is still present once the acute disorder has resolved.

One might question why, given these relationships between parameters and cognition, there weren't strong relationships between κ_1 or β and positive or negative symptom domains (negative symptoms were weakly predictive of β in dataset 2 only). One reason may be that the symptom analyses – conducted only on patients – were underpowered, but it is also possible that other pathological factors contribute to symptoms, beyond those measured here (e.g. striatal dopamine availability and positive symptoms). Of note, another study demonstrating clear WM parameter differences between subjects with schizophrenia and controls also failed to detect any relationship between those parameters and symptom domains (Collins et al. 2014).

An important future challenge will be to link belief updating parameters to those of spiking network models, to understand how NMDAR function on both pyramidal cells and inhibitory interneurons and neural 'noise' contribute to attractor instability, response stochasticity and inference in general (Lam et al. 2017; Soltani and Wang 2010). Beyond that, a true understanding of the disorder will probably emerge once we gain a better understanding in computational terms of how the thalamus, striatum and cortex all interact with each other, and with dopamine, in performing inference.

6.4 Chapter Summary

In this chapter, we have covered the positive, negative and cognitive symptoms of schizophrenia and attempts to model them in computational terms. Negative symptoms – broadly, the failure to act to obtain reward – have been modelled using RL models with some success. Positive symptoms – delusions and hallucinations – are less straightforward to understand. Attempts to model them have concentrated either on abnormal dopamine signaling in striatum or abnormal synaptic gain at both ends of the cortical hierarchy. Abnormal dopamine signaling would contribute to delusional thoughts (through aberrant salience, aberrant RPEs or aberrant gating of thoughts). Abnormal synaptic gain would lead to an imbalance (or alternatively a loss of adaptability) in the encoding of precision in the brain's model of the world, such that prior beliefs are underweighted, and sensory evidence is overweighed. Such models have not yet given a full account of positive symptoms, however. There is a sizable literature on psychological biases (e.g. jumping to conclusions) in schizophrenia, and these are beginning to be understood in modelling terms. Relevant to this may also be the modelling of cognitive symptoms – e.g. concentration and working memory problems – using spiking network models with attractor dynamics. In such models, NMDAR hypofunction, perhaps resulting in an increased E/I ratio (due to disinhibition), can make attractors unstable and easily affected by random fluctuations in neural firing. These changes can explain spatial working memory performance in subjects with schizophrenia, as well as apparent biases in probabilistic inference. Ultimately, to understand schizophrenia, we will need a deeper understanding of how the thalamus, striatum and cortex all interact with each other, and with dopamine, in performing inference.

6.5 Further Study

Strauss, Waltz, and Gold (2014) provide an excellent summary of RL models of negative symptoms.

Maia and Frank (2016) offer the most detailed and developed account of dopamine's potential contributions to positive and negative symptoms.

Adams, Huys, and Roiser (2015) contains a simplified version of the hierarchical predictive coding account of schizophrenia: for more equations and models, see Adams et al. (2013).

Rolls et al. (2008) – an excellent review of spiking and neural network models and how they relate to a dynamical system view of schizophrenia, containing unstable attractor states, etc. For more on this theme, see also **Chapter 3** of this volume.

Collins et al. (2014) is a first rate behavioral modelling paper, demonstrating the importance of including WM function in RL models, as performance in schizophrenia is explained by pathology in only the former.

Chapter 7: Depressive Disorders from a Computational Perspective

Samuel Rupprechter[1], Vincent Valton[2], Peggy Seriès[1]

1. University of Edinburgh, UK. 2. University College London, UK

7. 1 Introduction

Depression and anxiety disorders are the two most common psychiatric disorders around the world (Alonso et al. 2004; Ayuso-Mateos et al. 2001; Üstün et al. 2004; Vos et al. 2012) and display a high level of comorbidity: patients suffering from one of these illnesses are often affected by the other one as well (Kessler et al. 2003). In the United States, Kessler et al. (2003) estimated the lifetime prevalence of major depressive disorder (MDD) at over 16%. Similar figures have been reported for Europe at 13% (Alonso et al. 2004).

Diagnosis for MDD is commonly based on the Diagnostics and Statistical Manual of mental disorders (DSM-V; American Psychiatric Association (2013)). The manual lists two core symptoms of MDD: *depressed mood* and *loss of interest or pleasure* (anhedonia), of which at least one has to be present for diagnosis. Other symptoms include *a significant change in weight*, *insomnia*, *hypersomnia*, *psychomotor agitation or retardation*, *fatigue* or *loss of energy*, *feelings of worthlessness or guilt*, *a diminished ability to think or concentrate*, and *recurrent thoughts of death or suicide*. Overall, five or more symptoms have to be present for at least two weeks, cause significant impairments in important areas of daily life, and should not be better explained by other psychiatric disorders. The International Classification of Diseases (ICD-10; World Health Organization (1992)) has similar criteria for diagnosis of (single) *depressive episodes* and *recurrent depressive disorder*.

Strikingly, according to the DSM definition, it is possible for two people to receive the same diagnosis of MDD without sharing a single symptom. One MDD patient may experience depressed mood, weight gain, constant tiredness and fatigue, and regularly think about ending their life. Another MDD patient

may experience anhedonia, lose a lot of weight, and go through psychomotor and concentration difficulties while being unable to sleep properly. The existence of these non-overlapping profiles partly stems from the fact that categories and symptoms of depression originated from clinical consensus and do not necessarily have a basis in biology (Fried et al. 2014). As a consequence, research often focuses on individual symptoms - for example anhedonia (Pizzagalli (2014); see also our case study) - in addition to categorical group differences. In the clinical and drug trial literature, Hamilton Depression Rating (HRSD-17) and MADRS are by far the most important rating scales. In research environments, the Beck depression inventory (BDI; Beck et al. 1961) is a popular choice to measure overall depressive severity and a sub-score can be extracted from items of the questionnaire to quantify anhedonic symptom severity.

Cognitive neuroscience of depression

Patients often show deficits on a broad range of tasks probing executive function and memory (Snyder 2013; Rock et al. 2014), and impairments often remain (to some degree) after remission (Rock et al. 2014).

An early influential theory, inspired by a wealth of animal studies, is that of learned helplessness (Seligman 1972; Maier and Seligman 1976; Abramson, Seligman, and Teasdale 1978). The theory suggests that continued exposure to aversive (stressful) environments over which animals do not have any control lead to behavioral deficits similar to those observed in depression. In such a framework, the patients' distress is believed to stem from their perception of a lack of control over the environment and ensuing rewards or penalties. This, in turn, could explain patients' distress and lack of motivation to initiate actions. Stress has been proposed as a mechanism for memory impairments in depression (Dillon and Pizzagalli 2018) and Pizzagalli (2014) hypothesized that dysfunctional interactions of stress with the brain reward system can lead to anhedonia.

An alternative influential theory about depression concentrated on the prevalence of negative biases involved in the development and maintenance of depression (Beck 2008), which led to the emergence of cognitive behavioral therapy (CBT). This line of research hypothesized that negative schemas about the self, the world, and the future would form due to adverse childhood experiences. According to this framework, negative schemas could lead patients to downplay the magnitude of positive events, or

attribute negative valence to objectively neutral events. Patients would effectively perceive the world through "dark tainted" glasses.

It has been suggested that negative biases play a *causal* role in the development and maintenance of depression (Roiser, Elliott, and Sahakian 2012) and that antidepressant medications target these negative biases rather than targeting mood directly (Harmer, Goodwin, and Cowen 2009).

Recently, much cognitive research has focused on decreased sensitivity to reward in depression. There are at least two important reasons for this focus: First, reward processing appears to align with a lack of interest or pleasure (anhedonia), a core symptom of depression and one to which we will come back again in the case study of this chapter. Second, reward processes are better understood than mood processes, both at the neurobiological and at the behavioral level. Indeed, cognitive neuroscience has started to dissociate and delineate different sub-domains of reward processing, which can be studied independently in relation to anhedonia (Treadway and Zald 2013). For example**,** "incentive salience" ("desire" or "want") can be distinguished from "motivation" and "hedonic response" (enjoyment) and we may want to independently study the association of each of these sub-domains with depression. For instance, your driving attention and focus on a piece of chocolate (a potentially rewarding stimulus) is different from how much you enjoy that piece while you are eating it. These two subdomains may also be independent from your willingness to expand effort to obtain that piece of chocolate.

Cléry-Melin et al. (2011) tested depressed patients and healthy controls on a task in which they could exert physical effort (through grip force on a handle) to attain monetary rewards of varying magnitudes. They found that depressed participants did not exert more physical effort to obtain higher rewards (as opposed to lower rewards). However, they *believed* they had exerted more effort for higher rewards, as evidenced by their higher effort ratings. Controls, on the other hand, objectively exerted more effort for greater rewards, but reported subjectively reduced effort ratings for higher rewards compared to lower rewards. In another study (Treadway et al. 2012), participants were able to obtain varying amounts of money if they managed to make a large number of button presses within a short time window. Depressed patients exerted less effort (made less button presses) than controls in order to obtain reward. Together these studies suggest that depression, and anhedonia in particular, may be related to impairments in the motivation and willingness to exert effort for rewards. This may also explain why

behavioral activation therapies have been reported to work well for depressed patients[12]: these practices specifically target decreased motivation (Treadway et al. 2012).

Overall, there is large overlap between different theories of depression. Most cognitive theories place a large emphasis on biases influencing emotional processing (Gotlib and Joormann 2010), but some differ in their explanation of the development of these biases; for example whether they develop in response to early stressful life experiences (Beck 2008, Pizzagalli 2014) or stem from biased perceptual and reinforcement processes (Roiser, Elliott, and Sahakian 2012).

Several neurotransmitters, most commonly serotonin and dopamine, are implicated in reward and punishment processing in depression (Eshel and Roiser 2010). Dopamine is heavily implicated in reinforcement learning processes (Schultz 2002) and has consistently been associated with depression in humans and animals (Pizzagalli 2014). Serotonin has long been implicated in the processing of aversive stimuli and learned helplessness and depression may be related to a failure of stopping such aversive processes (Deakin 2013). Antidepressant medications commonly work by altering serotonin levels (Eshel and Roiser 2010). Neuroimaging studies have revealed abnormal activation and connectivity of many cortical and subcortical brain regions in depression (Pizzagalli 2014, Chen et al. 2015). Reporting of blunted striatal response to reward in MDD has been particularly consistent (Pizzagalli 2014, Arrondo et al. 2015). The orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (vmPFC) are implicated in the representation of internal values (Chase et al. 2015). Depression is associated with abnormal activation in these regions (Pizzagalli 2014, Cléry-Melin, Jollant, and Gorwood 2018), possibly related to abnormal use of reward values during decision-making (Rupprechter et al. 2018). Large meta-analyses have concluded that MDD is associated with reduced hippocampal volume (Schmaal et al. 2016) and alterations in cortical thickness, especially in OFC (Schmaal et al. 2017).

7.2 Past and current computational approaches

A variety of different computational approaches, ranging from connectionist and neural networks, to drift diffusion models, reinforcement learning and Bayesian decision theory, have been used to study the behavior of MDD patients. We will, in turn, briefly describe findings from each of these approaches.

[12] How such psychological therapies can be applied successfully in real clinical environments is still debated, however.

7.2.1 Connectionist Models

One early approach that has been used to model depression is a connectionist approach, which is inspired by the idea that complex functions can naturally arise from the interaction of simple units in a network (see **Section 2.1**).

Siegle, Steinhauer, and Thase (2004) asked groups of depressed and healthy individuals to perform a Stroop color naming task. In this task, color words are presented on each trial with different ink colors matching or not matching the word (e.g. the word "red" written in blue ink), and participants have to name the ink color while refraining from reading the word itself (**Figure 7.1**). The task is typically used to probe attentional control. Pupil dilation measurements were used as an indicator for cognitive load, because pupils reliably dilate under cognitively demanding conditions (Siegle, Steinhauer, and Thase 2004). Previous studies had shown impairments within groups of depressed subjects, but the nature of these impairments varied, with patients sometimes showing slower responses and other times increased error rates. Siegle, Steinhauer, and Thase (2004) found similar performance patterns for the two groups, but differences in pupil dilation. Depressed individuals showed decreased pupil dilation, consistent with decreased cognitive control. A neural network was used to identify possible mechanisms that could have resulted in these group differences. The modelling suggested that decreased prefrontal cortex activity could lead to the observed cognitive control differences in this experiment. Such a disruption might also explain attentional deficits commonly observed in depression (Siegle, Steinhauer, and Thase 2004).

< insert Figure 7.1 around here>

Siegle and Hasselmo (2002) provided another example of how neural network models can be used to better understand deficits in depression during (negatively biased) emotional information processing. The task considered was one where emotional word stimuli were observed, which participants had to label as positive, negative, or neutral. Patients typically show biases in emotional information processing, for example quicker responses to negative information (Siegle and Hasselmo 2002). A neural network model was used to simulate classification of emotional stimuli. It could reproduce the typically observed behavior of depressed patients: it was quicker to identify negative information than

positive information and showed larger sustained activity when confronted with negative words. Different mechanisms could lead to these observed abnormalities in the network, including over-learning of negative information, which can be related to rumination, i.e. the tendency to repetitively think about the causes, situational factors, and consequences of one's negative emotional experience. A network that had over-learned on negative information could be retrained using positive information (akin to a cognitive behavioral therapy), which resulted in the normalization of network activity in response to negative information. The longer the network had "ruminated", the longer it took for the "therapy" (i.e. retraining) to work, providing insights into the recovery from depression using CBT and its interactions with rumination. Siegle and Hasselmo (2002) therefore suggested that rumination can be predictive of treatment response and should be routinely assessed in depressed individuals.

7.2.2 Drift Diffusion Models

Drift diffusion models (DDMs; see **Section 2.2**) have also been used to better understand the mechanisms underlying depressive illness. These models are especially useful when the modelling of reaction time and accuracy *in combination* is of primary interest.

For example, Pe, Vandekerckhove, and Kuppens (2013) modelled behavior on the emotional flanker task to analyze negative biases in depression. In this task, participants are shown a positively or negatively valenced word that they are asked to classify according to valence. The central stimulus is flanked by two additional words with positive, negative or neutral valence (**Figure 7.2**). The authors hypothesized that higher depressive symptomatology and rumination (as measured by self-report questionnaires) are related to negative attentional biases (i.e. a bias towards negative target words). Classical analyses showed that the higher the rumination score, the stronger the facilitation effect (computed from accuracy scores) of negative distracters on negative targets and the weaker the facilitation effect of positive distracters on positive targets. After controlling for depression, only the former effect remained. A DDM analysis on the other hand revealed more effects involving the drift rate, which corresponds to the rate at which information is being processed. The drift rate was negatively correlated with rumination scores on trials where a negative target word was flanked by positive words and was positively correlated with rumination scores on trials where negative words flanked a negative or positive word. After controlling for depression scores, rumination still predicted attentional bias for negative information, but depression scores were no longer predictive after

controlling for rumination. The computational modelling therefore revealed that rumination was associated with an enhanced processing of words flanked by negative words and decreased processing in the presence of positive flankers.

< insert Figure 7.2 around here>

In addition to negative biases, depression is also associated with impairments in executive function (Snyder 2013). Dillon et al. (2015) used a combination of three drift diffusion processes to account for behaviour on a different (non-emotional) version of the flanker task**.** In this version, stimuli and distracters were three arrows pointing left or right. The central and flanking arrows could either be congruent (pointing in the same direction) or incongruent. Depressed and healthy participants had to indicate the direction of the arrow in the middle. The authors' goal was again to address inconsistent findings of previous studies, which had sometimes found enhanced executive functioning in depression during tasks that demand careful thought. Depression can lead to increased analytical information processing (c.f. rumination), which results in worse performance during tasks requiring fast decisions but can also lead to increased accuracy when a careful approach is necessitated and when reflexive responses need to be inhibited. Dillon et al. (2015) found that depressed participants were more accurate but slower than controls on incongruent trials. They decomposed behavior on the flanker task into three different mechanisms that might be affected by depression, and which were modelled by separate drift diffusion processes: (1) a reflexive mechanism biased to respond according to the flankers, (2) a response inhibition mechanism able to suppress the reflexive response, and (3) executive control responsible for correct responses in the presence of incongruent flankers. The analysis of model parameters showed that the drift rate for the executive control mechanism was lower in depression, which on its own would lead to slower, but also less accurate responses. However, this executive control deficit was offset by an additional decreased drift rate in the reflexive mechanism. This could explain impaired executive function but highly accurate responses in MDD (Dillon et al. 2015).

One more example comes from Vallesi et al. (2015), who used DDMs to better understand deficits in the regulation of speed-accuracy trade-offs in depression. At the beginning of each trial, a cue signaled whether participants should focus on speed or accuracy. It was found that MDD patients, unlike controls, adjusted their decision threshold based on the instructions for the *previous* trial, with speed instructions decreasing the decision boundary (independently of the cue for the current trial). That is, patients had difficulties overcoming instructions from the previous trial and flexibly switching between

fast and accurate decision-making. In addition, drift rates within the patient group were generally lower than in the control group, indicating a slowing down of cognitive processing, which is commonly found in MDD patients.

7.2.3 Reinforcement Learning Models

In reinforcement learning models, behavior is captured on a trial-by-trial basis. An agent makes a decision based on some internal valuation of the objects in the environment, observes an outcome, and then uses this outcome to update the internal values (see **Section 2.3** and **Chapter 4**). There exists substantial behavioral and neural evidence, often supported by computational modelling, for impaired reinforcement learning during depression (see Chen et al. 2015 for a review).

Chase et al. (2010) fitted a Q-learning model to the behavior of MDD patients and healthy controls on a probabilistic selection task. On each trial, one of three possible stimulus pairs was displayed and participants had to choose one of the stimuli, which were followed by positive or negative feedback according to different probabilities. They did not find evidence for their initial hypothesis that patients would preferentially learn from negative outcomes due to a tendency in depression to focus on negative events. Participants' anhedonia scores, however, negatively correlated with positive and negative learning rate as well as the exploration-exploitation (softmax) parameter. The study therefore provided evidence that depression, and specifically anhedonia, is related to altered reinforcement learning.

Huys et al. (2013) performed a meta-analysis on the Signal Detection Task (Pizzagalli, Jahn, and O'Shea 2005). In contrast to the previous study, they concluded that anhedonia is principally associated with blunted sensitivity to reward as opposed to an impaired ability to learn from experienced rewards. The task and their approach will be covered in detail in the case study section of this chapter.

Temporal difference (TD) prediction-error learning signals have been linked to the firing of dopamine neurons in the brain (Montague, Dayan, and Sejnowski 1996; Schultz 1998; Schultz 2002; O'Doherty et al. 2004) and there exists substantial evidence that these neurons play an important part in the experience of pleasure and reward (Dunlop and Nemeroff 2007). Using fMRI and a Pavlovian reward-learning task, Kumar et al. (2008) investigated whether TD learning signals would be reduced in MDD patients. The authors indeed found blunted reward prediction error signals in the patient group and additionally a correlation between such blunting and illness severity ratings. This provides a link

between an impaired physiological TD learning mechanism and reduced reward learning behavior as observed in anhedonia.

The previous study (Kumar et al. 2008) investigated Pavlovian learning during which participants passively observed stimulus-outcome associations. An early study to look at instrumental learning through active decision-making in depression was performed by Gradin et al. (2011). Stimuli were associated with different reward probabilities, which slowly changed. Prediction errors and expected values of a Q-learning model were regressed against fMRI brain activity. Compared to healthy controls, depressed patients did not display behavioral differences. However, physiologically they showed reduced expected reward signals as well as blunted prediction error encoding in dopamine-rich areas of the brain. This blunting correlated with anhedonia scores. This shows that model-based fMRI can reveal differences in reward learning; even in the absence of behavioral effects.

7.2.4 Bayesian Decision Theory

At a more abstract level**,** Bayesian decision theory (BDT) has been used to explain common symptoms of depression such as anhedonia, helplessness and pessimism (Huys et al. 2008; Trimmer et al. 2015; Huys, Daw, and Dayan 2015). Bayesian decision theory allows to formulate optimal behavior during a task and then to analyze how sub-optimal behavior can arise (see **Section 2.4**).

Huys et al. (2008) fitted a Bayesian reinforcement learning model to the behavior of depressed and healthy participants in two reward learning tasks. Importantly, their formulation of the model included two parameters, describing sensitivity to reward and a prior belief about control (cf. helplessness). Higher values of the control parameter corresponded to stronger beliefs about the predictability of outcomes following an action. Individuals who believe they have a lot of control over their environment would predict that previously rewarded actions will likely be rewarded again, while someone with a low control prior would expect weaker associations between action and reward. Huys et al. (2008) showed how a linear classifier could be used to distinguish between healthy and depressed participant after they had played a slot machine game, based purely on the two values of individuals' parameters. This suggests that model parameters obtained by fitting a behavioral task, such as a probabilistic learning task, could be used to classify MDD to a high accuracy. The classification of diseases is an important goal of computational psychiatry (Stephan and Mathys 2014).

A comprehensive evaluation framework formulated through BDT was introduced by Huys, Daw, and Dayan (2015), in which they discuss how depressive symptoms can arise from impairments in utility evaluation and prior beliefs about (the control over) outcomes. They argued that it is primarily model-based reinforcement learning, rather than model-free learning, which is abnormal in depression.

A theoretical description of how optimal decision-making can lead to (seemingly) depressed behavior and inaction similar to learned helplessness in a probabilistic environment can also be found in Trimmer et al. (2015). They concluded that to understand a patient's current depressed behavior, the history of the individual should be considered by describing it much further back in the past than what is the current norm. Imagine, for example, that Bob gets fired from his job due to "corporate restructuring" due to an economic crisis. Further, no other company seems interested in hiring while the economy is in this downswing, which is unlikely to change for the foreseeable future. Best efforts and repeated attempts to get a new job fail and adverse events in the environment increase (e.g. he loses friends or family or becomes homeless). Bob starts to learn that his actions do not seem to influence his environment. Negative outcomes appear unavoidable and over time his willingness to try to escape his situation decreases. Distressed and desperate, Bob starts to show symptoms reminiscent of depression. He has "learned to be helpless".

7.3 Case study: How does reward learning relate to anhedonia?

The case study in this chapter is a meta-analysis published by Huys et al. (2013) of a behavioral task that has consistently revealed reward-learning impairments in depressed and anhedonic individuals and other closely related groups.

Anhedonia is a core symptom of depression. Different behavioral tasks have been used to show that reward feedback *objectively* has less impact on participants who *subjectively* report anhedonia (Huys et al. 2013). However, there are different ways through which such a relationship could be realized. The goal of the meta-analysis was to find out whether anhedonia was principally associated with the initial *rewarding experience* of stimuli, or the subsequent *learning* from these rewards. The two mechanisms are important to disentangle, as they would likely correspond to distinct etiologies and different strategies for therapies (Huys et al. 2013).

7.3.1 Signal Detection Task

The Signal Detection Task (see **Figure 7.3**) consists of many (often 300) trials. In each trial one of two possible stimulus pictures (cartoon faces) is shown and the participant is prompted to indicate which picture was observed. This can be quite difficult, because the stimuli look very similar---they only differ slightly in the length of their mouth---and are only displayed for a fraction of a second. If participants correctly identify a stimulus, they sometimes received a reward (e.g. in the form of points) and sometimes receive no feedback. Participants are told to maximize their reward.

< insert **Figure 7.3** around here>

The most important aspect of the task is the *asymmetrical* reward structure. Unbeknownst to participants, one of the stimuli (called the "rich" stimulus) is followed by reward approximately three times as often as the alternative "lean" stimulus. If participants are not certain about the stimulus, they can incorporate knowledge about their reward history into their decision and choose the rich stimulus so as to maximize their chances to accumulate rewards. Healthy individuals have consistently shown to develop a response bias towards the rich option (Huys et al. 2013).

Using this task, Pizzagalli, Jahn, and O'Shea (2005) found a reduced ability in (healthy) participants with high depression (BDI) scores to adjust their behavior based on their reward history, while low BDI participants developed a stronger response bias towards the rich stimulus. Similarly, worse performance has been observed in MDD patients (Pizzagalli, Iosifescu, et al. 2008), stressed individuals (Bogdan and Pizzagalli 2006), euthymic (i.e. neutral mood) bipolar outpatients (Pizzagalli, Goetz, et al. 2008), as well as volunteers receiving medication (Pizzagalli, Evins, et al. 2008), and even healthy participants with a history of MDD (Dutra et al. 2009; Pechtel et al. 2013).

These studies used signal detection theory and summary statistics from raw behavior to analyze the data. Huys et al. (2013) extended this by using trial-by-trial reinforcement learning (RL) modelling to better understand the evolution of the behavior through time and get to a finer granularity in the analysis of the behavior.

While anhedonia has been associated with a diminished ability to use rewards to guide decision-making (such as in studies listed above), there exist varied possibilities for this impairment. Of primary interest in this case study was the distinction between the primary reward sensitivity, the immediately experienced consummatory pleasure following reward, and the *learning* from reward. Huys et al. (2013) included these two factors as parameters into a reinforcement learning model. **Figure 7.4** shows how changes in either reward sensitivity (ρ) or learning rate (ε) could lead to the empirically observed changes in response bias.

<insert Figure 7.4 around here>

7.3.2 A basic RL model

As described in **Chapter 2.3**, a standard Q-learning update rule incorporates learning rate in the following way:

$Q_{t+1}(a_t, s_t) = Q_t(a_t, s_t) + \varepsilon \times \delta_t$ *(Eq. 1)*

where s_t is the displayed stimulus on trial t, a_t is the action on trial t (i.e. which button was pressed), $Q_t(a_t, s_t)$ denotes the internal value assigned to the stimulus action pair (a_t, s_t) at trial t, $r \in \{0,1\}$ is the observed outcome, and $\delta_t = \rho r_t - Q_t(a_t, s_t)$ is the prediction error. Note that Huys et al. (2013) included a reward sensitivity parameter ρ that scales the true value of the reward. A lowering of the learning rate ε increases the time needed to learn about the stimulus-action pairs, while a lowering of the reward sensitivity ρ alters the asymptotic (average) values of Q that are associated with each pair.

