From “Computational Psychiatry: New Perspectives on Mental Illness,”
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

12

Candidate Examples for a Computational Approach to Address Practical Problems in Psychiatry

Rosalyn Moran, Klaas Enno Stephan, Matthew Botvinick, Michael Breakspear, Cameron S. Carter, Peter W. Kalivas, P. Read Montague, Martin P. Paulus, and Frederike Petzschner

Abstract

Scientists and clinicians can utilize a model-based framework to develop computational approaches to psychiatric practice and bring scientific discoveries to a clinical interface. This chapter describes a general modeling perspective, which complements those derived in previous chapters, and provides distinct examples to highlight the scientific and preclinical research that can evolve out of a computational framework to offer new tools for clinical practice. It begins by reviewing areas of theoretical and modeling studies that have reached a critical mass and outlines the pathophysiological insights that have been revealed. Three particular models are used to demonstrate how clinical questions, relating to understanding disease mechanisms and predicting treatment response, could be potentially addressed using an integrated computational framework. First, the phasic dopamine temporal difference model shows how neurophysiological and neuroanatomical research, incorporated into a learning circuit model, provides a constrained hypothesis testing framework, related to the likely multiple mechanisms contributing to addiction. Second, a potential application of generative models of neuroimaging measurements (dynamic causal models of EEG data)

Group photos (top left to bottom right) Rosalyn Moran, Klaas Stephan, Cameron Carter, Michael Breakspear, Read Montague, Frederike Petzschner, Matthew Botvinick, Martin Paulus, Peter Kalivas, Read Montague, Rosalyn Moran, Matthew Botvinick, Martin Paulus, Peter Kalivas, group discussion, Michael Breakspear, Frederike Petzschner, Klaas Stephan, Matthew Botvinick and Read Montague, Rosalyn Moran, Cameron Carter

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

is described to predict individual treatment responses in patients with schizophrenia. The third example offers a novel approach to quantifying patient outcomes under a "recovery model" of psychiatric illness. This involves a dynamical system appraisal of allostasis, using the amygdala-HPA axis with its role in anxiety disorders and depression as a clinical target syndrome to which the model could be applied. In conclusion, consideration is given to the community efforts needed to support the validation of these and future applications.

Introduction

The promise of computational approaches to psychiatric clinical practice is evidenced by the breadth and scope of developments from computational and systems neuroscience that are targeted directly at understanding the etiology (Winterer and Weinberger 2004), pathogenesis (Kheirbek et al. 2012), and clinical course (Huys et al. 2015a) of psychiatric illnesses. We outline where these scientific points of contact are concentrated and how their development could further evolve into pragmatic tools for the practicing psychiatrist. The primary motivation for this endeavor is the lack of diagnostic technologies that could be used to probe whether the symptoms of a particular patient are more likely to be explained by one putative pathophysiological process or another. In other words, unlike cardiologists, for example, who are equipped with a mechanistic understanding of how the heart works and have access to a broad arsenal of tools for differential diagnosis, psychiatrists face a multitude of competing pathophysiological theories and lack the technology to disambiguate alternative disease mechanisms in individual patients. So far, the translation of the current corpus of neuroscience into the clinic has had limited penetration and practical value (Kapur et al. 2012; Millan et al. 2012). A fundamental goal of computational psychiatry is thus to develop a framework that situates the algorithmic properties of the brain (its information-processing capacity and the supporting neural substrates) as the basic scientific level of inquiry (Maia and Frank 2011) and to provide tools for inferring individual disease mechanisms and predicting individual clinical trajectories and treatment response (Stephan and Mathys 2014).

In this chapter, we consider a general computational approach that could be applied to produce interventions and characterizations across the history of a patient's disease (Figure 12.1). In much the same way as Flagel et al. (this volume) developed a computational model to formalize current psychiatric nosology and putative disease trajectories, our framework similarly employs models populated by discoveries from basic science for use by clinicians and patients to monitor disease risk, progression, and recovery. Specifically, we propose that unobservable biological parameters (B) affect unobservable computational parameters (C), which can be estimated using observable behavioral symptoms and signs (S), biological measurements (M), and diagnosis (D). Our model allows for treatment response prediction by incorporating

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

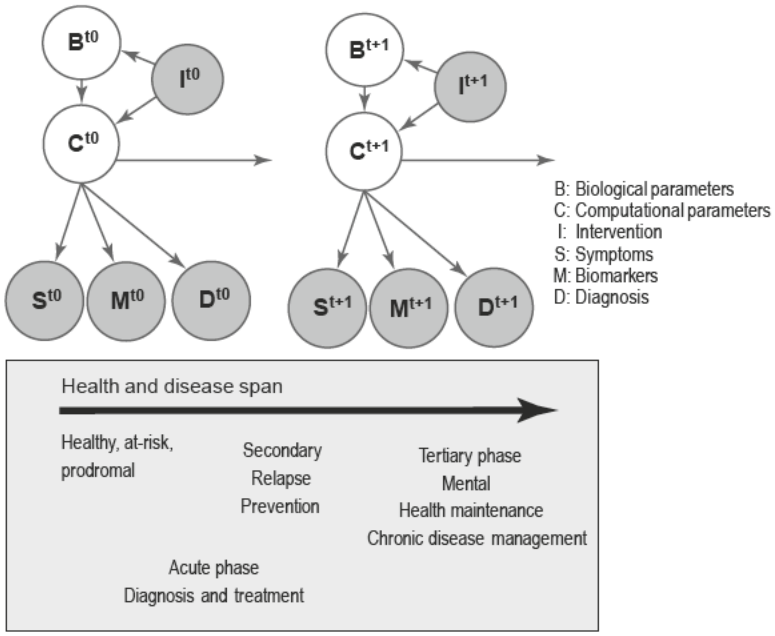

Figure 12.1 General model-based framework made up of a generative model to enable inference on the biological and computational causes of illness, and a formulation where therapeutic interventions affect model parameters. Model outputs are symptoms, biological measurements such as imaging data, and diagnosis. In principle, given interventions, symptoms, measurements, and diagnosis, a model inversion could be performed to infer disease causes, progression, and intervention effects as well as their trajectories over time.

intervention states (I) that can influence the underlying biological or computational variables. We envisage that this framework offers both a platform for testing putative computational and biological substrates of disease as well as a prototypical system that could eventually bridge basic science to a clinical end user, and present concrete examples of where this type of modeling platform could prove useful.

Different instantiations of this general model should be aimed at different stages of medical interventions. For example, when considering the challenge of primary prevention in healthy, at-risk, and prodromal stages, circuit and behavioral models of normal function are essential. From here, consideration of the acute disease phase necessitates developments in diagnostic and treatment prediction models, whereas the stage of secondary prevention is more concerned with relapse and recurrence susceptibility, prognostic classification, and monitoring. Finally, tertiary prevention during an emergent chronic phase mandates recovery-based approaches aimed at assisting a patient's

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

ability to function in daily life (e.g., learning to live with negative symptoms of schizophrenia).

Current developments span all of these stages with focused advances clustering around diagnosis and early detection. These have arisen from over a decade of work in computational neuroimaging (for review, see Stephan et al. 2015), which has produced normative descriptions of computational neural substrates in healthy populations, their developmental, and aging trajectories (Eppinger et al. 2013; Plichta and Scheres 2014; Thomas et al. 2014); the identification of unique neuronal correlates of high risk and prodromal states (Breakspear et al. 2015); and the identification of deviations from health in a range of psychiatric conditions (Frank et al. 2011; Morris et al. 2012; Robinson et al. 2012). In particular, algorithmic approaches to midbrain-striatal interactions have been used to uncover aberrancies in latent valuation and reward processing. These developments have been buttressed by biological circuit models that capture the downstream effects of these signals, helping to understand later, chronic disease-stage processes, such as the psychiatric symptoms of impulse control disorders that emerge following long-term dopaminergic treatment in Parkinson disease (Voon et al. 2011).

Ideally, the field of computational psychiatry will provide a broad armamentarium: from computational stress tests for teens at risk of schizophrenia, to a computational model that predicts pharmaceutical response in mania. In pursuit of these long-term goals, we develop specific examples of the general framework presented in Figure 12.1, to illustrate the "value added" by our burgeoning field. Overall, our objective pertains to inducing a conceptual shift in the qualification of psychiatric illness, placing computational formalism as a developing language designed to illuminate the link between the psychological and behavioral symptoms of psychiatric disorders and their neurobiological underpinnings. Our examples are designed to incorporate the translational potential of a shared mathematical brain language, whereby animal models of psychiatric disease can be compared and adjudicated in light of commonalities between theoretical and model-based constructs.