In addition, Huys et al. (2013) included a term, $\gamma I(a_t, s_t)$, encoding participants' ability to follow the task instructions (i.e. press one key for the short mouth stimulus, and the other key for the long mouth stimulus), where:

$I(a_t, s_t) = 1$ if stimulus s_t required action a_t, and

$I(a_t, s_t) = 0$ if action a_t is the wrong response to stimulus s_t

Higher values for the parameter γ indicate a better ability to follow instructions and will result in generally higher accuracy. The two terms for I and Q were added together to form a "weight" for a particular stimulus-action pair (on trial t):

$$W_t(a_t, s_t) = \gamma I(a_t, s_t) + Q_t(a_t, s_t) \textit{ (Eq. 2)}$$

These weights are related to the probability of choosing action a when stimulus s was presented. From the above equation we can see that the probability of choosing an action does not only depend on following the task instructions (I), but also on the internal value based on previous experience (Q). Huys et al. (2013) used the popular SoftMax decision function to map these weights to action probabilities:

$$p(a_{\square}|s_t) = \frac{1}{1+exp\left(-\left(W_t(a_t,s_t)-W_t(\overline{a}_t,s_t)\right)\right)} \textit{ (Eq. 3)}$$

$W_t(\overline{a}_t, s_t)$ is the weight associated with choosing the wrong action for stimulus s at trial t. The softmax gives the probability that individuals choose the correct action given a certain stimulus. While individuals' parameters are not directly accessible, it is possible to infer them by *fitting* the model to their sequence of actions, i.e. by finding parameters that maximize the probability that the model would produce a similar sequence of actions when presented with the same sequence of stimuli (see **Section 2.5**).[13]

7.3.3 Including uncertainty in the model

The above model ignores one central aspect of the Signal Detection Task: stimuli are only displayed very briefly and so participants can never be certain about which of the two stimuli they actually observed. To account for perceptual uncertainty about the stimulus, Huys et al. (2013) expanded the model to assume that when participants compute their internal weights that guide their decision, they incorporate the possibility for both stimuli to have been presented. This leads to an updated equation for the weights, which now includes a term for stimulus s as well as a term for the alternative stimulus $\overline{s}$:

$$W_t(a_t, s_t) = \gamma I(a_t, s_t) + \zeta Q_t(a_t, s_t) + (1-\zeta) Q_t(a_t, \overline{s}_t) \textit{ (Eq. 4)}$$

13 One might wonder about the fact that the softmax function here does not include an (inverse) temperature parameter. However, it can be shown that such a parameter would be equivalent to ρ in these models (Huys et al. 2013).

Huys et al. (2013) use the parameter ζ to capture the average certainty (i.e. their belief) about which stimulus they actually observed and called this model "Belief".

7.3.4 Testing more hypotheses

Reinforcement learning models can be used to describe specific hypotheses about the behaviour of participants while performing the task. Model comparison (see also **Section 2.5**) then allows one to find the model that "best fits" the data, by which is generally meant that the model is neither too simplistic nor too complex and can explain how the data was generated. Usually, model comparison is used to test different hypotheses, heuristics, or strategies that participants may employ to solve the task. One other such hypothesis about performance in the Signal Detection Task is that participants could feel as if they are being punished when they do not receive a reward on a given trial. In the models described above, the reward r was coded as 1 or 0 (presence or absence of reward). Huys et al. (2013) changed the model to test the possibility that participants would perceive a lack of reward as punishment by including a punishment sensitivity parameter ρ^{-}. The prediction error term therefore becomes

$$\delta_t = \rho r_t + \rho^{-}(1 - r_t) - Q_t(a_t, s_t) \quad \textit{(Eq. 5)}$$

A final possibility is that participants might completely ignore the stimuli and only focus on the values of actions. Huys et al. (2013) formalized an "Action" model by setting the ζ parameter of the model "Belief" (in Eq. 4) to 0.5, which results in the weights equation

$$W_t(a_t, s_t) = \gamma I(a_t, s_t) + \frac{1}{2} Q_t(a_t, s_t) + \frac{1}{2} Q_t(a_t, \overline{s}_t) \quad \textit{(Eq. 6)}$$

The ζ parameter captures the average 'belief' about which stimulus they actually observed. By fixing the parameter at 0.5, participants are assumed to (on average) ignore the stimulus and only update the value of their actions. This means they would only learn about the values of "left" or "right" button press.

Table 7.1 summarizes all four models.

< Insert Table 7.1 around here>

7.3.4 Results

Huys et al. (2013) found that the model "Belief" best explained the data and therefore focused further analysis on this single model (**Figure 7.5A**). The authors also performed additional checks. For example, they confirmed that the model could explain more choices than a null model that assumed

participants always chose options randomly. Huys et al. (2013) then attempted to relate the estimated model parameters to measures of depressive symptoms severity, and in particular to anhedonia. The authors used the anhedonic depression (AD) questionnaire. They performed a correlation analysis to investigate whether primary reward sensitivity (ρ) or learning (ε) was most associated with AD (**Figure 7.5B**). They found a negative correlation between ρ and AD, but no significant correlation between ε and AD. This suggested that reward sensitivity rather than learning rate is primarily impaired in anhedonic depression.

< Insert Figure 7.5 around here>

There are limitations to these results. For example, Huys et al. (2013) found that reward sensitivity and learning rate were strongly negatively correlated. Additionally, the reward sensitivity parameter could not be distinguished from a temperature parameter typically included in the SoftMax decision rule. This means that differences in the reward sensitivity parameter might have masked differences in the exploration-exploitation behavior of participants. Another aspect of reward processing that the study did not touch on is effort, which is a large part of everyday decision making. Because in the signal detection task participants always have to exert the same amount of effort (a button press) independent of the stimulus they chose, it was not possible to address this here.

7.4 Discussion

Depression is a devastating disease with a major societal impact and rising prevalence (Vos et al. 2012), which make it an important area of study. Due to unclear boundaries between categorical definitions of psychiatric disorders, current research often focuses on individual personality traits such as neuroticism or depression symptoms such as anhedonia, both of which have been identified as promising endophenotypes of depression (Pizzagalli 2014). However, it has been noted that anhedonia itself encompasses various subdomains (e.g. hedonic response to pleasurable stimuli, but also motivation to pursue such stimuli) and these also need to be teased apart (Treadway and Zald 2013).

Patients suffering from depression routinely display impairments in a range of different experimental paradigms (Snyder 2013; Rock et al. 2014; Chen et al. 2015; Rupprechter et al. 2018). Different computational tools and techniques (connectionist models, diffusion models, reinforcement learning

techniques, Bayesian decision theory) have been used to describe this (abnormal) behavior and brain activity in depression, to gain insight into cognitive and neural processes, and to make predictions.

An important aim for computational psychiatry is the development of computational assays that can be used to separate patients into subgroups, generate treatment recommendations, and make predictions for the outcome of those treatments (Stephan, Baldeweg, and Friston 2006; Stephan and Mathys 2014; Chekroud et al. 2016). As Huys et al. (2016) put it, "Aspects of decision-making that have predictive value may become useful for the guidance of treatment or for alternative (and complementary) classifications of psychiatric disorders and individual patients." Reinforcement learning has been described as especially promising in this regard (Hitchcock et al. 2017) and has indeed shown potential for classification of depression from purely behavioral data without the need for (subjective) questionnaires (Huys et al. 2008).

Commonly observed pessimistic cognitive biases in depression have been explained using prior beliefs within the framework of Bayesian decision theory (Huys, Daw, and Dayan 2015; Stankevicius et al 2014). Simulations of neural network models have shown that biases could arise from a combination of different mechanisms including over-learning of negative information and rumination (Siegle and Hasselmo 2002). Drift diffusion models have been used to explain how aberrant behaviour relates to executive control deficits (Dillon et al. 2015; Vallesi et al. 2015) and rumination (Pe, Vandekerckhove, and Kuppens 2013).

RL models in which behavior is fitted on a trial-by-trial basis make it possible to measure group differences in behavior that are not obvious from raw data. Our case study (Huys et al. 2013) pooled data from various studies using the same experimental paradigm and fitted different reinforcement learning models according to hypotheses of the behavior of participants. The goal was to better understand anhedonia and how it is related to aberrant reward processing. Results indicated that the symptom is primarily associated with the initial experience of reward, rather than the reward learning mechanism.

On the neuronal level, there is substantial evidence that dopamine neuron activity encodes reward prediction errors (among other things; Schultz 1998; Iglesias et al. 2017). Work by Kumar et al. (2008) and Gradin et al. (2011) revealed that in depression prediction error signals appear reduced in the

striatum and other dopamine rich regions of the brain, suggesting that symptoms of depression are associated with an abnormal encoding of reward learning signals.

It is worth noting that in the meta-analysis of Huys et al. (2013), the authors found the two parameters of interest (reward sensitivity and learning rate) to be highly negatively correlated. Small changes in one of the parameters could therefore be compensated by changes in the other parameter, and Huys et al. (2013) had to perform additional analyses in order to increase their confidence in the fitted parameter values. The authors used the popular SoftMax function to model decision probabilities but decided against adding a temperature (or exploration-exploitation) parameter, because it would have traded off against the important reward sensitivity parameter. Changes in one of these parameters could have been compensated by changes in the other parameter. The larger question here is how to reliably distinguish between parameters. At least some computational variables are thought to be encoded in the brain (Iglesias et al. 2017), for example dopamine neurons' activity is believed to encode prediction errors. However, to discover these biological correlates we need reliable estimates that are not confounded by other parameters. The signal detection task was not initially designed with RL modelling in mind for example, and one could think about running a subtask to isolate exploration-exploitation behavior and estimate the temperature parameter independently. Replication of results, especially involving larger number of participants, will also be important before useful computational assays can be developed. Paulus, Huys, and Maia (2016) published a pipeline describing additional phases necessary for computational psychiatry to support the development for new drugs.

Current research has often focused on reward. While the omission of a reward might be felt as punishment by participants (as was assumed in Huys et al. 2013), Chen et al. (2015) point out that reward and punishment processing involve different neural bases. They hypothesize that depression might be characterized by a gain-loss asymmetry, so that patients experience decreased reward sensitivity but increased punishment sensitivity. As mentioned above, reward processing can also further be sub-divided into different domains. The association between anhedonia and the motivation to exert effort could not be addressed in our case study. In natural settings, patients weigh the pros (reward outcome) against the cons (effort required) to make a decision (cost-benefit analysis). Therefore, when an individual displays an abnormally large effort sensitivity, perceiving efforts as more effortful than they objectively are, they may decide against engaging in a potentially rewarding activity. The effort cost might be perceived as outweighing the potential reward outcome. This is also related to what is

observed in Parkinsons' patients who display high levels of apathy (a symptom akin to anhedonia; Husain and Roiser 2018). In the future, scientists may want to design tasks that enable them to test hypotheses about different reward learning domains such as effort sensitivity and reward sensitivity.

While much research points towards behavioral deficits of patients suffering from MDD, there is also evidence for improved performance in depression (Beevers et al. 2013). Replications and robust (computational) techniques will be needed to pinpoint exactly when impairments occur and how they relate to aberrant brain activity. Memory impairments are common in depression (Rock et al. 2014; Snyder 2013), but computationally they seem as of yet still largely unexplored. Notably, Dombrovski et al. (2010) included a memory parameter in their reinforcement-learning model and found that depressed suicide attempters discounted previously observed rewards more than healthy controls. It has been proposed that many observed impairments in schizophrenia could potentially be explained by deficits in the memory of patients (Strauss et al. 2010; Collins et al. 2014). Future research might want to consider whether memory impairments could also be a (partial) explanation for many of the observed abnormalities in depression.

< insert Box 7.1 here, Open Questions>

7.5 Chapter Summary

Behavioral impairments are prevalent in depression and computational methods provide a useful tool to tease apart different (neural) mechanisms that might influence learning and decision-making. Computational modelling of behavior in participants with depression has provided refinement and additional evidence for theories of MDD, which suggest that negative (perceptual) biases, deficient cognitive control, impaired reward learning, and beliefs about the controllability of the environment are all important aspects of the disease. Clever task design and replication involving larger samples, combined with robust computational techniques, are now needed to advance the field. It is important as well not to neglect the study of patients with moderate-severe mood disorder (rather than participants with low mood or mild forms of depression, who are often easier to study) and even of treatment-resistant patients. We want to move from findings that are able to distinguish between groups of

patients and healthy control participants to results that show convincing individual differences along symptom dimensions. This will ultimately be necessary to make treatment recommendations and predictions of outcomes for individuals based on non-invasive measurements.

7.6 Further Study

Chen et al. (2015) review a large number of computational studies in depression, focusing on reinforcement learning approaches. Early model-based neuroimaging studies showing altered brain activity during Pavlovian and instrumental learning in depression can be found in Kumar et al. (2008) and Gradin et al. (2011). Huys, Daw, and Dayan (2015) provide a compelling decision-theoretic analysis of depression and its symptoms. A recent study by Pulcu and Browning (2017) suggests that affective biases (i.e. the tendency to differentially prioritise the processing of negative events relative to positive events) - commonly observed in depression - may be related to individuals attributing higher information content to negative events than positive events. Recently, the availability of large amounts of data has enabled machine-learning approaches to be used for treatment outcome predictions (Chekroud et al. 2016).

Is there a more objective way to diagnose major depression, which does not rely on (subjective) interviews? Can we build automated assessment or screening tools using (computational modelling of) behaviour during decision-making tasks? Should we focus on categorical definitions, individual symptoms, or networks of symptoms (cf. Borsboom and Cramer 2013)? How are symptoms of depression related to other psychiatric disorders -- especially anxiety?
How far will brief experimental studies in the lab or clinical setting take us in the quest to better understand depression? How important is it to assess behaviour within more ecologically valid environments (e.g. using mobile phones to collect data during day-to-day activities)? How can we combine machine learning (data-driven) approaches with theory-driven computational modelling (cf. Huys, Maia, and Frank 2016) to make use of vast amounts of data?
When is it sufficient to look at behaviour and at what point do we need to include the analysis of brain activity? How are abnormalities in brain function related to alterations in brain structure?
What are the sub-domains of reward and punishment processing and how to these sub-domains (e.g. "liking" and "wanting") relate to symptoms of depression?
Can memory impairments explain many of the observed behavioural abnormalities?
Can we use the knowledge gained through the computational approach to depressive disorders to develop better pharmacological or psychological therapies or prevention strategies?

Box 7.1: Open questions in computational research regarding depressive disorders.

Chapter 8: Anxiety Disorders from a Computational Perspective

Erdem Pulcu and Michael Browning

Department of Psychiatry, University of Oxford

8.1 Introduction

Anxiety disorders are among the most common psychiatric diagnoses, with the lifetime prevalence of any of the disorders estimated to be as high as 33% (Alonso, Lépine, and ESEMeD/MHEDEA 2000 Scientific Committee 2007). A range of specific diagnoses is included under the umbrella term of anxiety disorders (). Many of these specific diagnoses are based around the context in which symptoms of anxiety are evoked. For example, social anxiety disorder describes a condition in which anxiety is evoked by social situations whereas agoraphobia describes a condition in which anxiety is evoked by being in situations from which it is difficult to escape from (or where help is not available). There has been some debate about the precise set of diagnoses which should be included as anxiety disorders with the recent version of the Diagnostic and Statistical Manual (DSM-V) opting to move obsessive-compulsive disorder and post-traumatic stress disorder out of the anxiety category and into their own categories (see for summary of anxiety related diagnoses in recent diagnostic manuals). Generally a diagnosis of one of the anxiety disorders requires that significant symptoms of the disorder are present, often for at least 6 months, that the symptoms cause significant difficulties in everyday life, and that they cannot be better accounted for by other psychiatric or medical conditions or by the effects of drugs or alcohol.

< insert Table 8.1 around here>

A second approach to subdividing the anxiety disorders, other than the context in which symptoms are evoked, is whether symptoms of fear or worry are predominant in the presentation of the disorder. Fear

describes a set of responses, including physiological, behavioral and subjective, to a well-defined threat and is characteristically seen in the specific phobias, such as phobias of animals like spiders, or situations such as darkness. In contrast, worry describes a set of responses to less well-defined, often potential future threats and is characteristically seen in generalized anxiety disorder. It is generally easier to elicit fear responses in a laboratory setting or in animal models than it is to induce worry. As a result of this much of the etiological work relevant to anxiety, including that reviewed below, has focused on the systems responsible for the production of fear responses rather than those implicated in worry.

Lastly, as with many psychiatric conditions, it is worth noting that symptoms of anxiety in the population appear to occur on a continuum with little evidence of qualitative shifts in symptoms between "clinical" and "non-clinical" groups. Because of this, studies that examine "trait anxiety", the tendency to experience symptoms of anxiety in everyday life, can be informative when considering etiological processes in the anxiety disorders.

In this chapter we provide a brief overview of the relevant conceptual background and results of recent studies which have taken a computational approach to study anxiety (see also; Raymond, Steele, and Seriès 2017; Grupe, D.W. 2017; Bishop and Gagne 2018 for review) before describing one study (Browning et al. 2015) in more detail. We end by briefly summarizing the state of the literature and suggesting how it may most effectively be developed.

8.2 Past and Current Computational Approaches

The observation that underpins much of the mechanistic work investigating anxiety disorders is that individuals can learn to fear stimuli or situations that they previously did not fear and, equally, can learn that previously feared stimuli or situations are in fact safe. This was memorably demonstrated a century ago in the experiments carried out by John Watson and Rosalie Rayner on the 9-month old child known as "Little Albert". In these studies, Albert was allowed to play with a white laboratory rat, to which he showed no fear. Following this, whenever he touched the rat, the experimenters made a sudden loud noise by banging a hammer against a steal bar, which startled Albert. Subsequently, when the rat was shown to Albert he would react with fear, even though no loud sounds were made. In other words, Albert had associated a neutral stimulus (the rat; in conditioning parlance, called the conditioned

stimulus; CS+) with an aversive stimulus (the loud sound or unconditioned stimulus; US) and had thus learned to show a fear response (crying or the conditioned response; CR) to the rat. Notwithstanding developments in the ethical oversight of experimental studies that have curtailed psychologists' freedom to traumatize infants, the same general experimental procedure has formed the basis of a large body of fear conditioning studies in humans and animals. The methodology of the studies has been developed by including control stimuli (CS-) which are not paired with aversive outcomes and by examining extinction (i.e. the reduction of a previously learned fear association which occurs when the CS+ is presented in the absence of a US), which allows these studies to test some simple hypotheses about the etiology of anxiety disorders:

a. Do patients with anxiety disorders demonstrate an enhanced learning of fear association to the CS+?
b. Do patients with anxiety disorders demonstrate a reduced extinction of fear associations?
c. Do patients with anxiety disorders demonstrate a greater generalization of the fear CR (i.e. do patients respond to safe stimuli, CS-, as if they were associated with the aversive outcome)?

A recent meta-analysis of fear conditioning studies in anxious participants (Duits et al. 2015) did not find evidence for enhanced fear learning to the CS+ (although see the earlier meta-analysis reported by Lissek et al. 2005 for slightly different conclusions), but did find evidence for reduced extinction of the CS+ and for increased generalization from the CS+ to the CS-.

The relative ease with which fear conditioning paradigms can be deployed in animal models has stimulated a well-developed mechanistic literature on the amygdala-based neural systems which support fear learning (Duvarci and Pare 2014; Johansen et al. 2011) and a parallel clinical neuroimaging literature in anxious patients (LeDoux and Pine 2016; Craske et al. 2017). The overarching picture from the latter describes a tendency for anxious individuals to show increased limbic (including amygdala) and reduced frontal activity in response to aversive stimuli (Indovina et al. 2011). While this work has led to mechanistic models which describe specific roles for distinct neural systems in the anxiety disorders (LeDoux and Pine 2016), to date computational approaches have not been employed in this work. As a result, we focus in the rest of this chapter on studies, which examine the behavior of anxious individuals and how computational techniques have been used to investigate this.

Conditioning studies such as those described above are well suited to computational descriptions with much of the early models of reinforcement learning being used to capture learning behavior in animal

conditioning studies (Rescorla and Wagner 1972). However, computational approaches have rarely been applied to behavioral or physiological measures in human fear conditioning studies relevant to anxiety. One reason for this may be that traditional human fear conditioning tends to employ "strong situations" (Lissek, Pine, and Grillon 2006) in which a CS+ (e.g. a shape on a screen) is paired deterministically with a unconditioned stimulus such as a shock. When faced with this sort of radically simple study design human participants can generally learn the association between the CS+ and the aversive outcome in one or two trials. In this sort of simple learning situation, behavioral or physiological responses over only a handful of trials are generally collected. Such responses can be adequately captured using simple summary statistics and computational analysis tends not to add much. However, concern as to the ability of strong situations to capture the aspects of fear learning most relevant to anxiety has prompted recent studies to explore how anxiety is related to learning in more ambiguous situations. Such situations represent areas in which computational descriptions start to be more useful.

One approach to introducing ambiguity into fear conditioning studies has been to utilize strong fear conditioning procedures, with CS+/CS- stimuli strongly associated with the presence/absence of aversive outcomes, but then test participants' response to stimuli that are ambiguous with regard to their identity as CS+ or CS-. For example, Lissek and colleagues (Lissek et al. 2010) used a large ring stimulus as a CS+ and a small ring as an unambiguous CS- in patients with panic disorder and controls. Following this, participants were presented with stimuli of intermediate size, while startle response was measured using electromyography (EMG). In keeping with similar work in a variety of clinically anxious populations from the same laboratory, patients with panic disorder showed a greater degree of generalization of the CR than controls; that is, they reacted to a greater proportion of the ambiguous stimuli as if they were a CS+ than controls. This work suggests that anxiety may be associated with a difficulty in precisely representing states of the world, in accurately assigning credit for aversive outcomes or with a reduced belief in one's ability to avoid future aversive outcomes (Zorowitz, Momennejad, and Daw 2019). While finding the optimal approach to generalization and credit assignment is a core question tackled by the machine learning literature (Alpaydin, Ethem 2009), to date fear conditioning studies in humans which examine generalization (see; Dymond et al. 2015 for a recent review) have again tended to rely on summary statistics rather than computational approaches.

A second way in which ambiguity may be introduced to conditioning studies is by reducing the strength of the association between conditioned and unconditioned stimuli (i.e. by reducing the probability with

which an aversive outcome follows a cue) and/or by employing designs in which the strength of this association changes over time (i.e. by changing which cue is most predictive of an outcome; Yu and Dayan 2005). These designs begin to capture some of the complexity missing from simple conditioning studies and highlight the real-world challenges faced by an individual trying to learn what may harm them in the environment. While the specific challenges introduced by this ambiguity are described in more detail in the case study example below, their effect is straightforward— they vastly increase how difficult it is to learn about the causes of aversive outcomes and therefore to select the optimal behaviors which avoid such outcomes. A number of lines of evidence suggest that humans employ various heuristics, simplified decision rules, in order to render this problem more tractable (Tversky and Kahneman 1992; Kahneman and Tversky 1979). The degree to which use of these heuristics is associated with anxiety have been examined in a number of studies using computational techniques.

Two of the most consistently reported heuristics are risk aversion—the tendency to select certain over probabilistic outcomes even when the expected value (i.e. the probability multiplied by the magnitude) of the certain outcome is lower; and loss aversion—the tendency to be more influenced when making a decision by potential losses than potential gains. Avoidance of perceived threatening situations is believed to be a causal process in the anxiety disorders (Barlow, D. H. 2004) suggesting that both of these heuristics may be exaggerated in anxiety disorders and that reducing them may be an important component of treatment. In order to assess this possibility, Charpentier and colleagues (Charpentier et al. 2017) compared the behavior of a group of clinically anxious patients with non-anxious controls using a gambling task in which both risk and loss aversion could be independently estimated as parameters of a Prospect Theory (Tversky and Kahneman 1992) inspired decision rule. The authors reported significantly increased risk but not loss aversion in the anxious group suggesting that the core process associated with anxiety is an aversion of risk rather than a general overweighting of negative outcomes, although the same group also report an increased loss learning rate in response to aversive stimuli (Aylward et al. 2019).

A complementary view of behavioral heuristics during learning and decision making suggests that humans (and animals) combine both a flexible instrumental learning system, which learns the best action to take in response to specific stimuli, with a stereotyped Pavlovian system which responds in an evolutionary pre-specified manner to stimuli (Dickinson, T. and Balleine, B. 2002). The Pavlovian system leads to fast, rigid responses to stimuli such as generally withholding responses to punishments

while facilitating responses to rewards. Mkrtchian and colleagues (Mkrtchian et al. 2017) probed these systems in patients with anxious or mood disorder and controls, using a task in which, on some trials, participants had to respond in line with Pavlovian biases and other trials in which they had to generate opposing responses (e.g. withhold a response to gain a reward or respond to avoid a punishment). This design allowed the authors to separately estimate the impact of the instrumental and Pavlovian systems on participant behavior using a reinforcement-learning model that included parameters which estimated the influence of both systems. The patient group was found to be more strongly influenced by the Pavlovian bias to withhold responses to a punishment, with other model parameters unchanged. The authors suggested that this reliance on Pavlovian inhibition provided mechanistic insight into the behavioral avoidance that is characteristic of anxiety disorders, that is, the avoidance arises because anxious individuals are more influenced by the automatic tendency to withhold responses in the face of punishment.