We begin by reviewing the state of the art and highlight those domains where a critical mass of knowledge has accrued to enable new scientific hypothesis testing in psychiatric illness as well as implementation and potential clinical take-up in the near term. We then proceed to develop focused prototypes based on our general model. These include computational approaches that have led to new and testable hypotheses in addiction research (the dopamine "fruit fly"), a computational model for predicting treatment responses in schizophrenia, based on a dynamic causal model (DCM) of neural circuit dynamics, and a model of allostatic regulation with relevance for chronic disease management. These candidate examples are diverse and generally intended to illustrate potential ways of moving computational models forward into the clinical application domain.

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

Cases of Computational and Theoretical Neuroscience Offering Biological Insight

To define the space of models that may be fit for purpose, let us first consider the definition of "computation" or "information processing" as it is relevant for psychiatry. Marr's tri-level hypothesis (Marr 1982; Ullman and Poggio 2010) partitions information processing in the brain into computational goals, algorithmic solutions, and implementational/physical levels. (As discussed by Kurth-Nelson et al., this volume, this nomenclature may not be ideal, and an alternative would be to refer to levels of purpose, computation, and implementation.) While computational psychiatry may fit most naturally at the intermediate algorithmic level (Montague et al. 2012), the field has developed biophysical instantiations of circuits that can also link psychiatric dysfunction to neurobiological substrates directly (Cooray et al. 2015). Thus for our toolkit, we consider both algorithmic (i.e., information-processing models) *and* implementation levels (i.e., biophysical models) as two broad categories that link computation and pathophysiology (for an overview, see Kurth-Nelson et al., this volume). Here we focus on examples that have provided direct insights into pathophysiology.

Biophysical Models

Biophysical models describe the dynamic activity of neurons, neuronal circuits, and large neuronal ensembles typically using either conductance-based models based upon simplifications of the Hodgkin-Huxley equations (Hodgkin and Huxley 1952; McCulloch and Pitts 1943; Morris and Lecar 1981) or current-based neural mass models (Freeman 1975). Their utility lies primarily in characterizing or identifying (through model inversion) the biological substrates of information transmission, ion channels, synaptic weights, transmitter levels, etc. These models are typically agnostic to the type of information processing (transformation of cognitive variables) that emerges from their activity. Exceptions do exist, however, particularly in circuitry where cognitive variables have been well investigated (e.g., in the direct and indirect pathways of the basal ganglia; Frank et al. 2004). Thus, these models serve primarily to provide causal explanations of observed neurophysiological data (Friston et al. 2003).

At a microscopic level, models of synaptic dynamics consider the subcellular milieu in which information is communicated (Jaeger and Bower 1999). These efforts are important to identify molecular targets for pharmacological interventions and, more importantly, can accommodate a detailed understanding of the effectors (Luscher et al. 2000) and dynamic function (Rubinov et al. 2009) of synaptic plasticity. With relation to psychiatric pathophysiology, models at the synaptic level have provided new insights into maladaptive plasticity in neuronal circuits. For example, components of glutamatergic homeostasis

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

have been used to explain a crucial nexus of circuit dysfunction in addiction (Kalivas et al. 2005). This has provided important translational insights in the context of new treatments for addiction. In this approach (Pendyam et al. 2009), the model was initially parameterized, based on a collation of quantities representative of anatomical and cellular physiological data acquired over years of experimental rodent work. Thereafter new disease data features were simulated by investigating the parameter space with the intention of identifying a limited set of parameters that were altered by chronic cocaine administration. The data for which the model was optimized was made up of measured microdialysis levels of extracellular glutamate. Critically, using a biophysical model of cocaine-induced cellular changes, neuronal plasticity, and larger network effects, it was possible to elucidate the influence of prefrontal inputs on the nucleus accumbens (Kalivas 2009). This example demonstrates a model-based identification of a putative target for disease treatment.

Above synaptic-level dynamics, single neuron models can be used to describe cellular input–output transformations whereas neural network models simplify the cellular processes and employ, for example, integrate and fire dynamics (Rudolph and Destexhe 2006) to represent a cell in an ensemble of connected neurons. These models can be used to describe the pathology of network connections, for instance, along the perforant pathway (from the entorhinal cortex to hippocampus), where they have been used to simulate the effects of NMDA receptor antagonism on memory impairment in schizophrenia (Siekmeier et al. 2007). At a scale above these models lie neural mass and mean field models (Deco et al. 2008). Mean field approaches engage the statistical properties (typically first- and second-order moments) to model the evolution of neuronal ensembles probabilistically. These meso- and macroscale dynamics are governed by the interaction of their statistical quantities (e.g., the mean and variance of membrane depolarizations within a cortical macrocolumn; Marreiros et al. 2009) and can be described by Fokker–Planck or path integral formulations (Knight et al. 2000). In terms of insights to pathophysiological processes, these population equations have most widely served as models for understanding seizure activity (Breakspear et al. 2006; Jirsa et al. 2014). More recently they have been proposed as component models in large connectomic analyses of neuropsychiatric disorders with the intention of developing a field of "pathoconnectomics" (Deco and Kringelbach 2014; see also (Horga et al. 2015). This class of model is also used in *dynamic causal modeling* and has been applied to a range of psychiatric disorders to test network hypotheses. For example, in the study of schizophrenia, DCMs of both fMRI and EEG data were used to study hierarchical brain connectivity associated with visual illusions, revealing a reduction of top-down effects on visual processing (Dima et al. 2009, 2010), effects mirrored by rodent electrophysiological DCMs in a ketamine model of psychosis (Moran et al. 2015). Below we highlight the potential of the DCM approach for deriving treatment predictions, in the specific context of schizophrenia.

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

Information-Processing Models

In contrast to biophysical models, which are designed primarily to answer the question of *how* the brain performs a particular operation, information-processing models ask *what* it is that the brain—its neural circuits, cells, and molecules—is computing. In other words, information-processing models are designed to uncover the neuronal computations that drive behavior and expose latent states, which can be used to characterize an individual patient's traits, such as how emotion affects valuation of immediate relative to delayed reward (Lempert et al. 2015). A more complete characterization of the distinction and overlap between biophysical and information-processing models is given by Kurth-Nelson et al. (this volume). Here we offer examples of where information-processing models have informed pathologies in the brain's algorithmic performance.

One area where computational modeling has already had a profound and sustained impact on understanding clinical phenomena is dementia. The syndrome of semantic dementia (SD), which is associated with a subset of neurodegenerative disorders as well as herpes encephalitis, involves a disruption of conceptual knowledge, including both knowledge concerning specific facts (e.g., Paris is the capital of France) and richer patterns of associative knowledge and inference (e.g., penguins are birds, consistent with their having wings and beaks, but also have the atypical characteristic in that they do not fly). SD patients show impairments in tasks which tap into such knowledge and yet other forms of memory, including episodic and working memory, are relatively spared. In a series of studies beginning in the 1990s, Timothy Rogers, Jay McClelland, and colleagues developed a computational account of SD, leveraging the tool of neural network modeling (Rogers et al. 1999; Rogers and McClelland 2008). Neural network models (also referred to as connectionist or deep learning models), are closely aligned with biophysical models and involve simple neuron-like units, which carry activation levels analogous to neuronal spike rates and connect to one another through idealized excitatory and inhibitory synapses. A key aspect of neural networks is that they are associated with well-developed learning algorithms which permit the strengths of the synapse-like connections in a network to be adjusted to allow the network to perform target tasks, producing desired output patterns in response to particular inputs (McClelland et al. 2010).

McClelland and Rogers (2003) have modeled semantic knowledge as involving associative relations among object properties. For example, a network might be trained to map from inputs representing *robin* and the relation *has* to the outputs *wings, beak,* and *feathers,* and from the inputs *goldfish has* to *fins* and *gills.* Following such training, if a new item *bluejay* is introduced and the network is trained to respond to *bluejay* and *has* with *beak,* the network is likely to infer that *bluejays* also have wings and feathers. Beyond reproducing such intuitive patterns of learning and inference, the Rogers–McClelland

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

model captures detailed patterns of behavioral data related to human category knowledge and conceptual development. More germane to the present topic, however, the model reproduces and explains detailed aspects of task performance in SD. When synaptic connections in the network are weakened or removed, simulating the effect of disease, the network model displays a degradation of conceptual and category knowledge that parallels the pattern of progressive memory loss seen in SD. Furthermore, detailed analysis of the conditions under which such impairments arise in the model have given rise to novel ideas, subsequently validated, about the anatomical locus of the critical lesion in SD—pointing to the importance of highly convergent multimodal inputs into a "hub" region, which leads to a resulting focus on the temporal pole as a candidate region (Hoffman and Ralph 2011; Irish et al. 2014).