The final way in which computational approaches have been used in studies of anxiety is as a tool to further decompose cognitive processes that are associated with the disorders. Beyond the learning-based work described above, cognitive accounts of anxiety suggest that habitual threat-related biases, i.e. the tendency to prioritize threat-related information at the expense of non-threatening information, are causally linked to the disorders (Mathews and MacLeod 2005). For example, experimental studies suggest that a greater tendency to direct attention towards threat-related information is causally associated with anxiety (MacLeod et al. 2002). Three separate studies have used drift-diffusion models to decompose reaction time data from tasks investigating such negative biases in anxiety. As described in **Section 2.2**, Drift-diffusion models attempt to capture the process by which decisions (generally perceptual decisions) are made (Ratcliff et al. 2016), breaking this process into an initial non-decision making stage and a later stage in which evidence is noisily accumulated over time until a decision boundary is crossed. Firstly, White and colleagues (White et al. 2010) reported that high anxious subjects demonstrated a higher drift rate for threatening, relative to neutral, stimuli, with a later result providing similar evidence using a slightly different metric (White et al. 2016). Aylward and colleagues (Aylward et al. 2017) report similar findings using positive outcomes with higher anxiety being associated with a lower drift rate for positive stimuli. Together these results are consistent with two possible interpretations: a) that anxious participants view the threatening stimuli as more threatening (and positive stimuli as less positive) or b) that anxious individuals use a lower threshold to classify a stimulus as threatening and a higher threshold to classify positive stimuli. By re-parameterizing the

traditional measures of negative bias reported for anxious participants, these studies hint at the "where" in the process of evidence accumulation, biases may be created.

In summary, the centrality of learning and decision making to the mechanistic literature on anxiety and its disorders makes them well placed to benefit from the insights provided by computational approaches. However, to date relatively few studies of anxiety have employed computational techniques. In the following section we describe in more detail the results from a final study that investigated how anxious individuals deal with the uncertainty caused by learning about an association that changes over time.

8.3 Case Study Example: Anxious individuals have difficulty in learning about the uncertainty associated with negative outcomes (from *Browning et al. (2015)*)

8.3.1 Theoretical Background, Expected and Unexpected Uncertainty

As described in the previous section, learning in the real world is more challenging than that captured by traditional conditioning studies (see Pulcu and Browning 2019 for a general discussion of uncertainty estimation). Below we present an example to illustrate the different sources of uncertainty that complicate learning and then describe how this uncertainty can be dealt with.

Imagine trying to learn what mood your cat is in based purely on observing whether it does or does not scratch you when you stroke it (**Figure 8.1**). When the cat is in a good mood it will only scratch you when it is play-fighting, say on 10% of the times you stroke it (**Figure 8.1** green areas), whereas when it is in a bad mood it will scratch you on 80% of the times you stroke it (**Figure 8.1**, red areas). The cat's mood is therefore useful to know—because it will help you predict how likely you are to be scratched in the future. However, given that you can't directly observe the cat's mood, you need to infer it based on previous events (whether it scratches you when you stroke it). The first challenge in this task is that being scratched (or not) by the cat is an ambiguous measure of the cat's mood-- if you are scratched, it may be because the cat is in a bad mood (and is therefore more likely to scratch you) or it may be because it is play-fighting. Similarly, if you are not scratched, it may be because the cat is in a good mood or that you were just lucky and this time, for whatever reason, it chose not to scratch you even though it was in a bad mood. One way to get a better estimate of the cat's mood is to study its behavior over a longer time period; if you have stroked it 10 times and it has only scratched you once then it is

probably in a good mood (although you still can't be certain of this). However, a further challenge limits how useful collecting data over a longer time period is--the cat's mood is characterized by some degree of volatility, i.e. it will change over time, so even if the cat was in a good mood the last time you stroked it, it may be in a bad mood now, which means you can't rely on the cat's previous behavior, particularly in the distant past, as being representative of its current mood.

< Figure 8.1 around here>

In order to learn as accurately as possible what the cat's mood is and therefore how likely it is that you will be scratched you need to deal with the uncertainty generated by the two challenges described above; first the cat's behavior is probabilistic rather than deterministic, so even if you know precisely what its mood is, you will not be able to predict with certainty whether it will scratch you when you stroke it. This form of uncertainty is sometimes called "expected uncertainty" (Yu and Dayan 2005) as, after sufficient experience, it can be precisely determined (e.g. you can know that the probability of the cat scratching you is exactly 10% when it is in a good mood even though you can't say with certainty what will happen on each occasion you stroke it). Expected uncertainty erodes how informative each individual event is when you are learning. For example, imagine your cat's behavior had low expected uncertainty so that it never scratched you when it was in a good mood but always scratched you when it was in a bad mood. In this case you can instantly tell what mood the cat is in after you have stroked it once. On the other hand if the cat's behavior has high expected uncertainty so that it scratches you 40% of the time when it is in a good mood and 60% of the time when it is in a bad mood, it becomes much more challenging to estimate its mood. In other words, the higher the expected uncertainty the less informative each particular event (i.e. stroking the cat and observing whether it scratches you) is. A second form of uncertainty is produced by changes in the underlying association that you are learning and is sometimes called "unexpected uncertainty" (Yu and Dayan 2005). This occurs when the cat's mood changes from good to bad or vice versa, so that the probability that it will scratch you changes. The effect of unexpected uncertainty is to reduce how informative previous events are during learning. For example, imagine your cat's mood never changes and the probability that it will scratch you is always 30%. In this case the unexpected uncertainty is low and the most accurate way to precisely estimate how likely it is to scratch you (and therefore its mood) is to estimate over many trials the

average rate at which it scratches you. In contrast, when learning about a cat whose mood changes frequently, that is whose behavior has high unexpected uncertainty, you can't rely on distant events as it is likely that they occurred when the cat was in a different mood than currently and you have to rely more on recent events. In other words, previous events become increasingly less informative the higher the unexpected uncertainty. In the next section we introduce a simple learning model to illustrate how one should adapt to these sources of uncertainty during learning.

8.3.2 Learning as a Rational Combination of New and Old Information

The Rescorla-Wagner learning rule (Rescorla and Wagner 1972) provides a simple description of how one might learn what the cat's mood is:

$$r_{(t+1)} = r_{(t)} + \alpha(s_{(t)} - r_{(t)})$$

In this equation, $r_{(t)}$ is the model's estimate of the probability that the cat will scratch you at time t, which we will use as a metric of its mood (i.e. when r is 1 the cat's mood is as bad as it can be, when it is 0 the cat's mood is as good as it can be). We initialise this so $r_1 = 0.5$ and then update the model's belief every time the cat is stroked using the outcome information $s_{(t)}$ which equals 1 if the cat scratches and 0 otherwise. A single parameter is included, the learning rate α, which lies between 0 and 1. Generally the learning rate is treated as a free parameter or is arbitrarily set at some value. However, in order to learn as efficiently as possible, the learning rate used should adapt to the two sources of uncertainty describe above (or, at least, to the learner's estimates of these uncertainties).

Note that in the above equation, $s_{(t)}$, the outcome, represents the new information presented to the model each time the cat is stroked and $r_{(t)}$ is the model's current belief about the mood of the cat, which has been influenced by all the previous times it has been stroked. If we rearrange the above equation to separate these two variables we get:

$$r_{(t+1)} = (1 - \alpha)r_{(t)} + \alpha(s_{(t)})$$

This demonstrates that the Rescorla-Wagner model's belief after each event is simply a weighted mean of the information provided by the recent event ($s_{(t)}$) and the model's previous belief ($r_{(t)}$) with the learning rate acting as the weight. When the learning rate is 1 all the weight is placed on the new

information and the model discards its previous belief, whereas when the learning rate is 0, the model places all the weight on its previous belief and ignores the new information.

In the previous section, we described how expected and unexpected uncertainty influence how informative events are—a high expected uncertainty reduces how informative new events are, a high unexpected uncertainty reduces how informative previous events are. Efficient learning requires beliefs to be more influenced by informative than non-informative events indicating how expected and unexpected uncertainty should influence learning rate. High expected uncertainty (i.e. a noisy relationship between cue and outcome) reduces how informative current events are indicating that a lower learning rate should be used to reflect the fact that previous events are relatively more informative, high expected uncertainty (volatility) reduces how informative previous events are indicating that a higher learning rate should be used. The relationship between unexpected uncertainty and learning rate is illustrated in **Figure 8.2** in which it can be seen that a model with a high learning rate is better at learning about a volatile cat, whereas a low learning rate is better for a stable cat.

< Figure 8.2 around here>

8.3.3 Effect of Volatility on Human Learning

As explained above, learning about a volatile process is more efficiently achieved with a higher learning rate. A number of studies (Behrens et al. 2007; 2008; Nassar et al. 2012) have examined whether humans adapt their learning as described above, that is whether humans estimate the volatility of the process they are learning about and tune their learning rate to increase learning efficiency. In all of these studies, participants were required to learn about the association between a cue and a reward during periods in which the association between the two was either volatile or stable. The consistent findings of the studies are that participants adapted the learning rate they used precisely as described above employing a higher learning rate in volatile than stable contexts. Interest has also focused on physiological markers of this volatility estimation process. An early synthesis of animal work (Yu and Dayan 2005) suggested that phasic activity of the central norepinephrine system (NE) may contain an estimate of volatility or unexpected uncertainty. This proposition is consistent with current theories

(Aston-Jones and Cohen 2005) on the broader role of NE, which is argued to increase the gain of sensory representations and thus increase their impact on behavior (i.e. analogous to an increased learning rate which, as described above, is an appropriate response to volatility). Phasic activity of the central NE system is correlated with pupil dilation in primates (Joshi et al. 2016) suggesting that it may be possible to estimate activity of this system using pupillometry. Nassar and colleagues (Nassar et al. 2012) collected pupillometry data during their study and reported a positive correlation between the learning rate participants employed and the magnitude of pupillary dilation during the outcome phase of their task. These findings are in line with the proposal by Yu and Dayan (2005) and suggest that estimates of central NE may provide a physiological marker of the neural process that adapts learning rate to estimated volatility.

8.3.4. Summary of Browning et al. Study

The background presented above suggests that humans adapt their learning to statistical aspects of their environment—such as the stability or volatility of the association they are learning. This observation raises the possibility that anxiety may be associated with difficulties in implementing this adaptation, rather than (or as well as) gross differences in learning about or extinguishing fear associations.

In order to test this possibility Browning and colleagues (Browning et al. 2015) recruited a group of non-clinical participants who had been pre-screened to ensure a range of trait anxiety scores. Participants completed an aversive learning task (**Figure 8.3**) in which two shapes were probabilistically associated with receiving an electric shock while pupilometry data was collected. The crucial manipulation of the task is that it was formed of two blocks (**Figure 8.3b**)—one volatile and one stable.

The learning rate for each participant and each block was estimated by fitting a computational model to participant choice in that block. The model consisted of three stages with a single free parameter in each stage. First, a simple Rescorla-Wagner rule was used to learn the probability that the shock was associated with 'shape A':

$$r_{shapeA(i+1)} = r_{shapeA(i)} + \alpha\varepsilon_{shapeA(i)}$$

In this stage, $r_{shapeA(i)}$ is the model's belief on trial i that the shock would be associated with shape A (note that the belief for shape B is simply 1 – that for shape A), $\varepsilon_{shapeA(i)}$ is the prediction error and the

free parameter α is the learning rate. The second stage calculated the value g_{shape} of each shape by combining this learned probability, $r_{shapeA(i+1)}$, with the shock magnitude, $I_{shapeA(i+1)}$:

$$g_{shapeA\,(i+1)} = F\left(r_{shapeA(i+1)}, \gamma\right) * I_{shapeA(i+1)}$$

$$F(r,\gamma) = \max[\min[(\gamma(r-0.5)+0.5), 1], 0]$$

In this stage, $F_{(r,\gamma)}$ transforms the learned probability using the free parameter γ. The effect of this parameter is to either increase or decrease the relative weight of the probability vs. the magnitude when calculating the value ($g_{shapeA(i+1)}$) of each shape (i.e. this allows for the possibility that participants did not use the exact produce of probability and magnitude when making decisions but rather could be more influenced by outcome probability ($\gamma > 1$) or magnitude ($\gamma < 1$)). Finally, the two values are combined using a SoftMax equation:

$$P_{(choice=shapeA)} = \frac{1}{1+\exp^{(-\beta(g_{(shapeB)}-g_{(shapeA)})}}$$

This stage has a single free parameter, β, the inverse temperature that controls the degree to which the values influence choice. The results derived from fitting this model are displayed in **Figure 8.4**. The critical result is displayed in panel B, which shows the relationship between trait anxiety and the degree to which participants altered their learning rate between the volatile and stable blocks. As can be seen, participants with lower anxiety adjusted their learning rate to a greater extent than participants with high anxiety. In other words, anxiety was not associated with a grossly increased or decreased learning rate during this task; rather anxious participants were less sensitive to the volatility of the task and adjusted their learning rate less.

The second question addressed in the study was the degree to which the physiological measure of central NE function, pupil dilation, was related to the behavioral measure of learning rate and to trait anxiety (**Figure 8.5**). This was assessed using a two-stage analysis of the pupil data, similar to that employed in fMRI studies. At the first level, regression analyses were performed for each subject, which estimated the degree to which a range of explanatory variables, including estimated volatility, influenced pupil dilation on a trial-by-trial basis. Separate regression analyses were performed for each time point of pupillary data over six seconds after outcomes were presented. These analyses produced time series of beta weights which estimate the degree to which pupil dilation in an individual participant was

influenced by a particular explanatory variable. A second level of analysis then combined these time series across all participants to test whether the explanatory variables influenced pupil dilation across the population of participants. These analyses demonstrated that a greater differential pupil dilation between volatile and stable blocks was positively associated with a greater behaviorally estimated learning rate difference between blocks (results not shown) which is consistent with Yu and Dayan's proposal for the role of NE (Yu and Dayan 2005). Critically the analysis also revealed that the pupils of participants with higher anxiety differentiated between volatile and stable blocks less than those of low anxious participants.

< Figure 8.5 around here>

Overall this study provided initial evidence that computational approaches may be usefully used to unpick the abnormal fear learning associated with anxiety. Specifically, it suggested that anxiety may be associated with difficulties in estimating the unexpected uncertainty of an environment or in using these estimates to guide learning.

8.4 Discussion

As reviewed in this chapter, anxiety disorders are a prime target for investigations that utilize computational approaches largely because there is clear evidence for a role of abnormal learning and decision-making in their etiology. To date, relatively few studies using computational approaches in anxiety have been published. Although preliminary, the computational work which has been completed has begun to describe differential use of decision making and learning heuristics in anxious individuals (Charpentier et al. 2017; Mkrtchian et al. 2017), reduced sensitivity to statistical aspects of the environment (Browning et al. 2015) and biased evidence accumulation processes during threat perception (White et al. 2016; 2010; Aylward et al. 2017).

A common theme across these studies is that computational approaches have been used to identify and describe cognitive processes linked to anxiety that are not readily apparent using traditional analytic

strategies. For example, the study by Mkrtchian and colleagues sought to separately estimate the impact of an automatic Pavlovian and a more flexible instrumental learning system on decision making in anxiety, which would be difficult to achieve without some formal estimation of the effects of the two systems. The motivation for identifying and characterizing such cognitive processes is that they may improve our etiological understanding of anxiety which, it is hoped, will ultimately facilitate better patient care by improving our ability to stratify diagnoses and/or by guiding the development of novel treatments.

A related observation about the published computational studies is that they have all used case-control designs to investigate and delineate cognitive differences between anxious and non-anxious groups. This is clearly a reasonable first step in identifying processes that are perturbed in anxiety, however these designs rarely provide tangible clinical benefit as they don't provide strong evidence for a causal relationship between the processes and symptoms of anxiety or provide the sort of information that might usefully guide treatment. We believe that progress in realizing the clinical benefit of computational studies will require a broader range of study designs to be implemented in the future and suggest two specific examples. Firstly, having identified computationally defined processes associated with anxiety it will be important to test whether measurement of these processes may be useful in clinical situations, for example, do they predict prognosis or treatment response, suggesting that they may be used to guide treatment decisions? Longitudinal studies in depression have begun to find associations between computationally defined processes and treatment response suggesting that this approach is feasible (Huys et al. 2016) and may be usefully deployed in anxiety disorders. The second example concerns the development of novel treatments. Relationships between a cognitive process and anxiety, such as those described in this chapter, are particularly interesting when the relationship is causal. Causality is most clearly established using experimental designs in which the computational process is manipulated and the effects of this on symptoms are then measured. Conceivably, computationally defined processes which are causally related to symptoms may provide a new class of treatment targets for anxiety so it will be important to establish which of the identified processes are indeed causally related to symptoms rather than simply being associated with them. Manipulation of computational processes may be achieved using targeted cognitive interventions such as those used in the cognitive bias modification literature (Browning et al. 2012). However, an advantage of computational approaches generally is that they have been successful in linking cognitive processes to the underlying neural and neurochemical systems that produce them. This raises the possibility of also

using pharmacological interventions to manipulate the computational processes, such as using norepinepheric agents to alter learning rate (Jepma et al. 2016). Currently, the most commonly used pharmacological treatments for anxiety are benzodiazepines, which enhance central GABA transmission, and serotonin reuptake inhibitors, which increase synaptic serotonin (NICE 2007). While the molecular effects of these agents have been well characterized, their impact on cognitive processes is less clear. Computational approaches may also be used to extend our understanding of these drugs' mechanism of action, possibly by examining their impact on the computational outcomes described in this chapter.

A final observation is that, if a clinical impact is to be achieved it will be essential to ensure that published computational results are as robust and reliable as possible. Reliability is demonstrated by the replication of results and robustness requires the collection of large clinical data sets. Both of these goals will be facilitated by closer collaboration between disparate research teams in the computational field. Such collaborations are becoming increasingly feasible with developments in online communication technology as well as stimulus presentation and analysis software that facilitate the sharing of tasks and code as well as general communication between centers. The nascent field of computational psychiatry is well placed to take advantage of these developments (Browning et al. 2019).

To conclude, computational approaches in anxiety are in their infancy, they have shown early promise in being able to identify and describe novel cognitive processes related to anxiety. Translation of this promise into clinical benefit will require the adoption of robust study methodology and a willingness to employ a broader range of study design.

8.5 Chapter Summary

A large amount of previous work has demonstrated that anxiety and its disorders are associated with abnormal learning about aversive outcomes. Computational approaches can be particularly useful when investigating both learning and decision making although relatively few studies to date have employed these techniques in anxious populations. The studies that have been published suggest that anxious individuals utilize different decision making and learning heuristics, show reduced sensitivity to statistical aspects of the environment and biased evidence accumulation during threat perception. While

computational work in anxiety is in its infancy, it shows promise in being able to identify novel cognitive processes which are relevant to the disorders and which are not apparent using standard analytic approaches. Future work needs to employ robust methodology and a broader range of study design if this promise is to be realized.

8.6 Further Study

Craske and colleagues (Craske et al. 2017) provide a recent and broad (although not computational) review of diagnostic, mechanistic and treatment related issues in the anxiety disorders including a section on the neural, genetic and cognitive associations of the disorders. This paper would be of interest to those who want a broad introduction to issues in anxiety research. A more focused review on the neurobiology of subjective vs. behavioral fear is provided by LeDoux and Pine (LeDoux and Pine 2016). The example study in this chapter measured the changes in learning rate induced by volatility by fitting a simple learning model to a stable and a volatile block of trials. However, computational approaches can also be used to describe the underlying calculations necessary to estimate volatility. A number of previous papers have described different approaches to this problem. While these papers don't specifically focus on anxiety disorders, they provide a useful computational background on how the volatility effect described in this chapter may be conceptualized. Firstly a seminal study by Pearce and Hall (Pearce and Hall 1980) describe how the Rescorla-Wagner model may be modified such that it adapts to the how surprising stimuli are (one way of estimating volatility is as the frequency with which surprising outcomes are observed). Secondly a neuroimaging paper by Li and colleagues (Li et al. 2011) suggested that a smoothed version of a Pearce-Hall signal was present in the human amygdala. Behrens and colleagues (Behrens et al. 2007) describe a fully Bayesian approach to this problem.

Chapter 9: Addiction from a Computational Perspective

A. David Redish

University of Minnesota, USA.

9.1 Introduction: what is addiction?

Everyone knows what addiction is. We all know people whose lives have been ruined by drugs and we all have behaviors that we wish we could stop, but don't. However, the definition of addiction remains elusive. Early definitions related to a "lack of will" and suggested addiction was a moral failing. However, this theory did not lead to reliable treatments and left many incapable of ending their addictions. Later definitions defined addiction as a disease and suggested that behavioral and chemical treatments could alleviate it. In particular, these disease-related theories suggested that many drug addictions arose from biological responses to chemical imbalances that could be treated pharmacologically. Some of these pharmacological treatments, such as methadone treatment for heroin addictions (Meyer and Mirin, 1979) and the nicotine patch for smoking (Hanson et al., 2003), have been very successful, but other addictions (stimulants, alcohol) have been much more difficult to treat pharmacologically. Furthermore, pharmacological definitions do not include the possibility of non-chemical addictions, such as gambling, which is now seen as an addiction-like problem.

Current definitions of addiction are based on conceptualizations of addiction as a problem with decision-making systems (Heyman, 2009; Redish, 2013), often evidenced as continued use despite stated preferences (Goldstein, 2000, Ainslie 2001) and as continued use despite high cost (Robinson and

Berridge, 2003; Koob and Le Moal, 2006). The most recent models identify addiction as arising from vulnerabilities leading to failure-modes in decision-making algorithms (Redish et al 2008).

One of the most common popular descriptions of addiction lies in the addict's continued use despite making explicit statements of a desire to stop. Current theories of decision-making reject the hypothesis of the unitary decision-maker – each individual is actually a multiplicity of decision-making systems (algorithms, processes) competing for behavioral control (O'Keefe and Nadel, 1978; Daw et al., 2005; Rangel et al., 2008; Redish et al., 2008; Kahneman, 2011; van der Meer et al, 2012; Redish, 2013). While this theory provides an explanation for this conflict (Kurzban, 2010), computational models of addiction have not emphasized this conflict because it is hard to study in non-linguistic animals (i.e. non-humans), while human rights limitations make it difficult to do controlled studies of addiction in humans. Nevertheless, the study of decision-making systems and their interaction is well established in both human and non-human animals and has been used computationally to guide treatment.

One of the classic descriptions of addiction is based on the observation that addicts will continue to use even in the face of high costs. This can be quantified through the economic concept of elasticity as a measure of how much one's willingness to buy something changes by its cost (Bickel et al. 1993; Hursh et al., 2005). Things that diminish slowly by cost are inelastic. Researchers have suggested that drugs are fundamentally inelastic: as costs increase, the number of rewards paid for decrease less than they should. Of course, there are many things that are inelastic that are not considered addictive – oxygen, for example (where the withdrawal symptoms are particularly traumatic), but also some behaviors continued even in the face of high costs are celebrated, such as Kerri Strug's 1996 Olympic vault performed on a sprained ankle, or Osip Mandelstam continuing to write poetry even after Stalin had thrown him in the gulag for it.

A key to the question of addiction is to separate the science of why an agent continues its behavior from the decision to treat and change that behavior. This conceptualization parallels Jerome Wakefield's conceptualization of psychiatry as depending on harmful dysfunction (Wakefield, 1992). "Dysfunction" reflects a system not working as it was intended to. For example, mu-opioid activation signals pleasure in mammalian brains (Berridge and Robinson, 2003). These receptors were certainly not evolved to respond to heroin, but they do. "Harmful" reflects a society's choice of what to change. For example, American society is currently transitioning from treating marijuana as so dangerous as to be illegal with severe penalties to something that can better be handled under legal regulation. Things can be harmful

without being dysfunctional, such as tribal wars, which are extremely harmful, but likely reflect the natural evolution of human behavior (Turchin, 2003; Diamond, 2006), and dysfunctional without being harmful, such as synesthesia (Cytowic, 1998).

Computational models of addiction are aimed at understanding the science of why an agent continues its behavior and the science of how one could change that behavior if one so desired. Importantly, the decision of whether to change that behavior has not been computationally assessed. Such a decision would depend on sociological models, which are not the focus of this chapter. Instead, this chapter will focus on computational approaches to addictive behavior and its modification.

9.2 Past approaches

Past computational approaches to addiction can be divided into three broad categories: economic models, in which drugs are seen as economic objects that have feedback properties that make them overvalued; homeostatic models, in which drugs change intrinsic biological properties and shift allostatic set-points which subsequently require drugs to reach that set-point; and reinforcement learning models, in which drugs hijack learning algorithms to produce aberrant learning. Current views on addiction suggest that these three hypotheses are all failure modes of decision-making systems, and that there are many endophenotypes of drug addiction.

9.2.1 Economic models

Although popular descriptions of drug use (e.g. Reefer Madness [Gasnier, 1949], Long Day's Journey into Night [O'Neill, 1956], The Lost Weekend [Wilder, 1945], Sid and Nancy [Cox 1986]) suggested that drugs were overwhelming and addicts would spend any cost to achieve drug-taking, experimental studies have long suggested that drugs were economic objects and that drug use decreased with increasing costs (Bickel et al., 1993; Liu et al., 1999; Grossman and Chaloupka,1998; Hursh et al., 2005). The first economic model of drug use is Becker and Murphy's 1988 "Rational Addiction" model, which is an economic utility model in which subjects are assumed to select the most cost-effective choice with the highest value. Drugs are assumed to have a positive feedback so that the more one takes those drugs, the more valuable they become. Becker and Murphy show that under these assumptions, a

hypothetical user could be shown to become addicted when the positive feedback overwhelms the negative consequences of the drug use.

These models led to quantitative analyses of drug use, asking direct questions of the economic demand curves of drug use. Demand curves are quantitative measures of elasticity. This can be measured either through effort (how many lever presses will a non-human animal push for reward?) or through monetary costs (how many grams of drug will you buy?) In a typical demand curve (**Figure 9.1**), there is an inelastic portion, where increases in cost have little effect on number of rewards bought, and an elastic portion, where the number of rewards bought falls off very quickly. These are separated by an inflection point (*pMax*). Addicts can be defined as people where this inflection point has shifted far to the right, but nevertheless, their demand curves do have this typical, canonical shape.