A second class of information-processing models are prediction error-based models which have served as a basis for theoretic approaches to understanding learning, inference, and decision making (Botvinick et al. 2009; Friston 2009). Starting from Pavlov's conditioned stimulus reflexes, these models have evolved to where predicted future rewards are used to estimate the value of states and actions (Montague, this volume). Temporal difference reinforcement-learning models use errors in expected future reward to update expectations (Sutton and Barto 1998). The now iconic correlate of phasic activity in VTA dopamine firing levels with temporal difference updates (Schultz et al. 1997) offered a paradigm shift in terms of the discovery of formal equivalencies between computation and neuronal activity. Moreover, the discovery highlights a key point in computational approaches to understanding (patho)physiological mechanisms; namely, raw data can be difficult to understand without models. Pathophysiological consequences of dysregulated dopaminergic reward prediction errors have been used to explain behavioral observations of aberrant learning patterns in dopamine-associated disorders, for instance when medicated Parkinsonian patients exhibit an impairment in learning from negative predictions in the presence of high tonic levels of striatal dopamine (Frank 2006). Using model-based fMRI (O'Doherty et al. 2007), pharmacological studies, genetic associations, and PET studies, different positive and negative learning signals have been linked to D1 and D2 binding, respectively (e.g., Cox et al. 2015). This dopamine-reliant signaling has also been used to identify striatal dysfunction in human neuropsychiatric conditions, including attention-deficit/hyperactivity disorder, substance abuse, and schizophrenia (Whitton et al. 2015). Expansions of these habitual reinforcement-learning models have been used to unravel the role of serotonergic dysregulation in depression, where Markov chain-transition probabilities become value dependent (Dayan and Huys 2008), thereafter permitting a reconciliation of complementary serotoninergic processes in depression; that is, where its role in predicting aversive events (Paulus and Angela 2012) can lead to a bias toward optimistic prediction and, through altering stopping policy and pruning, of action options. These findings inform rumination, low mood, and perseverative thinking which pervade subjective descriptions of depression

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

(Dayan and Huys 2015). These approaches are closely linked to accounts where an agent comprises a dyadic structure of models representing both value and the causal structure of the environment. This allows for further goals and neural systems to be examined by our models and can be imbued with temporal hierarchical structure (see Botvinick and Weinstein 2014). Indeed, this "promiscuity of models" may be a useful metaphor for brain function and is expanded below in the dopamine "fruit fly," where we sketch a role for "model-based" versus "model-free" brain circuits in mediating addiction.

Bayesian inference has been used formally to instantiate neuronal codes such as predictive coding under the free energy principle (Friston et al. 2006). This account posits a particular neuronal machinery that is designed to perform probabilistic reasoning, whereby the brain models, learns, infers, and acts on its world so as to minimize precision-weighted prediction errors. The neurobiological circuits required for this type of predictive coding have been shown to recapitulate key anatomical features of canonical cortical microcircuits (Bastos et al. 2012). This account is appealing to computational psychiatry as it posits prediction error updating processes throughout cortex. Indeed, precision-weighted prediction errors have been found to be reflected by fMRI signals all over the brain, even for simple sensory tasks (Iglesias et al. 2013). Moreover, predictive coding makes testable predictions about neurobiological substrates of belief updating in cortex, based on precision-weighted prediction errors. For example, glutamatergic top-down connections are supposed to mediate predictions via NMDA receptors, prediction errors are believed to be signaled via both AMPA and NMDA receptors at bottom-up connections, and their precision weighting is thought to depend on postsynaptic gain control through neuromodulatory transmitters and local GABAergic mechanisms (Corlett et al. 2011; Adams et al. 2013). This framework has been applied recently to outline a "computational anatomy of psychosis" (Adams et al. 2013) and offers testable substrates of the misheld belief structures that pervade psychiatric symptomatology.

The Value of Generative Models for Building Clinical Application Prototypes

Developments in biophysical and information-processing models, together with advances in molecular, cellular, and systems neurobiology, are building the foundations of a basic science of computational psychiatry. They proffer deep mechanistic insights into mind–brain relationships, which might apply directly to psychiatric clinical practice, and offer an avenue to amalgamate detailed neurobiological accounts into clinically relevant process models. In other words, the models accommodate an important translational aim whereby they harness and apply findings from animal models of psychiatric illness to

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

build better, more detailed descriptions of algorithmic and circuit breakdowns. To produce methodologies for improving patient care, prototypical examples of the modeling framework should be prefaced by a simple question: Will any of these models help diagnose, enable prognosis, tailor treatment, or prevent psychiatric illness? Independent of "big-data" analytics and disease predictions, where clinical predictions rest on black-box statistical relations among descriptive data features, the computational psychiatry approach seeks a mechanistic understanding of how treatment can be improved. While it may take longer to develop our form of model for clinical practice compared to black-box counterparts, the mechanistic approach allows for interpreting a successful prediction, in terms of the underlying biology and/or computation, and helps identify targets for new treatment approaches.

Our rationale is that data generated from a probabilistic model is most efficiently described by the parameters that generated it, in the sense that those parameters capture everything there is to say about the data that is not noise. For instance, if one were to generate noisy behavioral output from a simulation of a learning model, then the most succinct and theoretically optimal description of that data would be in terms of the parameters that were used to generate the data in the first place. This means that a good model which describes a data set well can be used to condense the data set optimally (cf. generative embedding, described below), resulting in measures that most efficiently remove the noise.

The examples we develop below are designed to illustrate:

1. *the uncovering of disease mechanisms*, where a learning model applied to dopamine signaling is used to inform testable hypotheses related to the causes of addiction;
2. *treatment prediction*, in which a computationally derived nosology may help in subcategorizations and tailored therapies for patients with schizophrenia; and
3. *a monitoring system for chronic disease management*, where the goal of reaching a patient's "new normal" is instantiated in a dynamic model applied to markers of brain activity.

Based on the overall framework of this generative probabilistic model of disease, we now present distinct examples tailored to specific clinical problems and highlight how different types of data may be used to further inform these models.

Testing Computationally Informed Theories of Addiction

Addiction has been termed "a pathology of motivation and choice" (Kalivas and Volkow 2014), with dopamine implicated in its pathophysiology given the direct effects of drugs of abuse such as cocaine in blocking dopamine transport and enhancing striatal dopamine levels. Model-based accounts of dopamine's

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

role in learning have provided a detailed theory of how aberrant learning mechanisms could contribute to addiction. Here we examine how the temporal difference model of reinforcement learning (TDRL) (Schultz et al. 1997) offers precise computational counterparts (C) to specific neural substrates (B) (Figure 12.2). We examine how together these mechanisms inform particular aspects of addictive behaviors, specifically hypervaluation of drugs of abuse, while remaining relatively agnostic to other aspects of the illness, such as individual susceptibility to compulsive drug-seeking behavior. The fact that TDRL cannot computationally expose all of the signs, symptoms, and markers of addiction is viewed here as an opportunity for computational psychiatry and our modeling approach. In particular, it is clear that multiple neural systems, instantiating habitual, goal-based, and emotional control over behavior, play a role in this disease (Redish et al. 2008; Everitt and Robbins 2013). Thus in the future, our framework provides a formalism to consider, simulate, and test the precise interaction of these systems. Indeed, a recent review of potential decision-making vulnerabilities in addiction highlights ten system functions that could play a role in maladaptive choice. Rather than treating each vulnerability separately, the value added by a model such as that presented in Figure 12.1 lies in their joint consideration, accessible through the clarity of a mathematical description.