< Figure 9.1 around here>

A key insight from this economic perspective on drug use is that drugs provide fast rewards and slow consequences. All animals (human and non-human) discount future rewards, valuing rewards more if they are delivered in a shorter time frame (Ainslie, 1992; Madden and Bickel 2010). Economically, this makes sense as immediate rewards can be invested, and consequences can prevent the use of later rewards. Importantly, as described in **Section 5.2**, all animals (human and non-human) show non-exponential discounting curves (**Figure 9.2**), which means that preferences can cross – thus it is possible both to prefer to smoke the cigarette in your hand and to prefer to not smoke in the future. (Of course, when the future becomes now, one will want to smoke the cigarette now again.) Addicts show particularly fast discounting functions, which can exacerbate this problem (Bickel and Marsch, 2001). There is some evidence that successful treatment modifies these discounting rates in subjects with particularly fast discounting functions (Bickel et al., 2014) and that these discounting rates are predictive of relapse (Sheffer et al., 2014). It is possible to modify discounting rates, by guiding the subject's attention to delayed rewards by providing episodic cues about the delayed rewards to make those delayed rewards more concrete (Peters and Büchel, 2010). Recent evidence has suggested that these changes can reduce drug use (Stein et al., 2018; Snider et al., 2018). However, whether these changes

are due to changes in discounting rates per se or to changes in interacting multiple decision systems remains an open question.

< Figure 9.2 around here>

While the basic economic story that drugs are economic objects that are discounted quickly is clearly correct, drug use is context sensitive in ways that make these simple economic descriptions incomplete (Bernheim and Rangel, 2004). We will return to the question of these economics later, when we come to the interacting multiple systems models, below.

9.2.2 Homeostatic models

All drugs that are reliably self-administered, either by humans or other animals, are pharmacologically similar in some way to endogenous chemicals used in neural processing (Koob and Le Moal, 2006). For example, active opioids such as morphine, heroin, or oxycodone activate the mu-opiate receptor, cocaine blocks dopamine reuptake in the synapse, which increases dopamine in the synapse, amphetamine encourages release of dopaminergic vesicles, and nicotine activates acetylcholine receptors. Biological systems in general and neural systems in particular are very sensitive to levels of these endogenous chemicals and have extensive negative feedback processes (such as trafficking of receptors in and out of the synaptic membrane) that keep the sensitivity balanced. In situations where receptors are flooded, they will normalize their levels requiring more activation to produce the same effects.

For example, many self-administration experiments (in which animals are trained to press a lever for drug reward) can be described quantitatively in terms of maintenance of pharmacological levels of drug (Tsibulsky and Norman 1999, Keramati et al., 2017). Negative feedback processes driving maintenance interact with the positive feedback processes of drug utility (as suggested by Becker and Murphy) to produce dramatic differences in valuation between drugs and non-drug rewards (with drugs being valued much higher than non-drugs, leading to over-taking of drug rewards.)

Three quantitative models based on issues of homeostatic balance are the Tsibulsky and Norman (1999), the Keramati et al. (2015), and the Dezfouli et al. (2009) models (**Figure 9.3**). The Tsibulsky and Norman model explicitly hypothesizes that animals are attempting to maintain a specific level of cocaine, which explains quantitatively the observed shifts in response to changes in the dosages given

with each lever press. Keramati et al. notes, however, that there are short-term dynamics when the changes actually occur which are not explicable by a simple set-point hypothesis, particularly in the transition that occurs with increased access to drug. They therefore add in a learning component based on the reinforcement learning models detailed below. The Dezfouli et al. model is based on a homeostatic expansion of the Redish (2004) model (see below), particularly looking at the effect of homeostatic set-points driving pharmacological effects of dopamine on learning. While the Redish model is based on the temporal-difference-reinforcement learning dopamine-as-delta model of Montague et al. 1996, and is thus a hijacked-learning model, the Dezfouli et al. model is based on the average reward dopamine hypothesis of Daw and Touretzky (2000), and becomes a homeostatic model.

Opponent process theory

One of the earliest models of drug use is the opponent process theory of Solomon and Corbit (1974, see Koob and Le Moal, 2006 for extensive discussion of this model), in which drugs are assumed to produce a strong positive reward followed by a strong negative recovery. Homeostatic processes are assumed to normalize the excess drug to decrease the positive factors, and increase the negative factors, which leads to increased need for drugs to return the homeostatic process to baseline. These models have been supported by evidence that chronic drug use leads to enhancement of positive valuation neuron activity in the nucleus accumbens (Kourrich et al. 2007; Volman et al. 2013) and evidence that the emotional crash after drug use is an important factor in driving self-administration (Rothwell et al. 2010).

While the Solomon and Corbit and Koob and Le Moal models are not quantitative, Gutkin et al. (2006) proposed an opponent process model in which there is habituation of response processes to a continuous delivery of nicotine – a phasic increase at the start and a phasic decrease at the end, and a decrease in the overall tonic dopamine levels. The normalization caused by the assumed decrease in dopamine levels leads to a decrease in ability to learn non-drug related cues, which leads to an increase in attention to and learning of drug-related cues. Thus, Gutkin et al. shows how an opponent process model can hijack learning process by disrupting the difference between learning on and off drug.

< Figure 9.3 around here>

9.2.3 Reinforcement models

The third family of computational models is based on the concept that learning depends on physical processes, and those physical processes can be modulated by external chemicals and other processes. In animal learning theory, the concept of reinforcement is separate from the concept of reward. Reinforcement is any mechanism that makes an agent more likely to return to an action. An external chemical that increases reinforcement would increase drug-seeking and drug-taking (di Chiara 1999, Redish 2004).

In the 1950s, it was discovered that electrical stimulated of specific neural sites was reinforcing, in that both human and non-human animals would activate the stimulation (Olds and Milner 1954), even to the extent of avoiding many other rewards. Interestingly, in humans (who could rate "pleasure" linguistically) these studies found that the most reinforcing stimulations were not always the most pleasant (Heath 1963).

An important breakthrough in the understanding of reinforcement came when Berridge and Robinson directly measured reinforcement and pleasure in non-human animals and discovered that they were separable. It was well-known that many drugs of abuse affected dopaminergic functioning and that the stimulation drove dopamine release and it was thought that dopamine would drive pleasure signals. However, when Berridge and Robinson (2003) directly tested this hypothesis, it was discovered that this was wrong – dopamine and pleasure were dissociable. In their elegant studies, they measured facial expressions of pleasure and disgust in rats under manipulations of dopamine and opiate signals. Dopamine manipulations affected reinforcement but did not affect facial expressions of pleasure. In contrast, manipulations of opiate signals (e.g. mu-opiate and kappa-opiate agonists and antagonists) affected pleasure responses. This led them to hypothesize that drugs that affected dopamine increased the "incentive salience" or "value" of a reward, which drove seeking, independently of the pleasure experienced by that reward.

Around this time, a major breakthrough occurred in the understanding of dopamine function in animal learning – Wolfram Schultz and his team discovered that dopamine cells burst when provided a surprising reward but did not fire when the reward was predicted by a cue (Ljungberg et al., 1992).

Read Montague and colleagues (1996) realized that this signal was the value prediction error (VPE)[14] signal δ (delta) that underlay a theory of robotic learning called temporal difference reinforcement learning (TDRL) that had become very successful in the field of computer science[15] (Sutton and Barto, 1998; see also **Section 2.3**).

As described in **Section 2.3**, the temporal difference reinforcement learning algorithm (TDRL) defines value as the total reward one can expect to achieve given a policy of actions to be taken in given situations. TDRL maintains a representation of the currently believed value for each situation, and then calculates the difference between that remembered value and the observed value. This difference is the value prediction error or VPE. Positive VPE occurs anytime a value is better than expected and drives an increased willingness to take an action, while negative VPE occurs anytime a value is worse than expected and drives a decreased willingness to take an action. The concept of VPE is best understood through an example. Imagine a soda machine. If you put your money in the soda machine and get two sodas out, then you will be more willing to put money in that soda machine next time. (You have positive VPE.) If you put your money in the soda machine and get nothing out, then you will be less willing to put money in that soda machine next time. (You have negative VPE.) And, most importantly, if you put the correct amount of money in the soda machine, get your expected soda out, then you understand how that machine works and you don't need to learn anything about it. (You have zero VPE.) Notice that you still get the pleasure (such as it is) of drinking the soda, but you don't need to change your willingness to put money in that machine. VPE is about learning the value of actions. Computer simulations had shown that VPE would allow an agent to learn to behave in simulated environments (Sutton and Barto, 1998). These processes can be expressed in the following equations:

$$V(S_k) = \int_t^{\infty} \gamma^{\tau - t} E[R(\tau)] d\tau$$

[14] The dopamine signal is often described as a "reward prediction error" signal, but this is a misnomer, as bursts also occur when unexpected value appears. Thus it is better referred to as a "value prediction error" signal.

[15] Although many studies have supported the hypothesis that the phasic bursting of dopamine signals positive value prediction error and that pauses in firing signal negative value prediction error, recent experiments have suggested that the story may be more complex than previously thought. Recent experiments have found that not all costs are reliably included in this calculation (Gan et al. 2010; Wanat et al. 2010). And recent experiments looking at tonic levels have suggested that dopamine is actually signalling value, so that value prediction error would occur only from high-pass frequency filters of dopamine (Hamid et al. 2016).

$$\delta(t) = \gamma^d[R(S_l) + V(S_l)] - V(S_k)$$
$$V(S_k) \leftarrow V(S_k) + \eta\,\delta$$

Where $V(S_k)$ is the value of state S_k, γ^d is a discounting parameter[16], reflecting expected value decreases over observed delay d, $R(S_l) + V(S_l)$ is the value achieved on entering state S_l, and $\delta(t)$ is the value prediction error (the difference between the observed and expected value). By changing the value of state S_k towards the observed value (with learning rate η), $V(S_k)$ will approach the observed value. Theories hypothesized that dopamine signaled the value prediction error $\delta(t)$.

Redish (2004) proposed that if drugs were providing a dopamine signal pharmacologically, then taking drugs would lead to positive VPE, even if the neural calculation of VPE should have been 0 (**Figure 9.4**). Effectively, Redish's model predicted that the dopamine signal at reward contained two components, one from the calculation of $\delta(t)$, and the other from the pharmacological action of the drug. This meant that even with experience, there would always be a non-compensable VPE signal at the reward, which would increase the predicted value of the reward, driving that value to infinity. (Or with normalization, normalizing all other values to zero.)

$$\delta = \max\{\gamma^d[R(S_l) + V(S_l)] - V(S_k) + D(S_l), D(S_l)\}$$

where $D(S_l)$ reflects the effect of the pharmacological dopamine from the drug.

In his 2004 paper, Redish used computer simulations to show that this model would lead to developing inelasticity (as in the Becker and Murphy hypothesis) and made several untested predictions. The first prediction was that there would be a double surge of dopamine in drug experiments. In the TDRL theory, $\delta(t)$ first appeared at the time of reward (as it was initially unexpected) and then it shifted to earlier cues that reliably predicted the reward (because the reward was now expected – thus δ=0, but the cues indicated an unexpected increase in value – thus δ>0). Similarly, Schultz and colleagues (see Schultz, 2002) had found that dopamine shifted from the reward (when unexpected) to the cue (once the animal learned that the cue predicted the reward). In Redish's model, the extra pharmacological component would always appear, even as the dopamine signal appeared at the cue. Since then, this

[16] Note that these equations use exponential discounting to reflect value decreases over observed delay. It is possible to construct consistent TDRL equations that express behaviors that reflect non-exponential discounting, but this requires additional complexities beyond what is necessary for this chapter (Kurth-Nelson and Redish, 2009, 2010).

double surge of dopamine has been observed, but as with any theory, reality is more complex than the model, and each component of the double-surge occurs separately, with the reward-related surge appearing in accumbens shell and the cue-related surge appearing in accumbens core (Aragona et al. 2009).

<Figure 9.4 around here>

The two other key predictions of the Redish 2004 model were (1) that additional drug use would always lead to increased valuation of the drug and (2) that drugs would not show Kamin blocking. These predictions have since been tested and provide insight into the mechanisms of drug addiction.

In the Redish (2004) model, the excess dopamine provides additional value, no matter what. Marks et al. (2010) directly tested this hypothesis in an elegant experiment, where rats were trained to press two levers for a certain dose of cocaine (both levers being equal). One lever was then removed and the other provided smaller doses of cocaine. The Redish (2004) theory predicts that the second lever should gain value, while expectation or homeostatic theories like those discussed earlier would predict that the second lever should lose value (because animals would learn the second lever was providing smaller doses). The Marks et al. data was not consistent with the Redish excess-delta model. However, as noted above, a key factor in drug addiction is that not everyone who takes drugs loses control over their drug use and becomes an addict. Studies of drug use in both human and non-human animals suggest that most animals in self-administration experiments continue to show elasticity in drug-taking, stopping in response to high cost, but that a small proportion (interestingly similar to the proportion of humans who become addicted to drugs) become inelastic to drug-taking, being willing to pay excessive costs for their drugs (Anthony et al., 1994; Hart, 2013). One possibility is that the homeostatic models (like that of Tsibulsky and Norman, 1999) are a good description of non-addicted animals, which have a goal of maintaining a satiety level, but that addiction is different.

The Redish (2004) model also predicted that drugs would not show Kamin blocking. Kamin blocking is a phenomenon where animals don't learn that a second cue predicts reward if a first cue already predicts it (Kamin, 1969). This phenomenon is well-described by value prediction error (VPE) – once the animal learns that the first cue predicts the reward, there is no more VPE (because it's predicted!) and the

animal does not learn about the second cue (Rescorla and Wagner, 1972). Redish noted that because drugs provided dopamine and dopamine was hypothesized to be that VPE delta signal, then when drugs were the "reward", there was always VPE. Thus, drug outcomes should not show Kamin blocking. The first tests of this, like the Marks et al study, did not conform to the prediction – animals showed Kamin blocking, even with drug outcomes (Panlilio et al., 2007). However, Jaffe et al (2014) wondered whether this was related to the subset problem – that only some animals were actually overvaluing the drug. Jaffe et al. tested rats in Kamin blocking for food and nicotine. All rats showed normal Kamin blocking for food. Most rats showed normal Kamin blocking for nicotine. But the subset of rats that were high responders to nicotine did not show Kamin blocking to nicotine, even though they did to food, exactly as predicted by the Redish model.

9.3 Interacting multi-system theories

Studies of decision-making in both human and non-human animals have, for a long time, found that there are multiple decision-making processes that can drive behavior (O'Keefe and Nadel, 1978; Daw et al., 2005; Rangel et al., 2008; Redish et al., 2008; Kahneman, 2011; van der Meer et al, 2012; see Redish, 2013 for review). These processes are sometimes referred to as different algorithms because they process information differently. They are accessed at different times and in different situations; they depend on different neural systems. How an animal is trained and how a question is asked can change which system drives behavior. Damage to one neural structure or another can shift which system drives behavior.

The key to these different systems lies in how they process information. Decision-making can be understood as a consequence of three different kinds of information – what has happened in similar situations in the past (memory), the current situation (perception) and the needs/desires/goals (teleology). How information about each of these aspects is stored can change the selected action – for example, what defines "similar situations" in the past? What parameters of the current situation matter? Are the goals explicitly represented or not? Each system answers these questions differently.

Almost all current decision-making taxonomies differentiate between planning (deliberative) systems and procedural (habit) systems. Planning systems include information about consequences – if I take this action, then I expect to receive that outcome, which can then be evaluated in the context of explicitly encoded needs. Planning systems are slow but flexible. Procedural systems cache those actions – in this situation, this is the best action to take, which is fast but inflexible. As described earlier (**Section 2.3** &

5.2), many current computational models refer to planning systems as model-based (because they depend on a model of the consequences in the world), while procedural systems are model-free (which is an unfortunate term because procedural systems still depend on an ability to categorize the current situation, which depends on a model of the world [Redish et al. 2007; Gershman et al. 2010]). Some taxonomies also include reflex systems, in which the past experience, the parameters of the current situation that matter, and the action to be taken are all hard-wired within a given organism and are learned genetically over generations. Most taxonomies also include a fourth decision-system, variously termed Pavlovian, Emotional, Affective, or Instinctual, in which a species-important action (e.g. salivating, running away, approaching food) is released as a consequence of a learned perception (context or cue).

The importance of these systems is three-fold: (1) How a question is asked can change which system controls behavior; (2) Damage to one system can drive behavior to be controlled by another (intact) system, and (3) There are multiple failure modes of each of these systems and their interaction. We will address each of these in turn.

9.3.1 How a question is asked can change which system controls behavior

One way to measure how much rats value a reward such as cocaine is to test them in a progressive ratio self-administration experiment (Hodos 1961). In this experiment, the first hit of cocaine costs one lever press, but the second costs two, the third costs four, the fourth eight, etc. Eventually a rat has to press the lever a thousand times for its hit of cocaine. Measuring when the rat stops pressing the lever indicates the willingness-to-pay and the value of the cocaine to the rat. Not surprisingly, many experiments have found that rats will pay more for cocaine than for other rewards such as saccharine, indicating that cocaine was more valuable than saccharine. However, Serge Ahmed's laboratory found that if those same rats were offered a choice between two levers, one of which provided saccharine while the other provided cocaine, the rats would reliably choose the saccharine lever over the cocaine lever, indicating that saccharine was more valuable than cocaine [Lenoir et al. 2007, see Ahmed et al 2010]. The most logical explanation for this contradiction is that the progressive ratio accesses one decision system (probably procedural) while the choice accesses another (probably deliberative) and that the two systems value cocaine differently. Interestingly, Perry et al. (2013) found that a subset of rats will choose the cocaine, even in the two-option paradigm. These are the same subset of rats that over-value cocaine in other contexts, such as being willing to cross a shock to get to the cocaine (Deroche-

Gamonet et al. 2004). Whether they are also the high responders or whether they no longer show Kamin blocking remains unknown.

9.3.2 Damage to one system can drive behavior to another

Imagine an animal pressing a lever for an outcome (say cheese). If the animal is using a planning system to make its decisions, then it is effectively saying "If I push this lever, I get cheese. Cheese is good. Let's press the lever!" If the animal is using the procedural system, then it is effectively saying "Pressing the lever is a good thing. Let's press the lever!" – cheese never enters into the calculation. What this means is that that if we make cheese bad (by devaluing it, which we can do by pairing cheese with a nauseating agent like lithium chloride), then rats using planning systems won't press the lever anymore ("If I push this lever, I get cheese. Yuck!"), but rats using procedural systems will ("Pressing the lever is a good thing. Let's press the lever!"). (See, for example, Niv et al. 2006 for a model of this dichotomy.) Many experiments have determined that with limited experience, animals are sensitive to devaluation (i.e. they are using a planning system), while with extended experience they are not (i.e. they are using a procedural system), and that lesions to various neural systems can shift this behavior (Killcross and Coutureau 2003; Schoenbaum et al 2006). A number of studies have suggested that many drugs (cocaine, amphetamine, alcohol) drive behavior to procedural devaluation-insensitive systems, which has led some theoreticians to argue that drug addiction entails a switch from planning to habit modes (Everitt and Robbins, 2005).

Building on the anatomical data known to drive the typical shift from planning to procedural decision systems, Piray et al. (2010) proposed a computational model in which drugs disrupted the planning valuation systems and accelerated learning in the procedural valuation systems. This model suggested that known changes in dopaminergic function in the nucleus accumbens as a consequence of chronic drug use could lead to overfast learning of habit behaviors in the dorsal striatum and would produce a shift from planning to habit systems due to changes in valuation between the two systems.

9.3.3 There are multiple failure modes of each of these systems and their interaction

However, rats and humans will take drugs even when they plan. A drug addict who robs a convenience store to get money to buy drugs is not using a well-practiced procedural learning system. A teenager who starts smoking because he (incorrectly) thinks it will make him look cool and make him attractive

to girls is making a mistake about outcomes and taking drugs because of an error in the planning system (the error is in his understanding of the structure of the world.)

Some researchers have argued that craving depends on the ability to plan, because craving is transitive (one always craves *something*), thus it must be depend on expectations and a model-based process (Tiffany 1999; Redish and Johnson, 2007). In fact, there are many ways that these different decision systems could drive drug-seeking and drug-taking (Redish et al. 2008). Some of those processes would depend on expectations (i.e. would be model-based, depend on planning) and explicit representations of outcomes, and could involve craving, while other processes would not (i.e. would be model-free, depending, for example, on habit systems). [An important consequence of this is the observation that seems to get rediscovered every decade or so that craving and relapse are dissociable – you can crave without relapsing and you can relapse without craving.]

In 2008, Redish and colleagues surveyed the theories of addiction and found that all theories of addiction could be re-stated in terms of different failure modes of this multi-algorithm decision-making system. An agent that succumbed to over-production of dopamine signals (Redish 2004) from drug delivery would over-value drugs and would make economic mistakes to take those drugs. An agent that switched decision-systems to habit faster under drugs (Everitt and Robbins, 2005; Piray et al. 2010) would become inflexible in response to drug offerings and take drugs even while knowing better. An agent with incorrect expectations ("smoking makes you cool", "I won't get cancer") would make planning mistakes and take drugs in incorrect situations. An agent that discounted the future ("I don't care what happens tomorrow, I want my pleasure today.") would be more likely to take drugs than an agent included future consequences in its plans (Bickel and Marsch, 2001). All of these are different examples of vulnerabilities within the decision-making algorithms. Redish et al (2008) proposed that drug addiction was a symptom, not a disease – that there were many potential causes that could drive an agent to return to drug-use, and that efficacious treatment would depend on which causes were active within any given individual.

9.4 Implications

9.4.1 Drug-use and addiction are different things

At this point, the evidence that a subset of subjects have runaway valuations in response to drugs is overwhelming (Anthony et al., 1994; Deroche-Gamonet et al. 2004, Koob and Le Moal, 2006; Hart,

2013; Perry et al., 2013; Jaffe et al., 2014). This is true both of animal models of drug addiction and humans self-administering drugs. This suggests a very important point, which is that drug use and addiction are different things. If we want to treat the harm that drugs do, then we may want to address drug use rather than addiction, which would require sociological changes (Hart, 2013). As noted above, these sociological models are beyond the scope of this chapter, which is addressing computational models of addiction.

9.4.2 Failure modes

This chapter has discussed three families of models. The first family was *economic models*, which simply defined addiction as inelasticity, particularly due to mis-valuations. However, these models did not identify what would cause that mis-valuation. The second family was *pharmacological models*, which defined addiction as a shift in a pharmacological set-point which drove value in an attempt to return the pharmacological levels back to that set-point. The third family was *learning and memory models,* which suggested that addiction derived from vulnerabilities in the neural implementations of these algorithms, which drove errors in action-selection.

The multiple failure-modes model suggests that all three families provide important insights into addiction. It suggests that there were multiple potential vulnerabilities that could drive drug use (which could lie in pharmacological changes in set-points or in many potential failure modes of these learning systems). The multiple vulnerabilities model suggests that addiction is a symptom not a disease. Many failure modes can create addiction. Importantly, identifying which failure modes obtain within any given individual would require specially designed probe tests; this model suggests that it would not be enough to merely identify extended drug use. In fact, these failure modes are likely to depend on specific interactions between the drug and the individual and the specific decision processes driving the drug-seeking/drug-taking behavior.

9.4.3 Behavioral addictions

If addictions are due to failure modes within neural implementations of decision-making algorithms, then addiction does not require pharmacological effects (even if pharmacological effects can cause addictions) and it becomes possible to define behavioral problems as addictions. For example, problem gambling is now considered an addiction, and other behaviors (such as internet gaming, porn, or even shopping) are now being considered as possible addictions. As noted at the beginning of the chapter, the

definition of addiction is difficult. Nevertheless, computational models of addiction have provided insight into problem gambling and behavioral change in general, whether we call those behaviors addictive or not.

Classic computational models of problem gambling have been based on the certainty and uncertainty of reward delivery, but these models have been unable to explain observed properties of gamblers, such as that gamblers tend to have had a large win in their past (Custer, 1984; Wagenaar, 1988), that they are notoriously superstitious about their gambling (Griffiths 1994), or that they often show hindsight bias (in which they "explain away losses", Parke and Griffiths 2004), or the illusion of control (in which they believe they can control random effects, Langer, 1975).

Redish and colleagues (2007) noted that most models of decision-making were based on learning value functions over worlds in which the potential states were already defined. Furthermore, they noted that most animal learning experiments took place in cue-poor environments, where the question the animal faced was "*What is the consequence of this cue?"* However, most lives (both human and non-human) are lived in cue-rich environments, in which the repeated structure of the world is not given to the subject. Instead, the subject has to identify which cues are critical to the definition of the situation the subject finds itself in. They noted that this becomes a categorization problem and had been well studied in computational models of perception. Attaching a perceptual categorization process based on competitive learning models (Hertz et al., 1991) to a reinforcement learning algorithm, they built a model in which the tonic levels of dopamine (i.e. longer-term averages of $\delta(t)$) controlled the stability of the situation-categorization process. This identified two important vulnerabilities in the system depending on over- and under-categorization, particularly in the different responses to wins and losses. In their model, wins produced learning of value, while losses produced recategorizations of situations. Their simulated agents were particularly susceptible to near misses and surprising wins, leading to models of hindsight bias and the illusion of control.

In general, these multi-system models suggest that addiction is a question of harmful dysfunction – dysfunction (vulnerabilities leading to active failure modes) within a system that causes sufficient harm to suggest we need to treat it. They permit both behavioral and pharmacological drivers of addiction.

9.4.4 Using the multi-system to treat patients

However, the suggestion that different decision-making systems can drive behavior provides a very interesting treatment possibility, which is that one could potentially use one decision-system to correct for errors in another. Three computational analyses of this have been done –changing discounting rates with Episodic Future Thinking (Peters and Büchel, 2010; Snider et al., 2018; Stein et al., 2018), analyses of Contingency Management (Petry, 2012; Regier and Redish, 2015), and analyses of Precommitment (Kurth-Nelson and Redish, 2010).