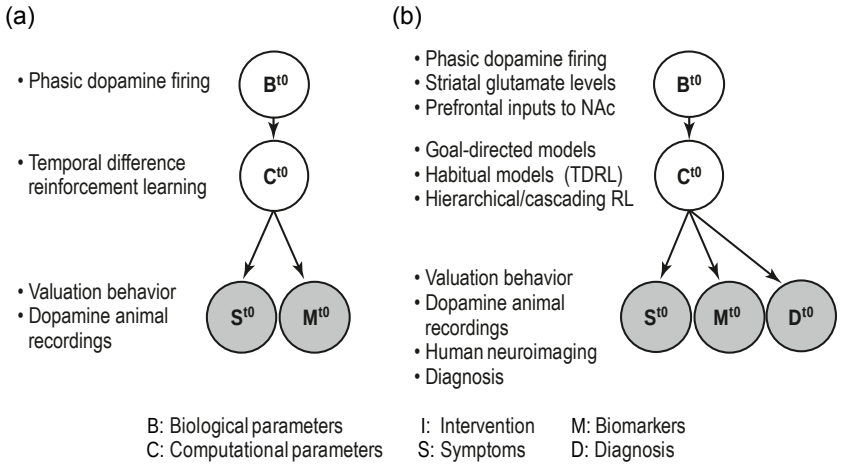

Figure 12.2 Developments in building a model for addiction. (a) The temporal difference model of reinforcement learning (TDRL) has formalized the role of phasic dopamine responses in evaluating states in the environment that predict reward and how the valuation process may be disrupted by pharmacologically enhanced dopamine. This model explains overvaluation in addiction and voltammetry findings in rodents. (b) Future accounts may use extended neurobiological and computational parameters to predict the full spectrum of symptoms associated with addiction and drug dependency.

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

The TDRL model (Sutton and Barto 1998) of phasic dopamine responses (Montague et al. 1996) goes beyond traditional reinforcement-learning and Pavlovian descriptions of stimulus-response associations, with two key components emerging from its particular mathematical formulation (Glimcher 2011). First, the formulation treats time as continuous discrete steps, removing the arbitrary notion of a "trial"; second, the formulation captures sequential associations, where an agent exposed to a series of conditioned stimuli, predicting the same unconditioned stimulus, can learn the history and redundancy of cues. These are both important constructs when considering addiction since real lives traverse through time not trials, and sequential operations may be bypassed by this system, given that reward predictions are transformed to the earliest patterns of predictive stimuli. Perhaps most importantly, the TDRL framework proffers a fundamental goal of a system instantiating its rules: it must carry zero prediction errors when a reward is encountered.

This mathematical goal has aided in explaining the apparent aberrancy in choices related to drug taking (Redish 2004), and in this context, addiction might be considered a valuation disease. In line with this model, prediction at synapses in the striatum is a potential neural substrate for instantiating the goal, which could be hijacked by an interacting pharmacological agent. Using a temporal difference model, the agent learns the value of each state in its world. Dopamine signals an error when things are better than expected, reduces its activity when outcomes are worse than expected, and stops modulating when rewards are predicted with zero error. However, imagine a case where reward is coupled with a pharmacologically magnified dopamine transient, via a drug: physiological assignment to predictive stimuli could reach ceiling levels and the agent would thus not reach its goal of zero error or learned state values. In other words, the values of the states which lead to drug receipt dwarf all other state trajectories, and maladaptive learning signals ensue. Based on this profound overvaluation, predictive environmental cues follow that enslave the agent to further drug-seeking behaviors.

This account has been used to explain several empirical features of addictive behavior and neurophysiology, including a decreasing elasticity to non-drug choice options in addicts over time, a resistance to blocking (learning associative redundancy) in animal models, and the concomitant dual dopamine signals observed in dopaminergic neurons in rodents (Redish 2004; Figure 12.2). As mentioned above, the model does not explain the whole range of phenomena encountered in addiction research, such as extinction (though the negative dip in firing is asymmetrically smaller than the phasic burst) and individual drug dependence. However, it stands as a canonical computational model system—a *computational fruit fly*—built directly on research which offered a model to explain neurophysiological firing patterns. Using this model as a basis, developments in formalizing additional decision systems—their parametric form and interactions (Servan-Schreiber et al. 1998)—will likely aid in building an expanded hypothesis set to uncover mechanisms of susceptibility,

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

chronic self-administration, relapse, and prevention. The idea of addiction as a shift from goal-directed to habitual networks has already been developed in a neurobiological circuit framework that includes shifting involvement from ventral to dorsal striatum, with a crucial role for prefrontal regions in drug reinstatement (Kalivas 2004; Everitt and Robbins 2005). This neurobiological extension (Figure 12.2) has been allied by theoretical developments which treat the formal adjudication between interacting "model-based" and "model-free" networks (Gläscher et al. 2010; Wunderlich et al. 2012b). The model also forms a basis from which hypotheses related to temporal extensions of a reinforcement-learning effect though hierarchical representation (Botvinick et al. 2009) and cascading corticostrital circuits (Collins and Frank 2013) could be studied (Figure 12.2).