Episodic future thinking is a process in which one imagines a future world (Atance and O'Neill, 2001), which is the key to planning and model-based decision-making, in which one simulates (imagines) an outcome, and then makes one's decision based on that imagined future world (Niv et al., 2006; Redish, 2013, 2016). Models of planning suggest that discounting rates may depend in part on the ability to imagine those concrete futures. Part of the discounting may arise from the intangibility of that future (Rick and Loewenstein, 2008; Trope and Liberman, 2010; Kurth-Nelson et al., 2012), which may explain why making future outcomes more concrete reduces discounting rates (Peters and Büchel, 2010). Other models have suggested that these discounting rate decreases occur through changes in the balance between impulsive and more cognitive decision systems (McClure and Bickel, 2014). Nevertheless, recent work has found that treatments in which subjects are provided concrete episodic future outcomes to guide episodic future thinking can decrease discounting rates (providing a more future-oriented attitude) and decrease drug use (Snider et al., 2018; Stein et al., 2018). Whether this effect comes from the changes in discounting rates per se or whether those changes are reflective of other processes (such as an increased ability to use planning and deliberative systems) is currently unknown.

Contingency Management is a treatment to create behavioral change (such as stopping use of drugs) through the direct payment of rewards for achieving that behavioral change – effectively paying people to stop taking drugs (Petry, 2012). Contingency Management was originally conceived of economically: if drugs have some elasticity (which they do, see **Figure 9.1**), then paying people not to take drugs increases the cost of taking drugs, by creating lost opportunity costs. In psychology, this would be called an alternate reinforcer.

However, Regier and Redish (2015) noted that the rewards that produced success in contingency management did not match the inelasticity seen in either animal models of addiction nor in real world measures of inelasticity due to changes of drug costs in the street. Building on the idea that choosing to

take a drug or not (a go/no-go task, asking one's willingness-to-pay) accesses different decision-making algorithms than choosing between two options (take the drug or get the alternate reward), Regier and Redish suggested that contingency management had effectively nudged the subject to use their deliberative decision-making systems. They then suggested that this could provide improvements to standard contingency management methods, including testing for prefrontal-hippocampal integrity (critical to deliberative systems) and providing concrete alternatives with reminders (making it easier to imagine those potential futures). Whether these suggestions actually improve contingency management has not yet been tested.

The fact that addicts show fast discounting functions with preferences that change over time suggest two interesting related treatments: bundling and precommitment. Bundling is a process whereby multiple rewards are grouped together so as to calculate the value of the full set rather than each individually (Ainslie, 2001). For example, an alcoholic may want to go to the bar to drink one beer but recognizing that going to the bar will entail lots of drinking, may reduce the value of going to the bar relative to staying home. This can shift the person's preferences from going to the bar to staying home.

A similar process is that of precommitment, where a subject who knows in advance that if given a later option, the subject will take the poor choice, prevents the opportunity in the first place. The classic example is that a person who knows they will drink too much at the bar decides not to go to the bar in the first place. Economically, precommitment depends on the hyperbolic discounting factors that lead to preference reversals (Ainslie, 2001). Preference reversals imply that the earlier person wants one option (to not drink) while the later person wants a different one (to drink). Although many experiments have found that the average subject shows hyperbolic discounting (Madden and Bickel, 2010), individuals can show large deviations from good hyperbolic fits. Computationally, an individual's willingness to precommit should depend on the specific shape of their discounting function (Kurth-Nelson and Redish, 2010).

Furthermore, Kurth-Nelson and Redish (2010) proved that neurophysiologically, precommitment depends on having a multi-faceted value function – that is, the neural implementation of valuation has to be able to represent multiple values simultaneously. One obvious possibility is that the multiple decision-making systems each value options differently, and conflict between these options can be used to drive precommitment to prevent being offered the addictive option in the first place.

<Figure 9.5 around here>

9.5 Chapter Summary

Because addiction is fundamentally a problem with decision making, computational models of decision making (whether economic, motivational [pharmacological], or neurosystem) have been important to our definitions and understanding of addiction. These theories have led to new treatments and new modifications that could improve those treatments.

9.6 Further Study

Koob and Le Moal. (2006) provides a thorough description of the known neurobiology of addiction.

Bickel et al (1993) is a seminal article showing that behavioral economics provides a conceptual framework that has utility for the study of drug dependence.

Redish (2004) was the first explicitly computational model of drug addiction and set the stage for considering addiction as computational dysfunction in decision systems.

Redish et al (2008) provides evidence that addiction is a symptom rather than a fundamental disease and proposed that the concept of vulnerabilities in decision processes offers a unified framework for thinking about addiction.

Chapter 10 : Tourette Syndrome from a Computational Perspective

Vasco A. Conceição and Tiago V. Maia
Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Portugal

10.1. Introduction

10.1.1. Disorder definition and clinical manifestations

Tourette syndrome (TS) is a disorder characterized by tics—repetitive, stereotyped movements and oral-nasopharyngeal noises—that are usually preceded by aversive sensations called premonitory urges (American Psychiatric Association 2013; Leckman, Walker, and Cohen 1993; Robertson et al. 2017). Tics have sometimes been characterized as involuntary, but they may instead be voluntary (or "semi-voluntary") responses aimed at alleviating the preceding premonitory urges (Hashemiyoon, Kuhn, and Visser-Vandewalle 2017; Jankovic 2001). Tics may be motor or phonic, and they are classified as simple, if they involve only a small group of muscles or simple oral-nasopharyngeal noises such as sniffing or grunting, or complex, if they instead involve several muscle groups or more elaborate phonic phenomena such as the utterance of words or phrases (American Psychiatric Association 2013; Robertson et al. 2017). TS has an estimated prevalence of 0.3–1% (Robertson et al. 2017).

10.1.2. Pathophysiology

TS is strongly (Conceição et al. 2017; Neuner, Schneider, and Shah 2013; Worbe, Lehericy, and Hartmann 2015) and likely causally (Caligiore et al. 2017; Pogorelov et al. 2015; Tremblay et al. 2015) mediated by disturbances in the motor cortico-basal ganglia-thalamo-cortical (CBGTC) loop, which seems to be strongly implicated in both simple and complex tics (Conceição et al. 2017; Pogorelov et al. 2015; Tremblay et al. 2015). The associative and limbic CBGTC loops are strongly implicated in attention-deficit/hyperactivity disorder (ADHD) and obsessive-compulsive disorder (OCD; Castellanos

et al. 2006; Fineberg et al. 2018; Maia, Cooney, and Peterson 2008; Makris et al. 2009; Norman et al. 2016; Tremblay et al. 2015), which occur in approximately half or even more of patients with TS (Hashemiyoon, Kuhn, and Visser-Vandewalle 2017; Robertson et al. 2017). Studies in animals, in fact, suggest that the same disruption in CBGTC loops may produce tics, complex tics and inattention with hyperactivity-impulsivity, or obsessive-compulsive symptoms depending on whether that disruption affects motor, associative, or limbic CBGTC loops, respectively (Grabli et al. 2004; Tremblay et al. 2015; Worbe et al. 2009).

The motor loop is implicated in the learning and execution of habits (Horga et al. 2015; Yin and Knowlton 2006). Habits correspond to stimulus-response (S-R) associations that initially are learned on the basis of outcomes but then become independent from such outcomes, thereby implementing "cached" action values (Daw, Niv, and Dayan 2006; Delorme et al. 2016; Yin and Knowlton 2006). Learning S-R associations bypasses the need to learn a model of the environment, so habit learning is often called "model-free" (Daw et al. 2011; Delorme et al. 2016). Such designation contrasts with that used for goal-directed learning, which relies on internal models of the world and is thereby often called "model-based" (Daw et al. 2011). The use of the term "model-free" can be somewhat confusing because many reinforcement learning (RL) models work in a model-free way (**Box 10.1**). The term "model-free" refers to the absence of an explicit internal model of contingencies in the world, not to the absence of a computational model.

The implication of the motor loop in both habits and tics is consistent with the idea that "tics are exaggerated, maladaptive, and persistent motor habits" (Maia and Conceição 2017, 401). Habit learning and execution, moreover, are strongly modulated by dopamine, which likely explains the role of dopamine in TS (Maia and Conceição 2017; 2018; Nespoli et al. 2018), as we will discuss in detail below (**Section 10.3**).

Cortical motor areas are organized hierarchically, with lower- and higher-order motor cortices being responsible for simpler and more complex movements, respectively (Kalaska and Rizzolatti 2013; Rizzolatti and Kalaska 2013; Rizzolatti and Strick 2013). This hierarchical organization likely explains why primary and higher-order motor cortices seem to be implicated in simple and complex tics, respectively (Worbe et al. 2010). Interestingly, and consistent with the implication of somatosensory regions in the premonitory urges that typically precede tics (Conceição et al. 2017; Cox, Seri, and Cavanna 2018), simple tics are associated with disturbances in somatosensory cortices that may be more

restricted to primary somatosensory cortex (Sowell et al. 2008), whereas for complex tics, those disturbances extend farther into higher-order somatosensory cortices (Worbe et al. 2010).

As briefly mentioned above, some evidence suggests that the associative CBGTC loop may also be implicated in complex tics (Tremblay et al. 2015; Worbe et al. 2012; 2009). The associative CBGTC loop is involved in goal-directed behaviors (Yin and Knowlton 2006), which may explain why complex tics often seem to have a more intentional character than simple tics do (American Psychiatric Association 2013).

10.1.3. Treatment

TS can be treated pharmacologically (Ganos, Martino, and Pringsheim 2017; Mogwitz et al. 2018; the ESSTS Guidelines Group et al. 2011a) or behaviorally (Fründt, Woods, and Ganos 2017; Robertson et al. 2017; the ESSTS Guidelines Group et al. 2011b). Consistent with the likely implication of dopaminergic hyperinnervation in TS (Buse et al. 2013; Hienert et al. 2018; Maia and Conceição 2018), patients with TS are typically prescribed antipsychotics (dopamine D_2 antagonists), because of their greater efficacy (Ganos, Martino, and Pringsheim 2017; the ESSTS Guidelines Group et al. 2011a), or α_2 agonists, which also reduce dopaminergic transmission (Maia and Conceição 2018), because of their more favorable side effects (Ganos, Martino, and Pringsheim 2017). Aripiprazole, an antipsychotic with a different mechanism of action (Casey and Canal 2017; Mailman and Murthy 2010), may be particularly efficacious for TS (Mogwitz et al. 2018). Indeed, in the striatum, aripiprazole may combine favorable actions both on postsynaptic D_2 receptors, where it may partially block the effects of endogenous dopamine, and on presynaptic D_2 receptors, where its effects may be more akin to those of an agonist, thereby reducing dopamine release (Maia & Conceição, 2018).

Behaviorally, the treatments with most evidence for efficacy are habit reversal therapy (HRT) and exposure with response prevention (ExRP), both of which are recommended as first-line treatments for TS (Fründt, Woods, and Ganos 2017). HRT trains patients to suppress tics by executing tic-competing responses, via the use of antagonistic muscles, following the detection of premonitory urges and/or early movements that precede tic execution (McGuire et al. 2014; Rizzo et al. 2018; Verdellen et al. 2011). In ExRP, patients are encouraged to suppress all tics while focusing on the premonitory urges,

so as to promote premonitory-urge habituation; in addition, the patient is often exposed to situations (in vivo or imaginarily) that tend to elicit tics, while being encouraged to suppress the tics and focus on the premonitory urges (Fründt, Woods, and Ganos 2017; Rizzo et al. 2018; the ESSTS Guidelines Group et al. 2011b). Although clinically the protocols for HRT and ExRP are different, their mechanisms of action might be similar or even the same (van de Griendt et al. 2013). Indeed, premonitory-urge habituation is also a key component of HRT (McGuire et al. 2014), and both therapies involve suppressing tics (through a competing response in HRT and through the patient's own strategies in ExRP). Both HRT and ExRP may also be administered within broader behavioral interventions (Fründt, Woods, and Ganos 2017). HRT, for example, is a primary component of the comprehensive behavioral intervention for tics (CBIT; Fründt, Woods, and Ganos 2017; McGuire et al. 2014), which is also currently recommended as a first-line treatment for TS (Fründt, Woods, and Ganos 2017). Although HRT and ExRP have the advantage of avoiding medication's side-effects, pharmacological treatment, or the combination of behavioral and pharmacological treatments, may be necessary for, at least, the most severely affected patients (Ganos et al., 2017).

Some patients are refractory to all pharmacological and behavioral treatments (Kious, Jimenez-Shahed, and Shprecher 2016). In such cases, for very severely affected patients, invasive treatments, such as deep brain stimulation (DBS; Akbarian-Tefaghi, Zrinzo, and Foltynie 2016; Baldermann et al. 2016; Hashemiyoon, Kuhn, and Visser-Vandewalle 2017) or even psychosurgery (Hashemiyoon, Kuhn, and Visser-Vandewalle 2017), may be justified. The best targets for these treatments are still under investigation but generally involve nodes or fibers in the CBGTC loops (Akbarian-Tefaghi, Zrinzo, and Foltynie 2016; Baldermann et al. 2016; Hashemiyoon, Kuhn, and Visser-Vandewalle 2017). Even such invasive treatments, however, can sometimes be only moderately successful (Akbarian-Tefaghi, Zrinzo, and Foltynie 2016; Baldermann et al. 2016; Hashemiyoon, Kuhn, and Visser-Vandewalle 2017).

10.1.4. Contributions of computational psychiatry

Despite substantial progress, fundamental questions concerning the etiology, pathophysiology, and, more importantly, the adequate treatment of TS remain (Hashemiyoon, Kuhn, and Visser-Vandewalle 2017; Robertson et al. 2017; Thenganatt and Jankovic 2016). There is a pressing need both for a more detailed and integrative mechanistic understanding of TS (and its treatment) and for practical, clinically relevant predictive tools (whether based on an understanding of mechanism or not). These two needs

align closely with the two main branches of computational psychiatry: theory- and data-driven, respectively (Huys, Maia, and Frank 2016; Maia 2015). Moreover, the potential of the combined fulfilment of these two needs relates to the potential of combining these two approaches to computational psychiatry (Huys, Maia, and Frank 2016; Maia 2015).

As described in the remainder of this chapter, computational-psychiatry work in TS has already started to address these needs. While these efforts are still in their early days, with much work remaining to be done, theory-driven computational psychiatry has already yielded a mathematically rigorous theory of multiple aspects of TS (**section 10.3**), data-driven computational psychiatry has started to yield proof-of-concept classifiers for automated TS diagnosis (**section 10.2.3**), and the combination of these approaches has started to characterize computationally the neurocognitive disturbances that may underpin TS (**section 10.2.1**).

10.2. Past and Current Computational Approaches to TS

Consistent with the implication in TS of disturbances in the dopaminergic system (reviewed below; **section 10.3**) and in the motor loop (reviewed above; **section 10.1.2**), multiple studies have reported alterations in RL and in habit learning in TS. Studies that have used data-driven approaches to automatically classify patients with TS, moreover, have offered additional evidence for the involvement of the motor loop in TS. In this section, we review the findings from these three lines of research: RL, habit learning, and automated classification in TS. Then, in the next section, we show how all those findings may be reconciled under the hypothesis that TS involves dopaminergic hyperinnervation.

10.2.1. Reinforcement learning in TS

Unmedicated patients with TS seem to have increased learning from rewards: they learned from rewards but not from punishments in a subliminal task (Palminteri et al. 2009), and they had increased internal reward values (R^+, **Box 10.1**) relative to controls in a motor skill-learning task (Palminteri et al. 2011). Two studies failed to find significant differences between unmedicated patients and controls in learning from rewards (Salvador et al. 2017; Worbe et al. 2011), including specifically in R^+ (Worbe et al. 2011). Both of those studies, however, had a substantially greater proportion of males in the patient group than in the control group: the ratio of males to females in patients vs. controls was 2.16 vs. 1.17, respectively,

in one study (Worbe et al. 2011), and 3.25 vs. 1.22, respectively, in the other (Salvador et al. 2017)[17]. The increased proportion of females in the control group in these studies could have masked increased learning from rewards in patients because females learn better from rewards than males do (Evans and Hampson 2015)—a finding that is consistent with higher striatal presynaptic dopamine synthesis capacity (Laakso et al. 2002) and possibly higher striatal dopaminergic innervation, as assessed by dopamine transporter binding (Wong et al. 2012), in females relative to males. Furthermore, the study that found no differences in R^+ between unmedicated patients and controls (Worbe et al. 2011) suffered from model identifiability issues (Maia and Conceição 2017) that we discuss briefly below (**section 3.2.2**). A third study did not find differences in learning from rewards between patients with TS, many of whom were unmedicated, and controls, other than differences due to ADHD comorbidity (Shephard, Jackson, and Groom 2016). That study, however, used a simple deterministic task, and accuracy throughout the task was very high for all participants; the task therefore likely engaged explicit, rule-based learning, which may be largely unaffected in unmedicated patients with TS (Maia and Conceição 2017).

Patients with TS on antipsychotics other than aripiprazole have consistently been reported to be impaired at learning from rewards: they learned from punishments but not from rewards in a subliminal task (Palminteri et al. 2009), and they had decreased R^+ relative to controls in two studies (Palminteri et al. 2011; Worbe et al. 2011). These findings are most likely due to the medication, because antipsychotics also decrease learning from rewards in healthy humans, patients with other disorders, and animals (Maia and Conceição 2017; Maia and Frank 2017). Patients on aripiprazole, unlike those on other antipsychotics, seem to have spared simple learning from rewards (Salvador et al. 2017; Worbe et al. 2011). The mechanisms of action of aripiprazole are different from those of other antipsychotics (Casey and Canal 2017; Maia and Conceição 2018), as aripiprazole is characterized by "functional selectivity" (Mailman and Murthy 2010), which may explain this difference in effects. Aripiprazole does impair more complex forms of learning—namely, counterfactual learning—in a dose-dependent manner, but that may be due to detrimental effects on executive function (Salvador et al. 2017). Indeed, counterfactual learning involves learning from the outcomes of actions that one did not take but could have taken, which requires more complex inference and therefore requires executive function.

[17] In both studies, statistical tests did not reveal a statistically significant difference in the sex distribution between the groups, but failure to reject the null hypothesis cannot be construed as proof of no differences.

10.2.2. Habits in TS

As noted above, the motor loop is associated with both habits and tics, which suggests that "tics are exaggerated, maladaptive, and persistent motor habits" (Maia and Conceição 2017, 401). Further support for that idea comes from a study that found that patients with TS over-rely on habits relative to goal-directed behaviors (Delorme et al. 2016). The same study, moreover, found positive correlations between (1) overreliance on habits and tic severity, (2) overreliance on habits and structural connectivity between motor cortex and putamen, and (3) tic severity and structural connectivity between supplementary motor cortex and putamen (the latter two in unmedicated patients only), thereby demonstrating an association among habits, tics, and increased structural connectivity within the motor loop. Other studies have also shown positive correlations between structural connectivity within the motor loop and both (1) tics in patients with TS (with patients, moreover, having increased structural connectivity within the motor loop relative to controls; Worbe et al. 2015) and (2) habit learning in healthy controls (de Wit et al. 2012).

Two older studies found that patients with TS performed worse than healthy controls did in the weather prediction task (Kéri et al. 2002; Marsh et al. 2004), a probabilistic classification task that was designed with the goal of probing the gradual learning of S-R associations (Knowlton, Squire, and Gluck 1994). Moreover, this impaired performance did not seem attributable to medication or comorbidities. Those articles interpreted the impaired performance as indicative of impaired habit learning; however, neither study included any of the tests that are now considered necessary to classify a behavior as a habit (Yin and Knowlton 2006), a particularly pertinent concern because performance in the weather prediction task may also rely on other cognitive processes (Price 2009). Neither study, moreover, disentangled learning from positive vs. negative prediction errors (**Box 10.1**); as we will discuss later (**section 10.3.2**), the reported impairments might therefore be a consequence of impaired learning from negative, but not positive, prediction errors in TS.

10.2.3. Data-driven automated diagnosis in TS

As noted above (**section 10.1.2**), somatosensory and motor regions are strongly implicated in premonitory urges and tics, respectively (Conceição et al. 2017; Cox, Seri, and Cavanna 2018; Worbe,

Lehericy, and Hartmann 2015). Consistent with such involvement, studies that have applied data-driven computational-psychiatry approaches to build classifiers using data from magnetic resonance imaging (MRI) suggest that sensorimotor regions are key to distinguish patients with TS from healthy controls or from patients with other neuropsychiatric disorders, as described next.

Three studies used MRI data to automatically distinguish medication-naïve children with TS from healthy children. The three studies were conducted by the same research group using substantially overlapping samples and the same machine-learning approach: support vector machines (SVMs) with cross-validation. The studies differed in the specific MRI modalities used: resting-state functional MRI (rs-fMRI; Wen et al. 2018), diffusion MRI (Wen, Liu, Rekik, Wang, Zhang, et al. 2017), and both structural and diffusion MRI (Wen, Liu, Rekik, Wang, Chen, et al. 2017). In one of the studies, children with TS had no comorbidities (Wen et al. 2018), and in the other two they had no comorbidities other than ADHD (Wen, Liu, Rekik, Wang, Chen, et al. 2017; Wen, Liu, Rekik, Wang, Zhang, et al. 2017). All three studies achieved classification accuracies above 85%, and they all implicated sensorimotor regions, or their connectivity, as key discriminating features between children with TS and healthy children. They also implicated several other regions—for example, the inferior frontal gyrus (IFG), which is strongly implicated in inhibitory control (Aron, Robbins, and Poldrack 2014)—or their connectivity as discriminating features.

Two additional studies used rs-fMRI data to distinguish children with TS from healthy children but using samples in which a considerable percentage of the children with TS was medicated (Greene et al. 2016; Liao et al. 2017). Both studies also used SVMs with cross-validation. One study used inter-hemispheric intrinsic functional connectivity and included only boys without comorbid ADHD or OCD; that study strongly implicated sensorimotor and limbic regions in successful discrimination, and it achieved a classification accuracy of over 90% (Liao et al. 2017). The other study, which did not exclude patients with comorbidities, strongly implicated connectivity within and between sensorimotor and/or cognitive-control regions in successful discrimination (Greene et al. 2016), but it had a much lower classification accuracy (~70%) than the other studies.

A final study used cortical and subcortical morphological variations to classify children or adults who were either healthy or diagnosed with one of several neuropsychiatric disorders, including TS (Bansal et al. 2012). The medication status of patients with TS was not reported. For each pair (or set) of

groups to be compared, the discriminating features were preselected as those that differed with high significance between those specific groups. For children with TS vs. healthy children, the chosen features involved the surface morphology of the right globus pallidus and hippocampus; for adults with TS vs. healthy adults, they involved only the surface morphology of the right hippocampus. Classification accuracy was remarkably high when grouping features into two groups (e.g. adults with TS vs. healthy adults), but not into three groups (e.g. adults with TS vs. adults with schizophrenia vs. healthy adults). Although the study used both leave-one-out cross-validation and multiple independent split-half replication analyses, the preselection of features for each discrimination seems to have used the full sample of subjects to be discriminated—including, as far as we can tell, the subjects held out for test in the leave-one-out cross-validation and split-half analyses—which may have introduced overfitting.

In short, consistent with the main theme of this chapter, automated classification studies highlight the importance of sensorimotor regions for the classification of patients with TS, although they also point to other potentially relevant regions. In terms of clinical application, however—the aim of applied computational psychiatry (Huys, Maia, and Frank 2016; Paulus, Huys, and Maia 2016)—this work has important limitations. Arguably, the most fundamental limitation of this work is that the nearly exclusive focus on classification of patients with TS vs. healthy controls does not respond to a real clinical need; clinicians face many difficult tasks in which they could use the help of computational psychiatry—e.g., prognosis, prediction of treatment outcome, or differential diagnosis (Huys, Maia, and Frank 2016)—but distinguishing patients from controls usually is not one of primary concern. Amongst all of the studies reviewed above, only one tried to tackle the more realistic problem of distinguishing between different disorders, and it even tried to tackle classification into more than two groups (Bansal et al. 2012). Unfortunately, as noted above, that study might have suffered from overfitting; moreover, even with the possible overfitting allowed by the feature-selection process, the study had very limited success in the classification into more than two groups, which highlights the difficulties inherent in that process. Ultimately, we hope that more researchers interested in automated classification turn their attention to problems with potential for real clinical impact (Huys, Maia, and Frank 2016; Paulus, Huys, and Maia 2016).

10.3 Case Study: An Integrative, Theory-Driven Account of TS

CBGTC (Neuner, Schneider, and Shah 2013; Worbe, Lehericy, and Hartmann 2015) and dopaminergic (Buse et al. 2013) disturbances have long been implicated in TS. To the best of our knowledge, however, there was no cohesive, integrated account capable of explaining the multiple findings in TS obtained with various methods: molecular imaging, pharmacology, structural imaging, tic-related and resting-state functional imaging, and experimental behavioral data. Given that parsimony is a fundamental principle of science, we recently suggested a mechanistic, integrated theory of TS that provides a unified explanation for these multiple findings (Conceição et al. 2017; Maia and Conceição 2017; 2018). First, we conducted a systematic review of all positron emission tomography (PET) and single-photon emission computed tomography (SPECT) studies of the dopaminergic system in TS and considered related postmortem studies and the mechanisms of action of all medications with proven efficacy in TS; we showed that the hypothesis that TS involves dopaminergic hyperinnervation—i.e., an increased number of dopaminergic terminals—provides a simple and unified explanation for all of those findings (Maia and Conceição 2018). Second, we used insights concerning the computational roles of phasic and tonic dopamine in action learning and selection (Collins and Frank 2014; Maia and Frank 2017) to formulate a computational description of how increases in phasic and tonic dopamine—themselves resultant from dopaminergic hyperinnervation—may promote tic learning and expression and also explain the findings from the studies that have assessed RL and habits in TS. This formulation also allowed us to explain detailed observations concerning the time course of action of antipsychotics in the treatment of TS that had previously been unappreciated (Maia and Conceição 2017). Third, we reviewed studies that used anatomical imaging, resting-state functional imaging, and tic-related functional imaging in TS, focusing on the relation between such data and the genesis and severity of tics and premonitory urges (Conceição et al. 2017); using that information, we expanded the theory that we had formulated (Maia and Conceição 2017) to explain the neural substrates of premonitory urges and the computational roles of such urges in tic learning and execution (Conceição et al. 2017). We review each of these three steps below (sections 10.3.1-10.3.3).

10.3.1. Dopaminergic hyperinnervation as a parsimonious explanation for neurochemical and pharmacological data in TS

PET/SPECT studies of the dopaminergic system suggest that patients with TS have increases in dopamine transporter (DAT) binding, amphetamine-induced dopamine release, and possibly also in vesicular monoamine transporter 2 (VMAT2) binding and F-Dopa accumulation (Maia and Conceição

2018). Dopaminergic hyperinnervation would be expected to cause all of these findings (**Figure 10.2;** Maia and Conceição 2018). The full ensemble of findings of PET/SPECT studies of the dopaminergic system in TS presents important interpretational challenges, and typically there are about as many studies with null findings as studies with positive findings supporting the disturbances we mentioned above. Careful consideration of the studies with null findings, however, shows, first, a widespread and unmistakable lack of power—most studies used extremely small samples—and, second, in several studies, important age-, sex-, and/or medication-related confounds (Maia and Conceição 2018). At the moment, therefore, the dopaminergic-hyperinnervation hypothesis seems to successfully reconcile all extant PET/SPECT findings (Maia and Conceição 2018). This hypothesis has recently received additional support from a meta-analysis that confirmed increased striatal DAT binding in TS (Hienert et al. 2018). Furthermore, the dopaminergic-hyperinnervation hypothesis explains why all medications with well-established efficacy for TS—antipsychotics, low-doses of certain dopamine agonists like pergolide (which act mostly on presynaptic D_2 receptors), ecopipam (a selective D_1 antagonist), VMAT2 inhibitors, and even α_2/α_{2A} agonists—reduce dopaminergic transmission (Maia and Conceição 2018). Moreover, if indeed TS involves dopaminergic hyperinnervation, then it can be expected to involve increased tonic and increased phasic dopamine. As we will discuss in the next section, such increases explain a wide range of clinical and experimental findings in TS.

10.3.2. The roles of phasic and tonic dopamine in TS

Extensive evidence implicates phasic and tonic dopamine (**Box 10.2**) in action learning and selection (**Box 10.1**; **Figure 10.1**), and these effects have been elegantly captured computationally in the **OpAL model** (Collins and Frank 2014; Maia and Frank 2017). We have used these ideas, albeit under a slightly different mathematical instantiation from that in prior formulations of OpAL, to suggest that increased phasic and tonic dopamine in TS—themselves due to dopaminergic hyperinnervation (**Figure 10.2;** Maia and Conceição 2018)—may promote tic learning and expression (Maia and Conceição 2017).

The CBGTC-inspired RL model

OpAL (Collins and Frank 2014) expands the *actor* component of the actor-critic model (Barto 1995; Sutton and Barto 1998) to explicitly account for the existence of direct (Go) and indirect (**NoGo**) CBGTC pathways (**Box 10.1**; **Figure 10.1**). It does so by subdividing state-action preferences, $p(s,a)$ (where s and a denote a state and an action, respectively), into two "sub-preferences," $G(s,a)$ and

$N(s,a)$, which denote positive and negative parts of the preference, respectively, and are coded by Go and NoGo pathways, respectively (Collins and Frank 2014). Loosely speaking, the preference then becomes equal to the difference between these two sub-preferences: $p(s,a) = G(s,a) - N(s,a)$. As we will discuss next, however, these sub-preferences are differentially modulated by dopamine.

Go and NoGo striatal medium spiny neurons (MSNs) express mostly D_1 and D_2 dopamine receptors, respectively, which are excitatory and inhibitory, respectively (Soares-Cunha et al. 2016). For this reason, dopamine modulates the excitability (or gain) of Go and NoGo striatal MSNs in opposite directions: higher striatal dopamine levels increase and decrease the excitability of Go and NoGo MSNs, respectively, and lower dopamine levels have the opposite effects (**Figure 10.1**). The positive and negative action sub-preferences (*G* and *N*, respectively) are therefore differentially modulated by dopamine. In OpAL, this differential modulation is captured by using different gains for the positive and negative action sub-preferences: β_{G} and β_{N}, which represent the gains of the Go and NoGo pathways, respectively, and which are modulated by dopamine in opposite directions (Collins and Frank 2014; Maia and Frank 2017). This opposite modulation in OpAL is achieved in a simple formulation by making β_{G} and β_{N} depend on the level of dopamine, ω, with different signs:

$$\beta_{\mathrm{G}} = \beta(1 + \omega),$$

and

$$\beta_{\mathrm{N}} = \beta(1 - \omega),$$

where β is a constant. The dopamine-modulated preference then becomes: $p(s,a) = \beta_{\mathrm{G}} G(s,a) - \beta_{\mathrm{N}} N(s,a)$. Actions can then be selected, for example, with the softmax, as in other RL models (**Box 10.1**), but using these dopamine-modulated preferences.

We previously associated mostly tonic dopamine with action selection (Maia and Conceição 2017), under the assumption that phasic firing of dopamine neurons occurs only in specific circumstances and its corresponding transients are short-lived (Venton et al. 2003). Tonic dopamine levels, which are in the nanomolar range (Sulzer, Cragg, and Rice 2016), likely act mostly on the NoGo pathway because, in the striatum, D_1 and D_2 receptors appear to be predominantly in low- and high-

affinity states, respectively (Dreyer et al. 2010; Sulzer, Cragg, and Rice 2016). Thus, we had suggested that, in most cases, the gain parameters could be simplified to:

$$\beta_{\mathrm{G}} = \beta,$$

$$\beta_{\mathrm{N}} = \beta(1 - \tau),$$

where τ represents tonic dopamine, which has a limited effect on β_{G}. On the other hand, we had already suggested that if action selection occurred shortly following phasic dopamine firing, it would be modulated by the corresponding dopamine transients (Maia and Frank 2017). In other words, we had suggested that $\omega = \tau + \rho$, where ρ represents phasic dopamine (Maia and Frank 2017), which could then affect β_{G}. Recent evidence (da Silva et al. 2018), added to other evidence (Howe and Dombeck 2016; Syed et al. 2016), suggests that phasic responses might commonly occur prior to self-initiated action, increasing the probability of, and invigorating, subsequent movement. Thus, we now favor the formulation in which action selection is commonly modulated by both tonic and phasic dopamine components, which can affect both β_{N} and, if there is a phasic dopamine component, β_{G}.

In addition to its effects during action selection, dopamine also has differential effects on plasticity (Lerner and Kreitzer 2011; Shen et al. 2008), and therefore learning, in the Go and NoGo pathways (Maia and Conceição 2017). Specifically, dopamine increases cause long-term potentiation (LTP) in corticostriatal projections to the Go pathway and may cause long-term depression (LTD) in corticostriatal projections to the NoGo pathway, whereas dopamine decreases may have the opposite effects, causing long-term potentiation (LTP) in corticostriatal projections to the NoGo pathway and possibly causing long-term depression (LTD) in corticostriatal projections to the Go pathway (Lerner and Kreitzer 2011; Shen et al. 2008). Thus, in our formulation of the OpAL model, prediction errors affect learning in the Go and NoGo pathways in opposite directions:

$$G_{t+1}(s_t, a_t) = \begin{cases} G_t(s_t, a_t) + \alpha_{\mathrm{G,LTP}}\delta_t, & \text{if } \delta_t \geq 0 \\ G_t(s_t, a_t) + \alpha_{\mathrm{G,LTD}}\delta_t, & \text{if } \delta_t < 0 \end{cases}$$

and

$$N_{t+1}(s_t, a_t) = \begin{cases} N_t(s_t, a_t) - \alpha_{\mathrm{N,LTD}}\delta_t, & \text{if } \delta_t \geq 0 \\ N_t(s_t, a_t) - \alpha_{\mathrm{G,LTP}}\delta_t, & \text{if } \delta_t < 0 \end{cases},$$

where the parameters $\alpha_{G,LTP}$, $\alpha_{G,LTD}$, $\alpha_{N,LTP}$, and $\alpha_{N,LTD}$ are learning rates between 0 and 1, and the sub-preferences $G(s_t, a_t)$ and $N(s_t, a_t)$ represent the strength of the corticostriatal synapses onto MSNs of the Go and NoGo pathways concerning the current state, s_t, and the selected action, a_t (Maia and Conceição 2017). Given that these sub-preferences are meant to represent synaptic weights, they are constrained to be greater than, or equal to, 0 (Collins and Frank 2014).

The CBGTC loops have exactly the right anatomy to implement these computations (**Figure 10.1**; Maia and Conceição 2017; Maia and Frank 2017).

Mechanistic explanation of behavioral findings on RL in TS

From the five RL studies reviewed above (**Section 10.2.1**), two seem particularly consistent with the dopaminergic-hyperinnervation hypothesis of TS (Palminteri et al. 2009; 2011). In one of these studies, which applied an RL task in which the cues were presented subliminally, unmedicated patients with TS learned from rewards, but they did not learn from punishments (Palminteri et al. 2009). Both of these effects are consistent with dopaminergic hyperinnervation: increased learning from rewards is consistent with increased phasic dopamine, given the role of increases in phasic dopamine in the learning from positive prediction errors (**Box 10.2**); decreased learning from punishments is consistent with increased tonic dopamine, which might blunt the signaling of negative prediction errors by phasic decreases in dopamine. Furthermore, in that same study, the opposite pattern was found in unmedicated patients with Parkinson's disease (PD), who learned from punishments but not from rewards (Palminteri et al. 2009). The finding of opposite patterns in unmedicated patients with TS and PD is particularly relevant because PD is characterized by dopaminergic hypoinnervation, and we hypothesize that TS is characterized by dopaminergic hyperinnervation—hence, the two disorders are hypothesized to have the opposite dopaminergic disturbances. Further evidence for the dopaminergic-hyperinnervation hypothesis comes from the finding that patients with PD on levodopa and dopamine agonists became like unmedicated patients with TS, learning from rewards but not from punishments.

In the other study that supports the dopaminergic-hyperinnervation hypothesis, unmedicated patients with TS had higher internal reinforcement, R^+ (**Box 10.1**), for a monetary reward as compared with healthy controls (Palminteri et al. 2011). Dopamine does not seem to be implicated in the hedonic value of the reinforcements (Berridge 2007), which, at first sight, relates more closely to R^+. However, unless the task is designed carefully and appropriate parameter-recovery simulations are conducted—see

discussion in the Supplemental Materials in (Maia and Conceição 2017), —the biological interpretation of RL parameters is often complex and can be misleading. Specifically, in the context of our discussion here, R^+ can potentially relate to the signaling of positive prediction errors, and hence to phasic dopamine, rather than to hedonic value—a possibility to which we now turn.

In tasks in which the only non-negligible reinforcement is a positive reward, r, prediction errors, δ, can be described by $\delta_t = R^+ - V(s_t)$, where R^+ is the internal value of r (**Box 10.1**)[18]. We have noted that dopaminergic hyperinnervation is expected to lead to an increase in phasic dopamine release in TS. Suppose that such increase is additive; in other words, suppose that the increase in phasic dopamine release in TS is well captured by an additive parameter, a, that scales δ_t into $\delta_t^{\mathrm{TS}} = \delta_t + a$. We can rewrite δ_t^{TS} as follows: $\delta_t^{\mathrm{TS}} = \delta_t + a = R^+ - V(s_t) + a = (R^+ + a) - V(s_t)$. In other words, the change in phasic dopamine release would be well captured by a change in R^+ to $R^+ + a$ (i.e., the change in phasic dopamine release would be captured as a change in the R^+ parameter). Of course, we do not know if the increase in dopamine release is additive. In fact, the dopaminergic-hyperinnervation hypothesis may suggest that the increase is multiplicative because more fibers would be available to release dopamine for the same signal. Even if the change is multiplicative, however, that may still lead to a change in R^+[19].

We cannot overstate the importance of considering in detail the meaning of RL parameters and of conducting appropriate tests to ensure that, in a given task, parameters are identifiable and capture the intended meaning (Maia and Conceição 2017; see in particular the Supplemental Materials). For example, whereas one study found increased R^+ in unmedicated patients with TS, as discussed above,

[18] For simplicity, but without loss of generality, we focus our exposition on prediction errors as calculated by the critic in bandit tasks.

[19] To see this, suppose that the increase in phasic dopamine release in TS is well captured by a multiplicative parameter, m, that scales δ_t into $\delta_t^{\mathrm{TS}} = m\delta_t$. In RL equations (**Box 10.1**), δ_t is multiplied by a learning rate (e.g., α), so this multiplicative transformation should in principle be better captured by a change in α such that $\alpha^{\mathrm{TS}} = m\alpha$. However, in models with a single learning rate (α) for both positive and negative prediction errors—as was the case in the study under consideration (Palminteri et al. 2011)—this learning rate cannot adjust to increase only learning from positive prediction errors. In fact, if α increases, that will increase learning from both positive and negative prediction errors—and, as already discussed, learning from negative prediction errors might be blunted, rather than increased, in TS. Hence, it is not too surprising that the improved learning from positive prediction errors is captured, at least in part, by an increase in R^+.

another found no alterations in R^+ (Worbe et al. 2011). We have shown through simulations, however, that the parameters in the latter study were not identifiable (Maia and Conceição 2017). Moreover, we also showed that blunted learning from negative prediction errors—i.e., a reduced α^-—would, by following the model-fitting procedures in that study, erroneously be reflected in a reduced value for R^+ (Maia and Conceição 2017, Supplemental Materials). Now, note that dopaminergic hyperinnervation in TS would cause both (1) increased learning from positive prediction errors, which, as noted above, could be captured as an increase in R^+, and (2) reduced learning from negative prediction errors, which, as we have shown in simulations, could be captured as a decrease in R^+. These two opposing effects might therefore cancel each other out, leading to the observed finding of no alterations in R^+ in unmedicated patients with TS vs. controls (Worbe et al. 2011).

In addition to these computational arguments, there is also empirical evidence that R^+ in these studies may have captured dopaminergic effects. Indeed, patients with TS on antipsychotics, which block dopamine, had reduced, rather than increased, values of R^+ in both studies (Palminteri et al. 2011; Worbe et al. 2011). Such reductions in R^+ values, like the finding that medicated patients with TS, contrary to unmedicated patients with TS, failed to learn from rewards in the aforementioned subliminal task (Palminteri et al. 2009), likely are explained, at least in part, by the fact that, when administered chronically, antipsychotics decrease the firing of dopaminergic neurons and decrease phasic and tonic dopamine (Maia and Conceição 2017). Antipsychotics also have other, more complex effects that further explain why they blunt Go learning (Maia and Conceição 2017).

Relatedly, the finding that medicated patients with TS, like unmedicated patients with PD, but unlike unmedicated patients with TS, learned from punishments in the subliminal task that we first discussed (Palminteri et al. 2009) may be explained by considering other effects of antipsychotic administration. Except for aripiprazole, antipsychotics seem to exert their beneficial effects in TS by blocking postsynaptic D_2 receptors (**section 10.1.3**). Computationally, the blockade of D_2 receptors in NoGo MSNs translates into an increase in the excitability (or gain, $\boldsymbol{\beta_N}$) of the NoGo pathway, as well as into a tendency for strengthening of corticostriatal synapses onto NoGo MSNs, given that LTP and LTD in such synapses respectively depend on the lack of stimulation and stimulation of D_2 receptors (**Figure 10.1**; Maia and Conceição 2017). Such effects therefore explain why patients with TS under

antipsychotics learn better from punishments than unmedicated patients with TS do[20].Mechanistic explanation of behavioral findings on habits in TS

As reviewed in **section 10.2.2**, unmedicated patients with TS seem to over-rely on habitual, compared to goal-directed, behavioral control. As mentioned above, dopamine mediates habit learning and execution (**section 10.1.2**), with (1) increased phasic dopamine promoting excessive Go learning and (2) increased tonic dopamine, or increased phasic-dopamine release prior to action selection (da Silva et al. 2018), promoting excessive execution of the most ingrained motor actions (**Figure 10.1**). Thus, dopamine hyperinnervation provides a natural explanation for the reported over-reliance of unmedicated patients with TS on habits. In addition, the hypothesis that tics themselves are "exaggerated, maladaptive, and persistent motor habits" (Maia and Conceição 2017, 401) explains the observed positive correlation between the overreliance on habits and tic severity (Delorme et al. 2016).

As mentioned in **section 10.2.2**, two older studies found impaired habit learning in TS, but those studies did not disentangle learning from positive versus negative prediction errors (Kéri et al. 2002; Marsh et al. 2004). As mentioned above, under the dopaminergic-hyperinnervation hypothesis, one expects impaired learning from negative prediction errors because phasic dopamine decreases become blunted. The findings of those studies are therefore consistent with the dopaminergic-hyperinnervation hypothesis if they are driven mostly by impaired learning from negative prediction errors. A more recent study (Shephard, Groom, and Jackson 2018) lends further credence to this hypothesis. That study found that patients with TS, most of whom were unmedicated, were not impaired in a sequence learning task. However, they were impaired in switching from a sequenced block to a non-sequenced one, arguably the process that most relied on negative prediction errors. Interestingly, patients with TS on that study were also faster overall, in both sequenced and non-sequenced blocks, without a decrement on accuracy, possibly due to increased tonic dopamine. Indeed, increased dopamine, by increasing the gain of the Go relative to the NoGo pathway, should lead to faster responses overall (Collins and Frank 2014).

[20] Two other studies failed to find differences between unmedicated (Salvador et al. 2017) or mostly unmedicated (Shephard, Jackson, and Groom 2016) patients with TS and controls in RL tasks, but those studies suffered from confounds that were already discussed (**section 10.2.1**).

Mechanistic explanation of tic learning and expression in TS

As mentioned in **section 10.1.2**, dopaminergic hyperinnervation seems to explain why there is an increased propensity for tics to be learned and expressed in TS (**Figure 10.3A**), via increased phasic and tonic dopamine. Indeed, tic learning may be driven either by maladaptive, aberrantly-timed phasic-dopamine release or by phasic-dopamine released following the cessation of premonitory urges by tic execution (Maia and Conceição 2017); thus, tic learning is likely facilitated by the higher striatal phasic-dopamine release that is predicted to occur under dopaminergic hyperinnervation[21].

Furthermore, tic execution, like the execution of other well-learned motor actions, is likely facilitated by higher striatal dopamine levels—including both tonic and phasic dopamine—provided that the Go values of tics are considerable (**Figure 10.1**). Considering the aforementioned evidence on the overlearning of habits in TS (Delorme et al. 2016; Shephard, Groom, and Jackson 2018) and the association between tics and habits, the existence of tics with considerable Go values is likely generally the case in TS.

Tics, however, are not necessarily dependent on the existence of higher striatal dopamine levels, but rather on the existence of an overactivation of the Go compared to the NoGo motor pathway (**Figure 10.1**). Consistent with this idea, chronic administration of quinpirole, which is a D_2/D_3 agonist and therefore suppresses the NoGo pathway, causes tics in a juvenile-rat model of TS (Nespoli et al. 2018). In that rat model, however, dopaminergic projections to the dorsal striatum had been lesioned previously, which in itself is sort of the opposite of dopaminergic hyperinnervation, so these findings have to be interpreted with care. Nonetheless, chronic quinpirole administration, without prior lesioning of the dopamine system, also induces compulsive checking in rats (Szechtman, Sulis, and Eilam 1998), which is interesting given the very high comorbidity of OCD in patients with TS.

Mechanistic explanation of the therapeutic effects of medication in TS

In addition to providing a mechanistic explanation for the role of dopaminergic hyperinnervation in tics, the proposed CBGTC-inspired RL model seems to explain both the fast (**Figure 10.3B**) and cumulative

[21] The mechanisms underlying tic learning following the cessation of premonitory urges are comprehensively explained in **section 10.3.3** and in Conceição et al. (2017).

(**Figure 10.3C**) therapeutic effects of antipsychotics in TS (Maia and Conceição 2017), as well as possible increases in tic expression following withdrawal from antipsychotics (**Figure 10.3D**). The proposed model, moreover, seems to explain the therapeutic effects of all other medications with well-established efficacy in TS because all such medications reduce phasic and/or tonic dopaminergic neurotransmission (Maia and Conceição 2018).

As we suggested previously, from the medications with proven efficacy in TS, ecopipam, a D_1 antagonist, could be particularly interesting from a scientific perspective because, given the lower affinity of D_1 receptors, it should mostly antagonize phasic dopamine (Maia and Conceição 2018). We had associated phasic dopamine with tic learning, but not necessarily with tic execution (Maia and Conceição 2017). The existence of novel, strong evidence implicating phasic dopamine in action execution (da Silva et al. 2018), however, indicates that ecopipam should target tic execution, in addition to tic learning, which further helps to explain its efficacy in TS (Gilbert et al. 2018).

The aforementioned mechanisms of TS medication also help to develop a rationale for the combination of pharmacological and behavioral treatments. Successful execution of tic-competing responses in HRT and successful tic suppression in ExRP (see **Section 10.1.3**) both are conditional on the probability of tic execution not approximating 1. At least for the most severely affected patients (Ganos, Martino, and Pringsheim 2017), therefore, increasing NoGo relative to Go activation pharmacologically (e.g., through antipsychotic medications) may permit sufficient inhibition of the tic to allow the behavioral therapy to work. For other therapies, such as contingency management and massed negative practice, which work by assigning a negative value to tics, co-adjuvant medication may have a more direct therapeutic effect, by increasing learning and expression of negative values (Maia and Frank 2011). Chronic antipsychotic administration, for example, shifts the plasticity of corticostriatal synapses onto (motor) NoGo MSNs towards LTP, compared to LTD (Maia and Conceição 2017), besides leading to an increased gain of NoGo MSNs). By facilitating the NoGo learning and expression of tics (**Figure 10.3**), chronic antipsychotic administration might therefore conceivably promote the success of contingency management or massed negative practice. Although these therapies are yet to present convincing results when administered as monotherapies (Fründt, Woods, and Ganos 2017), we are not aware of any systematic attempts to combine them with pharmacological therapies.

10.3.3. Premonitory urges and tics in TS: computational mechanisms and neural correlates

Premonitory urges are aversive, distressful sensations that often precede, and are ceased by, tics (Brandt et al. 2016; Leckman, Walker, and Cohen 1993). Phasic dopamine is released following positive prediction errors (Schultz 2016; Maia 2009), including those elicited by the avoidance, and cessation, of aversive stimuli (Maia 2010; Navratilova and Porreca 2014; Seymour et al. 2005); thus, premonitory-urge cessation, via tic execution, may lead to phasic dopamine release. Such phasic release is possibly a key driver of tic learning (Conceição et al. 2017), via negative reinforcement (Capriotti et al. 2014)—that is, reinforcement due to escape from, or avoidance of, an aversive stimulus. However, as previously mentioned, aberrant, ill-timed phasic bursts may also reinforce tics.

Computational mechanisms of premonitory-urge-driven tic learning and execution

Two RL approaches may be used to describe how positive prediction errors may arise following premonitory-urge cessation (Conceição et al. 2017): an average-reward RL approach (Mahadevan 1996) and a standard RL approach [based on the actor-critic or similar state-value learning models (like OpAL models), but not on Q-learning models, which do not seem adequate to capture escape- or avoidance-learning behavior (as detailed in **Box 10.1**)]. In average-reward RL, there is an ongoing computation of a recency-weighted average reinforcement, $\bar{r}_{t-1}$, which is used to evaluate the obtained reinforcements (Mahadevan 1996). Specifically, in average-reward RL, a positive reinforcement is not necessarily "rewarding" (nor is a negative reinforcement necessarily "punishing", respectively) unless it is higher (lower, respectively) than the online estimate of the average reinforcement. Average-reward RL therefore attempts to optimize action policies by strengthening the associations between states and actions that yield a reinforcement higher than the average reinforcement at the time of action execution and by weakening those that yield a reinforcement lower than the average reinforcement. Given that, in this approach, all reinforcements are evaluated according to a mean reinforcement value, there is no mathematical reason to use a temporal discount factor γ (Mahadevan 1996), leading to the following equation for prediction-error calculation:

$$\delta_t = r(s_t) - \bar{r}_{t-1} + V_t(s_t) - V_t(s_{t-1}),$$

where $r(s_t)$ denotes the fact that the obtained reinforcement may be state dependent. This equation has been shown to capture pain-termination-driven RL in humans (Seymour et al. 2005), which we have previously hypothesized to parallel tic learning due to premonitory-urge termination in patients with TS (Conceição et al. 2017), as described next.

Premonitory urges, U, are inherently aversive sensations [$r(U) < 0$] that build up in time (Brandt et al. 2016); thus, immediately before premonitory-urge termination, $\bar{r}_{t-1}$ should typically be much lower than 0 [$\bar{r}_{t-1} \ll 0$]. Therefore, unless premonitory-urge termination (via tic execution) is accompanied by a very negative reinforcement [$r(s_t) \ll 0$], the term $r(s_t) - \bar{r}_{t-1}$ should be sufficiently positive to guarantee that, following premonitory-urge termination, $\delta_t > 0$, irrespective of the difference in state values at that time, $V_t(s_t) - V_t(s_{t-1})$, thereby strengthening the tic. Thus, under average-reward RL, state values are likely not necessary to explain tic learning, although they may certainly play a role (Conceição et al. 2017).

In standard RL (based on state-value learning models; see above), no average reinforcement is computed. Instead, prediction errors are calculated by:

$$\delta_t = r(s_t) + \gamma V_t(s_t) - V_t(s_{t-1}),$$

where γ is the aforementioned temporal discount factor (**Box 10.1**). The termination of an aversive premonitory urge does not per se result in a reward; in other words, $r(s_t)$ will not in general be positive. Thus, in standard RL, the elicitation of the positive prediction errors that may underlie tic learning is explained in terms of differences in state values (Conceição et al. 2017). Specifically, the combination of the aversive character of premonitory urges and the fact that they predict their own continuation means that the state of having a premonitory urge has a negative value [$V(U) < 0$]. Given that a tic terminates (even if only temporarily) a premonitory urge, the tic elicits a transition from a state with a negative value to a state with a neutral value, which produces a positive prediction error that reinforces the tic (Conceição et al. 2017). In other words, if it is assumed, for simplicity, that, on average, (1) $V_t(s_t)$ has no intrinsic value—because s_t is no longer characterized by the presence of a premonitory urge—and that (2) premonitory-urge termination is accompanied by a null, or no, primary reinforcement, $r(s_t)$, the prediction-error equation is therefore simplified into: $\delta_t = 0 - V_t(U) \Rightarrow \delta_t > 0$. Strikingly, however, δ_t would still be positive if premonitory-urge termination was accompanied by a negative $r(s_t)$ (e.g., social embarrassment) and/or by a negative $V_t(s_t)$, provided that $r(s_t) + \gamma V_t(s_t)$ is less negative than $V_t(U)$, which seems a reasonable assumption for most cases, given that premonitory urges are so aversive that they are often considered more distressing and life-impairing than tics themselves (Leckman, Walker, and Cohen 1993).

The positive prediction error elicited by premonitory-urge termination will tend to strengthen the association between the preceding state—having the premonitory urge—and the tic. Thus, the state of having a premonitory urge will come to elicit the tic, so premonitory urges will themselves come to elicit tic execution.

Neural correlates of premonitory-urge-driven tic learning and execution

The insula and somatosensory cortices are strongly implicated in premonitory urges (**Figures 10.4 and 10.5**; Cavanna et al. 2017; Conceição et al. 2017; Cox, Seri, and Cavanna 2018). In fact, the insula is also strongly implicated in natural urges (Jackson, Parkinson, Kim, et al. 2011) and in urges in addiction (Naqvi and Bechara 2010). Moreover, the insula is strongly implicated in interoceptive processing (Quadt, Critchley, and Garfinkel 2018), whose interaction with exteroceptive processing, in which the somatosensory cortices are strongly implicated, seems to underlie premonitory urges (**Figure 10.5**; Cox, Seri, and Cavanna 2018). Abnormal interoceptive sensibility, in particular, correlates with both the severity of premonitory urges and tics in TS (Rae, Larsson, et al. 2018). Furthermore, the insula and somatosensory cortices are both structurally and functionally abnormal in TS, in addition to being aberrantly coupled structurally and functionally with regions from the motor CBGTC loop (**Figure 10.5**; Conceição et al. 2017; Cox, Seri, and Cavanna 2018; Rae, Polyanska, et al. 2018; Sigurdsson et al. 2018; Wen et al. 2018) that are implicated in tic learning and execution (Conceição et al. 2017; Maia and Conceição 2017). Like abnormal interoceptive sensibility, insular (structural and/or functional) connectivity has also been shown to correlate with both tic and premonitory urge severity (Conceição et al. 2017; Rae, Polyanska, et al. 2018; Sigurdsson et al. 2018). The aforementioned abnormalities involving the insula and somatosensory cortices and their connections to the motor CBGTC loop are thereby likely to provide the substrate for premonitory urges and premonitory-urge-driven tic execution (**Figure 10.4**; Conceição et al. 2017).

The insula, together with the ventral striatum (VS), is also strongly implicated in RL (Garrison, Erdeniz, and Done 2013; Palminteri and Pessiglione 2017; Seymour et al. 2004; 2005)—particularly, in the case of the insula, with aversive outcomes (Garrison, Erdeniz, and Done 2013; Palminteri and Pessiglione 2017; Palminteri et al. 2012). Indeed, the insula has been strongly implicated in the coding of aversive state values [$V(s) < 0$] (Palminteri et al. 2012; Seymour et al. 2004), aversive prediction errors ($\delta < 0$) (Garrison, Erdeniz, and Done 2013; Seymour et al. 2004; 2005), and aversive outcomes ($r < 0$), even when such outcomes are fully predicted and therefore do not elicit a prediction error

(Conceição et al. 2017; Nitschke et al. 2006). The insula is therefore a prime candidate to represent three of the tic-learning-related variables mentioned in the previous subsection: the intrinsic negative primary value of a premonitory urge [$r(U) < 0$], its associated negative state value [$V(U) < 0$], and the negative prediction errors ($\delta < 0$) associated with the onset of premonitory urges and their estimation over time (not to be confused with the *positive* prediction errors elicited by the *offset* of premonitory urges when a tic is executed; **Figure 10.4**; Conceição et al. 2017).

More speculatively, the shell of the nucleus accumbens, which is a part of the VS, may be a good candidate to represent the average reinforcement, $\bar{r}$, over time (**Figure 10.4**; Conceição et al. 2017) [see also Niv et al. (2007) for a related proposal]. Indeed, dopamine in the shell has a set of unique properties that should, in principle, allow the online computation of $\bar{r}$. Calculating $\bar{r}$ requires (1) inputs representing the (signed) primary reinforcers, *r*, and (2) a mechanism for the integration of those inputs over time. Dopamine in the shell might fulfill these two requirements: (1) appetitive and aversive stimuli have been shown to respectively cause phasic increases and decreases of dopamine in the shell, in a manner that does not seem to depend on how predictable or unpredictable such stimuli were (McCutcheon et al. 2012; Sackett, Saddoris, and Carelli 2017) [see also Roitman et al. (2008)]—so, dopamine in the shell might represent the signed value of primary reinforcers; (2) DAT expression is comparatively low in the VS (Haber 2011), thereby permitting the slow integration of the shell's dopaminergic inputs over time, which would not be possible if the amount of DAT in the shell was such that phasic dopaminergic changes were always rapidly nullified, via dopamine reuptake (Conceição et al. 2017).

The existence of direct and indirect projections from both the shell and the insula to the VTA, in turn, explains how the VTA may have access to all variables that are necessary to calculate the prediction errors that are implicated in tic learning, as well as in the learning of related state values (**Figure 10.4**; Conceição et al. 2017). Finally, striato-nigro-striatal spirals (Haber 2011) may allow the propagation of the prediction errors implicated in tic learning from the ventral to the dorsal striatum, where they can be used to update the Go and NoGo values of actions (in this case, the Go and NoGo values of tics) stored in the corticostriatal synapses onto D_1 and D_2 MSNs, respectively (**Figure 10.4**; Conceição et al. 2017).

Premonitory urges and tics: clinical implications

Considering the likely causal role of premonitory urges in tic learning and execution, we have previously suggested that optimal treatment strategies for TS would likely have to act upstream of tics, in premonitory-urge related processes, possibly by targeting the insula and/or the somatosensory cortices (Conceição et al. 2017). In line with such prediction, successful tic reduction via high-frequency DBS of the thalamus was shown recently to correlate with changes in the activity of both the insula and sensorimotor regions (Jo et al. 2018). This prediction also has potential implications for repetitive transcranial magnetic stimulation (rTMS) treatment in TS (Conceição et al. 2017). Indeed, rTMS over the motor cortices has yet to prove better than sham stimulation in TS (Hsu, Wang, and Lin 2018), possibly because it is acting too downstream; it certainly seems that it would be worth trying rTMS over the insula or the somatosensory cortices to see if, by acting farther upstream, the effect would be better.

10.4. Discussion

10.4.1. Strengths of the proposed theory-driven account: a unified account that explains a wide range of findings in TS

The hypothesis that TS involves dopaminergic hyperinnervation provides a parsimonious and integrated explanation for extant neurochemical and pharmacological data in TS (Maia and Conceição 2018). Such hypothesis, moreover, is supported by a recent meta-analysis that found significantly increased striatal DAT binding in patients with TS, compared to controls (Hienert et al. 2018). Still, additional research on this issue is necessary because those meta-analytic findings became non-significant when controlling for age (Hienert et al. 2018), and most studies of the dopaminergic system in TS had very small samples and were subject to various other confounds (Maia and Conceição 2018).

Extensive evidence implicates phasic and tonic dopamine in action learning and selection, respectively (Collins and Frank 2014; Maia and Frank 2011; 2017), with recent evidence also implicating phasic dopamine in action selection (da Silva et al. 2018). Dopaminergic hyperinnervation would be expected to increase both phasic and tonic dopamine, which, in turn, would thereby increase both tic learning and expression (Maia and Conceição 2017). Hyperdopaminergia should also increase learning from rewards and increase habit learning, which are the main findings from RL and habit-learning studies in TS, respectively (Maia and Conceição 2017).

Finally, a ubiquitous clinical characteristic of TS is the presence of premonitory urges, which are alleviated temporarily by tics (Brandt et al. 2016; Leckman, Walker, and Cohen 1993). We have argued that the termination of premonitory urges likely elicits positive prediction errors, which reinforce tics (Conceição et al. 2017). These positive prediction errors likely elicit phasic firing of dopamine neurons, which again links to hyperdopaminergia in TS. A detailed consideration of the neural substrates of premonitory urges and their interactions with the motor system further explains how premonitory urges might play a role not only in tic learning but also in tic execution (Conceição et al. 2017). In short, our theoretical account, reviewed in **Section 10.3**, provides a rigorous and comprehensive account of a wide range of findings in TS (**Figure 10.5**).

10.4.2. Limitations and extensions

Other regions and neurochemical disturbances

We focused on the role of the motor CBGTC loop in tics and on the roles of the somatosensory cortices and insula in premonitory urges (as summarized in **Figures 10.4** and **10.5**). Multiple other regions, however—among which, for example, the cerebellum [which is bidirectionally connected with the basal ganglia via disynaptic connections (Bostan and Strick 2018)] and the IFG—have been strongly implicated in TS (Caligiore et al. 2017; Jo et al. 2018; Neuner, Schneider, and Shah 2013; Wen, Liu, Rekik, Wang, Chen, et al. 2017; Wen et al. 2018; Wen, Liu, Rekik, Wang, Zhang, et al. 2017). We also did not address possible differences between the mechanisms and neural correlates underlying motor vs. phonic tics, but some evidence suggests that phonic tics might be specifically related to limbic regions (Foltynie 2016; Jo et al. 2018).

We also focused on the involvement of the dopaminergic system, and specifically dopaminergic hyperinnervation, in TS. Multiple neurochemical abnormalities, however, have been implicated in TS (Cox, Seri, and Cavanna 2016; Kataoka et al. 2010; Lennington et al. 2016; Robertson et al. 2017), and some evidence suggests that targeting neuromodulators other than dopamine may also be beneficial for patients with TS (Augustine and Singer 2019; Thenganatt and Jankovic 2016). We should therefore emphasize that our focus on dopamine is not meant to imply that dopamine is the only or even the primary disturbance in TS. Other disturbances may also lead to TS if, like dopaminergic

hyperinnervation, they have the same circuit-level effects: increased plasticity and excitability of the Go, relative to the NoGo, motor pathway (Conceição et al. 2017; Maia and Conceição 2017).

Inhibitory control in TS

There has been substantial interest in inhibitory control in TS due to the hypothesis that tics might result from impaired inhibitory control (Morand-Beaulieu et al. 2017). Alterations in inhibitory control in TS would be consistent with the implication of the IFG in classification studies discriminating patients with TS from controls (**Section 10.2.1**) because the IFG is strongly implicated in inhibitory control (Aron, Robbins, and Poldrack 2014). Indeed, recent meta-analytic evidence seems to suggest that patients with TS have impairments in inhibitory control[22]. However, the same meta-analysis reported that inhibitory control was significantly more impaired in patients with TS that had comorbid and in patients with TS under medication than in unmedicated patients with TS, who were themselves quite similar to controls (Morand-Beaulieu et al. 2017). Thus, impaired inhibitory control in patients with TS may be a consequence of comorbid ADHD—indeed, inhibitory control is substantially impaired in ADHD (Brocki et al. 2007; Willcutt et al. 2005)—or of the medications. Some researchers, in fact, have even suggested that TS might involve enhanced (compensatory) cognitive control, including enhanced inhibitory control (Baym et al. 2008; Jackson, Parkinson, Jung, et al. 2011; Jung et al. 2013).

Appearance of tics before premonitory urges in development

We emphasized tic learning through negative reinforcement due to premonitory-urge termination, but children often develop tics before they start to report premonitory urges (Cavanna et al. 2017; Sambrani, Jakubovski, and Müller-Vahl 2016). One possible explanation for tic learning without premonitory-urge termination is that tics may also be learned due to excessive, "random" phasic dopamine transients. An alternative explanation for this seeming contradiction in timing during development is that children's failure to report premonitory urges does not mean that such premonitory urges are non-existent; it may simply mean that children are impaired at reporting them (Conceição et al. 2017; Martino, Ganos, and Worbe 2018). In line with this statement, children with TS are even impaired at reporting tics

[22] Cohen's $d = 0.33$, $p < 0.001$, when comparing patients with TS and controls, and Cohen's $d = 0.26$, $p < 0.01$, when comparing patients with TS without comorbidities and controls; Morand-Beaulieu et al. 2017.

themselves (Conceição et al. 2017), and the intensity of premonitory urges seems to be similar in children and youth of different ages (Raines et al. 2018; Steinberg et al. 2010; Woods et al. 2005). Also in line with the learning of tics through negative reinforcement, a recent report of two cases describes the onset of premonitory urges before tics and as early as at the age of 5 (Li et al. 2019). Moreover, subliminal learning from rewards is increased in TS (Palminteri et al. 2009), which might imply that subliminal learning from positive prediction errors in general—including those due to negative reinforcement—might be increased in TS. Thus, it is certainly possible that negative reinforcement plays a role even when premonitory urges fail to be reported (Conceição et al. 2017).

A related consideration is that urges may only become especially notorious for the individual when the individual attempts to inhibit the corresponding behavior. A parallel might be drawn here to natural urges (e.g., to urinate, etc.). There, too, early in development, children might feel an urge that, however, is almost too fleeting to be noticed because it is immediately followed by the corresponding behavior (e.g., an urge to urinate, which leads immediately and necessarily to urination). It is only as children learn to inhibit the behavior that the urge may become more notorious and that the link between the urge and the behavior may become more apparent; prior to that, the urge and behavior may be linked into a single experiential unit that makes noticing the urge as a separate entity more difficult. The same thing may occur with tics: it may be only as children learn to inhibit tics that premonitory urges become more noticeable (even if they were there, and could support negative reinforcement, all along). In line with these ideas, premonitory-urge intensity is indeed higher during tic suppression (Brandt et al. 2016).

Feedback connections and oscillations in the basal ganglia

We focused on RL models formed by sets of equations that predominantly capture the feedforward functioning of CBGTC loops (**Box 10.1**; **Figure 10.1**). However, there are several phenomena, such as neuronal oscillations within CBGTC loops (Brittain and Brown 2014), that those models cannot capture due to the underlying circuit oversimplifications (see, e.g., Augustine and Singer 2019). Here, we address such oscillatory behavior because pathological, low-frequency neuronal oscillations have been implicated in hyperkinetic symptoms from several movement disorders (Ellens and Leventhal 2013; Neumann et al. 2018), including TS (Hashemiyoon, Kuhn, and Visser-Vandewalle 2017; Neumann et al. 2018). In TS, indeed, symptom severity has been shown to correlate with increased low-frequency

oscillations, namely with pallidal and thalamic oscillations in the theta and beta bands (Neumann et al. 2018), and some authors now believe that prolonged theta bursts may be specifically implicated in involuntary movements (Neumann et al. 2018). Furthermore, in TS, DBS-induced reductions of tic severity have been shown to correlate with a relative increase of oscillations in the higher-frequency, gamma band (Hashemiyoon, Kuhn, and Visser-Vandewalle 2017).

The subthalamic nucleus (STN) and the globus pallidus external segment (GPe) seem to be centrally implicated in oscillatory behavior within CBGTC loops (Frank 2006; Gatev, Darbin, and Wichmann 2006). The STN and Gpe are connected in a negative feedback loop, in which the STN stimulates the Gpe, which in turn inhibits the STN. This negative feedback loop, driven by other inputs, may underlie the oscillatory behavior within CBGTC loops (Frank 2006; Gatev, Darbin, and Wichmann 2006). The projections from both the STN and Gpe to the output nuclei of the basal ganglia, the globus pallidus internal segment / substantia nigra pars reticulata, in turn, allow the propagation of the aforementioned oscillations through CBGTC loops (Frank 2006). Simplified CBGTC-inspired RL models, like OpAL models (Collins and Frank 2014; Maia and Conceição 2017; Maia and Frank 2017; **section 10.3.2.1**), do not explicitly account for STN-Gpe bidirectional connectivity (see, e.g., **Figure 10.1**), which explains why those models cannot be used to describe oscillatory behavior within CBGTC loops. These models also do not capture other phenomena that may similarly contribute to such oscillatory behavior (see, e.g., Llinás et al. 2005).

Although the models we have emphasized do not themselves exhibit oscillatory behavior, our main hypothesis that TS involves dopaminergic hyperinnervation is consistent with the observed increase of low-frequency oscillations in TS (Neumann et al. 2018). Indeed, these oscillations are present not only in TS but also in levodopa-induced dyskinesias in PD, where they arise specifically with levodopa administration (Alonso-Frech et al. 2006), which shows a clear association with dopamine. Moreover, these oscillations are also present in dystonia (Ellens and Leventhal 2013; Neumann et al. 2017), which, as hypothesized in Neumann et al. (2018), again might relate to dopamine because dystonia involves increased D_1 receptor availability (Simonyan et al. 2017)—which, like hyperdopaminergia, should increase Go pathway activation. Thus, dopaminergic hyperinnervation in TS, and the consequent hyperdopaminergia, might be the cause of the observed low-frequency oscillations.

10.5. Chapter Summary

TS, a disorder characterized by tics, seems to be associated with dopaminergic hyperinnervation, which likely causes increases in both phasic and tonic dopamine (**section 10.3.1; Figure 10.5A**). Given the roles of phasic and tonic dopamine in habit learning and execution, these increases lead to an overactive habit system (**Figure 10.5B**), with tics likely being persistent, maladaptive motor habits. This relation between tics and habits explains why the motor loop is centrally involved in both (**Figure 10.5B**). The central role of hyperdopaminergia in the overactivity of the habit system in TS, with consequent tics, explains why all medications with well-established efficacy for TS reduce dopaminergic neurotransmission (**section 10.3.2; Figure 10.5A–B)**.

Tics are usually preceded by, and terminate, premonitory urges—aversive, distressful sensations in which the somatosensory cortices and insula seem to be strongly implicated (**Figure 10.5C**). Premonitory-urge termination likely elicits positive prediction errors, which are signaled by phasic dopamine and reinforce tics (**section 10.3.3**; **Figure 10.5C**). Under dopaminergic hyperinnervation, this signaling might be excessive, which, again, might contribute to tic learning.

In the context of this book, it is worth articulating briefly the overall strategy underpinning the theoretical proposals in this chapter. We started with a well-motivated computational model of the function of CBGTC circuits and the role of dopamine therein, for which there is much evidence independent of TS; to understand dysfunction computationally, it is fundamental to first understand function, to be able to understand how such function may be disrupted. We then used this model to investigate the multiple implications of a simple and parsimonious hypothesis about an underlying disturbance in TS—dopaminergic hyperinnervation. We found that this simple hypothesis, when considered in light of the model, provided an explanation for a very broad range of experimental and clinical findings in TS—ranging from experimental findings about reinforcement and habit learning in TS to clinical findings about the medications that are used to treat TS. This body of work therefore showcases one of the key uses of theory-based computational psychiatry: developing a rigorous and integrated mechanistic understanding capable of explaining and bringing together a wide variety of seemingly disparate findings.

Box 10.1. Commonly Used Reinforcement Learning Models

Two standard computational models from the machine-learning literature—*Q*-learning (QL) and the actor-critic (Barto 1995; Sutton and Barto 1998; Watkins 1989)—have been used commonly and with considerable success to capture reinforcement learning (RL) in animals, healthy humans, and patients with Tourette syndrome (TS) and several other disorders (see, for example, Frank et al. 2007; Maia 2009; Roesch, Calu, and Schoenbaum 2007; Worbe et al. 2011). We briefly review those models here for two reasons: (1) understanding these models is necessary to understand the alterations in model parameters that have been described in TS (reviewed in **section 10.2.1**); (2) these models—specifically, the actor-critic—provide the backbone for a more elaborate model that we have used to provide an integrated, mechanistic account of multiple aspects of TS (as discussed in **section 10.3**).

Both QL and the actor-critic perform "model-free RL": a potentially misleading term because these are computational models but a reflection of the fact that these models do not explicitly learn a model of the world contingencies. Instead, they use *prediction errors* (commonly represented by δ) to learn directly the equivalent of Thorndikian stimulus-response (S-R) associations. In RL, it is common to speak of *states* rather than stimuli; states are more general because they include stimuli, situations, and contexts (which may be external and/or internal). In addition, although in psychology "responses" can be distinguished from "actions" (Dickinson 1985), in RL there is typically no such distinction, so responses—the accurate psychological term in the context of S-R associations (Dickinson 1985)—are also called *actions*. Thus, in RL, learning S-R associations corresponds to learning weights linking states and actions, $w_t(s, a)$ (where the subscript t indicates that these weights will vary over time with learning). QL and the actor-critic learn such weights in slightly different ways (described below). In both cases, however, those weights can be converted into action probabilities: $P_t(a_i|s_t)$, which gives the probability of selecting action a_i in the state s_t at time t. A common formula to convert the weights into probabilities is the *softmax* (Sutton and Barto 1998):

$$P_t(a_i|s_t) = \frac{e^{\beta\, w_t(s_t,a_i)}}{\sum_{a_j} e^{\beta\, w_t(s_t,a_j)}},$$

where β is the inverse temperature or gain ($\beta \geq 0$). This equation ensures that actions with greater weights tend to be selected more often, with the degree to which that happens being controlled by β,

which therefore controls the degree of exploration (trying out random actions regardless of their weights) vs. exploitation (always selecting the action or actions with the greatest learned weights; Daw 2011; Sutton and Barto 1998).

QL and the actor-critic both learn the weights $w_t(s,a)$ in a way that seeks to maximize the expected total sum of future reinforcements (although, as noted above, they do so slightly differently):

$$E\left[\sum_{\tau=t}^{\infty}\gamma^{\tau-t}r_\tau\right],$$

where *t* denotes the current time, and r_τ denotes the reinforcement at time τ. This expected total sum of future reinforcements is formally called a *value*. The sum has an infinite number of terms because, formally, values consider all future reinforcements. The discount factor, γ ($0 < \gamma < 1$), which discounts future reinforcements, is therefore usually necessary to ensure that the sum converges.

We turn next to the specific meaning of the weights and the mechanism that supports their learning in QL (next subsection) and the actor-critic (subsequent subsection).

a) The standard QL model

In QL, the weight $w_t(s,a)$ corresponds to an estimate at time *t* of the value of performing action *a* in state *s*—i.e., the expected total sum of future reinforcements obtained by performing action *a* in state *s*. Such state-action values are commonly called *Q* values and represented as $Q_t(s,a)$ (Maia 2009; Sutton and Barto 1998; Watkins 1989). *Q* values are learned using prediction errors (δs) that consist of the difference between (1) the sum of the obtained reinforcement with the discounted estimated state-action value for the best action in the next state and (2) the state-action value estimated prior to action execution (Maia 2009; Watkins 1989):

$$Q_{t+1}(s_t,a_t) = Q_t(s_t,a_t) + \alpha\,\delta_t,$$

$$\delta_t = r_t + \gamma \max_{a_i} Q_t(s_{t+1},a_i) - Q_t(s_t,a_t),$$

where s_t and a_t are the state and the action executed at time t, respectively, α is a learning rate ($0 \le \alpha \le 1$), r_t is the reinforcement obtained at time t, γ is the future-discount factor, and the

$\max_{a_i} Q_t(s_{t+1}, a_i)$ term represents the estimated state-action value of performing the best action in the subsequent state (Maia 2009; Sutton and Barto 1998; Watkins 1989).

b) The standard actor-critic model

Instead of estimating state-action values, the actor-critic estimates the values of states, $V(s)$, which correspond to the expected sum of future reinforcements starting in state s (essentially marginalizing all actions that can be performed in that state). State values are stored and learned in the *critic* component of the actor-critic. As in QL, state values are also learned using prediction errors (δs) but correspond to the difference between (1) the sum of the obtained reinforcement with the discounted estimated value of the next state and (2) the prior value of the state (Barto 1995; Maia 2009; Sutton and Barto 1998):

$$V_{t+1}(s_t) = V_t(s_t) + \alpha_C \, \delta_t,$$

$$\delta_t = r_t + \gamma \, V_t(s_{t+1}) - V_t(s_t),$$

where α_C is the critic learning rate.

In the actor-critic, the weights $w_t(s, a)$ therefore do *not* directly correspond to state-action values. Instead, these weights, commonly called preferences, $p_t(s, a)$, and stored and learned in the *actor* component, are learned using the prediction errors calculated in the critic (Barto 1995; Maia 2009; Sutton and Barto 1998):

$$p_{t+1}(s_t, a_t) = p_t(s_t, a_t) + \alpha_A \, \delta_t,$$

where α_A is the actor learning rate. Contrary to state values, which, in time should converge to the value of the state, the preferences are unbounded.

c) Simplifying prediction-error calculation

In so-called *bandit tasks*, action execution under the current state (s_t) does not affect the transition to subsequent states (s_{t+1}, s_{t+2}, …; Sutton and Barto 1998). In such cases, the equations for prediction-error calculation may be simplified into

$$\delta_t = r_t - Q_t(s_t, a_t)$$

in Q-learning models or

$$\delta_t = r_t - V_t(s_t)$$

in actor-critic models.

For simplicity, in this chapter we generally adopt these simplified equations, except where otherwise noted.

d) Extending QL and actor-critic models to be more realistic biologically

Positive and negative prediction errors are signaled differently by dopaminergic neurons: positive prediction errors are signaled via burst-firing of dopamine neurons, and negative prediction errors are likely signaled via the duration of pauses in the firing of dopamine neurons (Maia 2009; Maia and Frank 2011). Specific dopaminergic disturbances may therefore affect the signaling of positive and negative prediction errors differently; thus, models intended to capture these effects need to distinguish between positive and negative prediction errors computationally. In addition, it is sometimes of interest to assess individual or between-group differences in the internal representation of reinforcements (e.g., a $1 reward may have a very different effect on different participants). These extensions may be captured by the following set of generalized equations:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha(\delta_t)\,\delta_t,$$

$$\delta_t = f(r_t) - Q_t(s_t, a_t),$$

$$\alpha(\delta_t) = \begin{cases} \alpha^+, & if\ \delta_t \geq 0 \\ \alpha^-, & otherwise \end{cases},$$

where α^+ and α^- are positive and negative learning rates, respectively, which capture learning following positive and negative prediction errors, respectively (Frank et al. 2007), and $f(r_t)$ denotes the internal value of the reinforcement obtained at time t. In tasks in which the only non-negligible reinforcement is a positive reward, r, $f(r_t)$ may be simplified to:

$$f(r_t) = \begin{cases} R^+, & if\ r_t = r \\ 0, & otherwise \end{cases}.$$

We mention this special case because the latter equation was used by two studies that assessed RL in TS (Palminteri et al. 2011; Worbe et al. 2011), and so we refer specifically to R^+ in **section 10.2.1.**

Unpacking the actor learning rate into two prediction-error-dependent learning rates, or capturing the internal values of reinforcements, can be done similarly in actor-critic models.

Although using two learning rates to capture differential learning from positive vs. negative prediction errors is a step in the right direction to make the models more realistic biologically, it is likely insufficient. Direct, or Go, and indirect, or NoGo, motor (and associative) CBGTC pathways respectively mediate motor (and cognitive) action facilitation and inhibition (**Figure 10.1**; Collins and Frank 2014; Maia and Frank 2011; 2017). Both phasic increases and phasic decreases of dopamine (see **Box 10.2**)—signaling positive and negative δs, respectively—may simultaneously affect the Go and NoGo motor (and associative) CBGTC pathways in opposite directions, and not necessarily with the same magnitudes. Thus, (at least) four learning rates—besides the critic learning rate(s) in the actor-critic, or analogous, frameworks—may be needed to appropriately model RL via the CBGTC loops (Maia and Conceição 2017).

It would similarly be possible to unpack the critic learning rate into two prediction-error-dependent learning rates or, indeed, into four learning rates that depended on both the sign of the prediction error and the pathway affected (Go or NoGo). Such unpacking, however, could be slightly trickier to interpret because the critic is often associated with the ventral (limbic) striatum (Maia 2009; O'Doherty 2004; Rothenhoefer et al. 2017), and the limbic indirect pathway presents considerable anatomical and neurochemical differences from motor and associative indirect pathways (Soares-Cunha et al. 2016). We therefore do not address α_{C} unpacking in this chapter.

Concerning action selection, there are alternatives to the *softmax*, some of which better orthogonalize the processes implicated in action selection (see, e.g., Guitart-Masip et al. 2012). Here, however, we focus exclusively on the *softmax* because it is widely used in the literature and because expanding it to use two inverse temperatures (or gains), rather than a single gain, has been used to model the differential effects of dopamine in the expression of the positive and negative values of actions learned through the Go and NoGo pathways, respectively (Collins and Frank 2014; Maia and Conceição 2017; Maia and Frank 2017). Striatal dopamine at the time of action selection promotes the expression of the learned positive values of actions while suppressing the expression of the learned negative values of actions, by increasing the gain of the Go pathway (β_{G}) while suppressing the gain of the NoGo pathway (β_{N}), respectively (as comprehensively explained in **Section 10.3.2**; **Figure 10.1**; Collins and

Frank 2014; Maia and Conceição 2017; Maia and Frank 2017). Irrespective of the chosen action-selection equation, several other processes (not addressed here) may also be considered during action selection (see, for example, Daw 2011; Guitart-Masip et al. 2012).

Although we have highlighted how models with more parameters might be more realistic biologically, increasing the number of parameters in a model can create problems with model identifiability—especially when the parameters are far from being orthogonal, as in RL. Either extreme care must be exercised in task design to ensure that the parameters in models with a larger number of parameters are identifiable, or one must resort to simpler models. Still, the physiological basis for the use of (at least) four learning rates and (at least) two inverse temperatures (Maia and Conceição 2017) means that careful interpretation of the results from simpler models is needed, even when they seem to nicely capture physiological processes. For example, suppose that a given pathology or medication causes an impairment in long-term depression of the Go pathway (which should normally occur following negative prediction errors; Maia and Conceição 2017). In a model with only two learning rates, α^+ and α^-, such an effect would likely be captured by a reduced α^-; if, naively, one assumed that α^- was only associated with the NoGo pathway, this might be interpreted as suggestive of a NoGo-pathway-related abnormality, which would not be correct in this specific case.

e) Differences between QL and actor-critic models

The differences between QL and actor-critic models, although seemingly subtle, have important implications. Indeed, QL and actor-critic models may, in some circumstances, lead systematically to prediction errors with distinct signs for the same action in the same state (because only actor-critic models consider the past outcomes of all actions in a given state, via state-values, when calculating prediction errors). Such a difference is extremely relevant when considering the role of dopamine in TS—and, indeed, in psychiatric disorders more generally—because, as noted previously, positive and negative prediction errors are coded differently by dopaminergic neurons.

One case that illustrates these differences is that of active-avoidance learning. In active-avoidance learning, animals have to learn to perform a response that avoids an aversive outcome that would otherwise occur. Under the actor-critic framework, the expectation of the aversive outcome elicits a negative value $[V(s) < 0]$. When the animal performs the avoidance response, the successful avoidance of the aversive outcome elicits a positive prediction error because the observed outcome is

null but the predicted outcome was negative [$\delta = 0 - V(s) > 0$ because $V(s) < 0$]; it is this positive prediction error that reinforces the avoidance response (Maia 2010). In QL, however, the prediction error is 0: the state-action value is 0 [i.e., $Q(s, a) = 0$ when a is the avoidance response] because the avoidance response has itself never been associated with a negative outcome, so the prediction error is also 0 [$\delta = 0 - Q(s, a) = 0$ because $Q(s, a) = 0$]. This null prediction error is therefore unable to reinforce the response. Thus, for the response to be learned, the subject needs to execute a potentially infinite number of candidate actions—all possible actions other than the avoidance response—and to learn that the execution of all such actions leads to a negative outcome. Only after all other actions have negative Q values will the avoidance response, with its zero Q value, become preferred to the other actions (see the *softmax* equation at the beginning of this Box). Although in very constrained laboratory situations the set of candidate actions may be fairly constrained—especially in experiments with humans, who can be instructed about the possible actions (e.g., two possible buttons to press)—such a learning process clearly does not generalize to the real world.

This line of reasoning, together with the fact that non-overlapping implementations of both the *actor* and the *critic* have been identified (Maia 2009; O'Doherty 2004), seems to suggest that actor-critic models capture the actual biological implementation in animals and humans better than QL models do. However, the superiority of actor-critic over QL models is not yet consensual, with some electrophysiological data in animals actually favoring the latter (Roesch, Calu, and Schoenbaum 2007).

Box 10.2. Tonic and Phasic Dopamine

In this chapter, we often mention *tonic* and *phasic* dopamine, as well as differences in the specific contributions of tonic and phasic striatal dopamine to action selection and learning. Here, we briefly explain the difference between tonic and phasic dopamine. This distinction arises because dopaminergic neurons may fire in two distinct manners: in a spontaneous, low-frequency, single-spike manner and in a high-frequency, burst manner (Grace and Bunney 1984b; 1984a).

Spontaneously active neurons fire at a baseline frequency of approximately 5 Hz (Grace and Bunney 1984b; Sulzer, Cragg, and Rice 2016; Wightman and Robinson 2002) due to the alternation between a slow, pacemaker-like depolarizing current and an ensuing afterhyperpolarization (Grace and Bunney 1984b). Such firing, together with the mechanisms underlying the synthesis, release, reuptake, and degradation of dopamine, defines the tonic dopamine levels, which are relatively stable and spatially homogeneous (Venton et al. 2003; Sulzer, Cragg, and Rice 2016). In the striatum, tonic dopamine levels range between 10–30 nM (Sulzer, Cragg, and Rice 2016).

Dopaminergic neurons that are tonically active can be driven to burst-fire, provided that there is incoming excitatory drive to those neurons (Grace and Bunney 1984b; Lodge and Grace 2011). Bursts correspond to a small number of action potentials (typically up to 10) at high frequencies, sometimes exceeding 30 Hz (Grace and Bunney 1984a; Wightman and Robinson 2002); thus, burst firing may cause abrupt, spatially heterogeneous, and massive increases in dopamine release (Grace and Bunney 1984a; Sulzer, Cragg, and Rice 2016; Venton et al. 2003). These large increases are called *phasic*; they are typically in the micromolar range and often—albeit not always (da Silva et al. 2018; Matsumoto and Hikosaka 2009; Wenzel et al. 2015)—signal positive prediction errors (Maia 2009; Schultz 2016). In addition to these phasic increases in dopamine, there are also phasic decreases. These occur when dopaminergic neurons temporarily pause firing, such as when a negative prediction error occurs (Maia 2009).

10.6. Further Study

Maia, T. V., & Frank, M. J. (2011) shows how a biologically detailed model of reinforcement learning in the basal ganglia, closely related to the models described in this chapter, sheds light on multiple

neuropsychiatric disorders: Tourette syndrome (TS), Parkinson's disease, attention-deficit/hyperactivity disorder, addiction, and schizophrenia. The article's proposals about TS are precursors for several of the ideas in this chapter. In addition, the article shows how a single model can help to understand not only TS but also multiple other disorders that similarly involve disturbances in the dopaminergic system and basal ganglia.

Maia, T. V., & Frank, M. J. (2017) reconciles evidence from studies of the dopaminergic system with behavioral and functional neuroimaging data from patients with schizophrenia, using a model and ideas akin to those that we later applied to TS and describe in this chapter. This inter-related treatment of TS and schizophrenia is particularly apt because, despite their very distinct clinical presentation, both are hyperdopaminergic disorders. Thus, for example, some of the insights on the effects of antipsychotics originally developed in the 2017 article have close parallels in this chapter's account of the effects of antipsychotics in TS.

Maia, T. V. (2010) shows how learning to (actively) avoid aversive outcomes relies on positive prediction errors, using an actor-critic framework. This link between negative reinforcement and positive prediction errors, which are signaled by phasic dopamine responses, provided much of the motivation for our suggestions linking the termination of premonitory urges with tic reinforcement in the context of hyperdopaminergia in TS.

Seymour, B et al., O'Doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005) showed how average-reward learning can explain the relief signal that occurs when a painful stimulus is terminated. This account formed the basis for one of our accounts of how the termination of premonitory urges can reinforce tics (although the two accounts reviewed in this chapter are closely related, as they both imply that the termination of premonitory urges elicits positive prediction errors that strengthen tics).

Hienert, M et al., Gryglewski, G., Stamenkovic, M., Kasper, S., & Lanzenberger, R. (2018) reports what is, to the best of our knowledge, the first and, to date, only meta-analysis of molecular-imaging studies of the dopaminergic system in TS—specifically, PET and SPECT studies of the dopamine transporter and D_2 receptor in the striatum. Consistent with the idea that TS involves dopaminergic hyperinnervation, the meta-analysis found significantly increased dopamine transporter binding in TS (although that finding was not entirely conclusive because it became non-significant after controlling for age).

10. 7. Acknowledgments

The authors acknowledge funding from Fundação para a Ciência e a Tecnologia, Portugal (Ph.D. Fellowship PD/BD/105852/2014 to VAC) and from the Tourette Association of America (to TVM). The authors also thank João Antunes for assistance in searching the literature for relevant articles, as well as their participation in a twinning project (SynaNet) from the European Union Horizon 2020 Programme (project number: 692340).

Chapter 11: Perspectives and Further Study in Computational Psychiatry

Peggy Seriès, University of Edinburgh

11.1 Processes and Disorders not covered in this book

This book described examples of methods and questions that are currently at the forefront of research in computational psychiatry, and which have offered new insights into the mechanisms that underlie several psychiatric disorders. The examples we have covered describe multiple levels of analysis, from the biophysically detailed level (**Chapter 3**) and network level (**Chapter 4**) to algorithmic and normative models that are more abstract using tools from reinforcement learning and Bayesian methods (**Chapter 5-10**).

There are a number of important topics that have not been covered in the current volume, however.

In terms of processes as defined by RDOC (see **Chapter 1**), **Chapter 3** and Chapter 4 cover cognitive and reinforcement processes respectively, and **Chapter 10** touches on motor processes. However, social processes (affiliation and attachment, social communication, perception and understanding of self and others) – a domain where research has grown significantly in the last years (see Hackel and Amodio (2018) for a recent review) - would deserve a much better treatment. Arousal and regulatory systems (circadian rhythms, sleep and wakefulness) are also absent from this volume but very little research exists in this area in the context of computational psychiatry.

In terms of disorders, similarly, the book only covers a subset of disorders. Notably absent are bipolar disorder and particularly mania, autism spectrum disorders (ASD), obsessive compulsive disorders (OCD), attention deficit hyperactivity disorder (ADHD), eating disorders, personality disorders (e.g., borderline, paranoid, antisocial personality disorder) and post-traumatic stress disorder (PTSD).

In the following, we offer some pointers on research that exists in those domains, as a starting point for the interested reader. This list is by no means exhaustive. While some of those topics are starting to attract a lot of interest from a computational point of view, others have only been addressed by a handful of studies to date. The idea we would like to convey is that understanding is still mostly lacking for those issues. Computational Psychiatry is still in its infancy and the field is wide open for interdisciplinary research progress.

11.1.1 Autistic spectrum disorder

There is a growing computational literature regarding autistic spectrum disorder (ASD).

A dominant theory is that ASD could be regarded as a disorder of prediction or Bayesian inference (Sinha et al. 2014; Palmer, Lawson, and Hohwy 2017). The general hypothesis is that the weight, also called 'precision' (see **Section 2.4.6**), ascribed to sensory evidence and prior expectations would be imbalanced in ASD, resulting in sensory evidence having a disproportionately strong influence on perception. This relatively stronger influence of sensory information could explain the hypersensitivity to sensory stimuli and extreme attention to details that are observed in ASD. The weaker influence of prior expectations would also result in more variability in sensory experiences. The desire for sameness and rigid behaviours could then be understood as an attempt to introduce more predictability in one's environment (Pellicano and Burr 2012). Furthermore, this could lead to prior expectations which are too specific, and which do not generalize across situations (Van de Cruys et al. 2014).

While all theories agree that the relative influence of prior expectations is weaker in ASD, the primary source of this imbalance has been debated: would it arise from increased sensory precision (i.e. sharper likelihood) or from reduced precision of prior expectations? While early authors argued for attenuated priors (Pellicano and Burr 2012), the hypothesis of increased sensory precision is currently gaining more traction (Lawson, Rees, and Friston 2014; Palmer, Lawson, and Hohwy 2017; Karvelis et al. 2018).

More recently, it has also been proposed that key differences in ASD could be in the extent to which participants can predict whether the environment is dynamically changing or whether it is

relatively stable (Lawson, Mathys, and Rees 2017). ASD may be associated with an overestimation of the volatility of the environment, which would then lead to a failure to make use of relevant priors.

Although the above theories have gained a lot of popularity, conclusive experimental evidence is still largely lacking.

The interested reader can consult Palmer, Lawson, and Hohwy (2017) and Haker, Schneebeli, and Stephan (2016) for recent reviews of the Bayesian approach.

At a more biological level, it has been proposed that deficits in prediction and inference could be related to an imbalance between excitation and inhibition in neural circuits (Rosenberg, Patterson, and Angelaki 2015). It is thought that, in the cortex, a key computation performed by neural circuits is that of "divisive normalization", which divides the net excitatory drive to a neuron by a measure of the local population activity. Alterations in divisive normalization, due to excitation/inhibition imbalances, may give rise to autism symptomatology (Rosenberg, Patterson, and Angelaki 2015). More experimental support is still needed to confirm this model.

11.1.2 Bipolar Disorder

Recent computational theories propose that bipolar disorder may be related to the perception of reward and its interaction with mood (Eldar et al. 2016; Mason, Eldar, and Rutledge 2017). When we are in a good mood, we may perceive rewards as better than they actually are. Reciprocally, when we are in a bad mood, we may perceive rewards as worse than they actually are. People whose moods bias their perception of rewards too strongly will be more likely to experience greater mood swings in reaction to the same sequence of good or bad events, potentially resulting in extreme behavior. Eldar et al. (2016) and Mason, Eldar, and Rutledge (2017) show that computational models based on such simple ideas can explain a range of symptoms observed in bipolar disorder.

11.1.3 Obsessive Compulsive Disorder

As described elsewhere in this book (**Section 2.3.3**, and **Chapter 5**), it is thought that decisions can arise from two distinct, parallel systems of instrumental control, called the goal-directed and habitual systems. In goal-directed control, choices are made depending on their likely affective outcomes as predicted by a model of the environment. In habitual control, on the contrary, choices aim to reproduce actions that were previously rewarded. Disorders of compulsivity have been associated with a bias towards model-free (habit) acquisition instead of towards the goal-directed system (Voon et al. 2015).

Additional insights into OCD might be gained by considering how beliefs and actions are coupled. For example, someone with OCD will tell you that they know their hands are clean, but nevertheless won't be able to stop washing them. Two things that are normally linked together—confidence and action—have become uncoupled. Using computational methods, it has been found that the degree to which action and confidence are uncoupled in a simple decision task correlates with OCD severity (Vaghi et al. 2017).

11.1.4 Attention Deficit Hyperactivity Disorder (ADHD)

Behaviorally, ADHD is best characterized by increased variability across multiple cognitive domains and timescales.

Ziegler et al. (2016) offers a detailed review of how drift-diffusion models (DDM, **Section 2.2**) of decision-making and reinforcement learning models (**Section 2.3**) have been applied to understanding individual differences in ADHD. They conclude that empirical studies agree with theories' prediction for a lower DDM drift rate and reduced reinforcement learning choice sensitivity ("noisier" SoftMax parameter).

At a more neurobiological level, Hauser et al. (2016) propose that this reduced choice sensitivity could be explained by impairments in neural gain i.e. the degree to which neural signals are amplified or suppressed, a computation commonly associated with catecholaminergic neurotransmitter systems (i.e., dopamine and noradrenaline). They suggest that impaired gain modulation could then explain ADHD abnormalities, in particular increased variability, spanning from behavior to neural activity.

11.1.5 Post-Traumatic Stress Disorder

Computational psychiatry is a promising tool for understanding PTSD (Seriès 2019). Indeed, it is commonly believed that PTSD results from abnormalities in learning during and after the traumatic event. Fear conditioning could explain why neutral stimuli (people, places, sounds, etc.) that have been associated with the traumatic event acquire the capacity to trigger and maintain anxiety long after the trauma itself. Why this association doesn't weaken over time could be explained either by the fact that it was abnormally strong in the first place or—more likely—due to deficits in extinction processes, i.e., a failure for the association to weaken when the same cues are re-encountered without leading to the traumatic event. This could be a result of patients' avoidance strategies: individuals with PTSD avoid encountering such cues again and thus may never experience them as being safe. Other theories assume, on the contrary, that PTSD is related to basic deficits in acquiring associations between specific cues and the traumatic event. This would result in associating the trauma with the environment as a whole, causing heightened contextual anxiety and/or overgeneralization of fear to all cues resembling the initial cues.

Despite the popularity of the theories mentioned above, the specific components of anomalous learning in PTSD remain unclear. Recently, however, research studies associating behavioral avoidance-learning tasks, computational modelling and fMRI have started to dissect how learning mechanisms could differ in PTSD. Homan et al. (2019) and Brown et al. (2018) found for example that combat-exposed veterans suffering from PTSD pay more attention to surprising aversive outcomes. The researchers could also identify the neural structures involved in these differences. This greater attention to perceived threat could in turn explain hypervigilant responses.

11.1.6 Personality Disorders

Computational perspectives in the fields of personality and personality disorders have been very limited. A recent review about the use of computational psychiatry methods in borderline personality disorder can be found in Fineberg, Stahl, and Corlett (2017). Lee (2017) offers a review of data and theories regarding paranoid personality disorder. Patzelt, Hartley, and Gershman (2018) also provide an

interesting discussion of the concept and promise of a computational phenotype—a collection of mathematically derived parameters that precisely describe individual differences in personality, development, and psychiatric illness.

11.1.7 Eating Disorders

Computational approaches have yet to provide detailed theories and models of eating disorders. However, a growing body of evidence suggests that patients with anorexia nervosa have impairments in value-based learning and decision making (Verharen et al. 2019). Similarly, binge eating disorders have been linked to impairments in making goal-directed decisions (Voon et al. 2015) and to biases towards exploratory behaviour (Morris et al. 2016).

11.2 Data-driven approaches

This volume focussed on theory-driven approaches, which, as described in **Chapter 1**, employ mechanistic models to make explicit hypotheses at multiple levels of analysis. On the other side of the spectrum, data-driven approaches use machine-learning to make predictions from high dimensional data and are generally agnostic as to underlying mechanisms. As the availability of large and multi-dimensional datasets is increasing, either through large neurophysiological studies, online behavioral studies and through the use of mobile devices like smartphones, data-driven approaches are currently getting a lot of attention. They are perceived as very promising ways to provide individual predictions of diagnosis, clinical outcome and treatment response.

Readers interested in learning about the advances and challenges in the use of big data and machine learning approaches in psychiatry can refer to the recent review by Rutledge, Chekroud, and Huys (2019). Steele and Paulus (2019) also discuss how machine-learning approaches applied to neuroscience data can impact clinical practice. Both reviews illustrate, based on recent studies, how objective and clinically useful predictions can be made for individual patients regarding diagnoses, illness severity, relapse, and psychotherapy or medication treatment outcomes. They also emphasize the fact that machine-learning techniques can be misapplied so care is needed in their use and interpretation.

It is important to note that data-driven and theory-driven approaches are not incompatible: theory-driven models can provide descriptions that efficiently summarize complex data and these summaries can provide inputs for machine learning algorithms. The combination of both methods has been found to outperform data-driven approaches alone (Huys, Maia, and Frank 2016).

11.3 Realizing the potential of Computational Psychiatry

As will hopefully be obvious from the previous chapters, Computational Psychiatry has already led to many insights into the neurobehavioral mechanisms that underlie several psychiatric disorders.

A number of tools have shown clinical potential (Paulus, Huys, and Maia 2016). For example, the development of theories and tasks related to model-free vs model-based learning has resulted in rich computational analyses and new insights in a variety of disorders, including substance abuse and OCD. Similarly, as described in many chapters of this book, theories about the role of dopamine in reinforcement learning have led to the development of tasks and models that have been applied to a wide range of disorders and can ultimately inform pharmacotherapy. It is often thought that computational assays, such as those based on Bayesian approaches, could help diagnostic tests (Haker, Schneebeli, and Stephan 2016). Computational psychiatry could also help psychotherapy: psychotherapy being a learning process, it may benefit from the rich computational understanding of learning processes (Moutoussis et al. 2018).

However, what is still lacking is a structured initiative to take computational psychiatry from the laboratory to the clinic.

As the field is maturing, there is a growing reflection about key developments required in the practice and infrastructure of computational psychiatry research to accelerate its clinical usefulness (Paulus, Huys, and Maia 2016; Browning et al. 2019; Teufel and Fletcher 2016).

These studies comment on the issue of measurement in computational psychiatry. Measurements usually involve choosing a behavioral task, to be modeled using an algorithm such as a reinforcement learning model and potentially used also in fMRI. It is important that the reliability and validity of such a computational assay be assessed and iteratively optimized. As mentioned in **Section 2.5**, parameter

identifiability needs to be assessed through analysis of parameter recovery, model recovery and model comparison. The reliability of the measurement also needs to be assessed, e.g. whether measurements are consistent across time can be assessed by test-retest reliability. Other important measures of assessment include clinical utility (is the measurement related to clinically important outcomes such as symptom scores, treatment response or illness course?) and convergent/divergent validity (does the measurement correlate with other measures of the same construct?). Meaningful measures for clinical purposes are then likely to consist not in one model parameter but in the relations between multiple such parameters within or across tasks. These relations can be obtained by collecting data from a range of related assays within a single population of participants and by using data-driven techniques such as clustering.

It is then crucial that the measured latent structures address clinically meaningful questions. This can be assessed by examining the predictive ability of the assay (e.g. can it predict response to treatment?) and/or the causal relationship between the process measured by the assay and clinically important outcomes such as symptoms. If causality can be established between the measurements and outcomes, the process measured by the assay can constitute a treatment target.

Related to this, an important issue is the recruitment of participants. For obvious practical reasons, emphasis has been in recruiting participants with mild and transient illness, rather than patients with severe and enduring illness or moderate-severe recurrent illness during periods of significant illness. Such bias in data collection could partly explain why progress in computational psychiatry has not yet been more significant.

Ideally, this process should be carried out at multiple sites involving individual laboratories that include a close collaboration between academic psychiatrists or psychologists and theoretical and computational neuroscientists. To be successful, the research environment must be developed to encourage large-scale, collaborative, interdisciplinary consortia.

11.4 Chapter Summary

Computational Psychiatry is still in its infancy. While the potential of the field is clear, there is still much to do to take computational psychiatry from the laboratory to the clinic. As the field matures,

improved and unified methodologies will be needed, as well as large-scale, collaborative, interdisciplinary consortia.

It is our hope that this book will inspire a generation of students who can make a difference.