Developing the neurobiological substrates of addiction may also refine model unknowns, such as the sort of ceiling effects that could be reached in overvaluing states. One crucial nexus is the role of the prefrontal cortex and its glutamatergic inputs for plasticity settings in the striatum. In other words, developing the latent biological parameters of the model will coincide with developing the latent computational parameters (Figure 12.2). In rodents, important clues regarding glutamate's role in striatal dysfunction has been elucidated using biophysical models. Pendyam et al. (2009) developed a structural and dynamic model of synaptic and extrasynaptic glutamate regulation to test *in silico* the effects of chronic cocaine administration and glutamate on neuroplasticity and withdrawal symptoms in the nucleus accumbens. Specifically, a differential-diffusion equation was used to understand the complex dynamics of extracellular glutamate levels. Diffusion properties were formalized by parameters, including diffusion coefficients given by the proposed local glial geometry ("tortuoisity"), while parameters of glutamate flux rates were controlled by synaptic and nonsynaptic glutamate release and exchange, reuptake transporters, and glutamatergic autoreceptors. By fixing model parameters to known empirical values—both physiological and cocaine-induced levels in glutamate exchange and autoreceptor signaling—different models of striatal geometry allowed the model to reproduce the basal reductions in extracellular glutamate observed in rodents after chronic cocaine administration. In addition to this withdrawal effect, the model predicted that enhanced extracellular glutamate levels which occur during rodent drug-seeking behavior could result from a specific change in the model's parameter space; namely, an alteration of the astrocytic XAG transporter. Thus the model predicted a molecular cause of synaptic overflow during drug-seeking behavior, providing a mechanism for reduced effectiveness in glutamatergic synaptic transmission in the nucleus accumbens. This susceptibility to extracellular glutamate accumulation has been proposed as a pathophysiological adaptation mechanism that could impair larger-scale corticostriatal communication (Kalivas 2009) and fits comfortably in an extended model-based framework of addiction phenomena. Couching these biophysical properties as computational effectors

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

could provide crucial links that enable a deeper understanding of the computational circuitry of addiction.

Dynamic Causal Modeling of Mismatch Negativity Circuitry for Treatment Prediction in Schizophrenia

As described by Huys (this volume), it would be intriguing to use computational models to predict which treatment should be assigned to an individual patient (e.g., in depression, psychotherapy vs. pharmacotherapy). Here we describe a potential (so far fictitious) concrete application in the domain of schizophrenia: predicting individual treatment response to a switch in pharmacotherapy. Clinically this is a highly relevant issue in the management of schizophrenia because at present there are no predictors to inform us as to which patient will respond to which drug. In clinical practice, antipsychotic drug treatment rests on trial and error processes: after several weeks of treatment, drugs (often chosen based on relevant side-effect profiles rather than predicted efficacy) are exchanged if no beneficial effect has been achieved.

Here, we consider a potential application of computational modeling to address this clinical prediction problem, using a concrete scenario. This potential application is guided by theories which highlight the pathophysiological role of NMDA receptors (Lisman et al. 2008; Gonzalez-Burgos and Lewis 2012) and, more specifically, their interaction with neuromodulatory transmitters (dopamine, acetylcholine, and serotonin) (Friston 1998; Stephan et al. 2006). This "dysconnection hypothesis" postulates that individual variability in clinical trajectories and treatment response results from individual variability in dysfunctional interactions between NMDA receptors and neuromodulators (Stephan et al. 2009a). This implies that a tool capable of inferring NMDA receptor function and its regulation by neuromodulatory effects within disease-relevant circuits should have predictive power with regard to outcome and treatment response.

Methodologically, the approach described below conforms to the notion of "generative embedding" (Brodersen et al. 2011). The general approach of "embedding" is at the heart of a computational rationale, where model parameters have been demonstrated to better capture and classify patients than raw data alone (Wiecki et al. 2015).This entails using a generative model of measured data to obtain subject-specific parameter estimates of mechanisms (with a physiological or computational interpretation) for use in unsupervised learning procedures (e.g., clustering) to detect mechanistically defined subgroups. One can then test, in a second step, whether the assignment of individual subjects to subgroups has prognostic value; that is, whether belonging to one subgroup or another predicts differential response to treatment. In the potential application described here, the idea is to use a DCM (a generative model of neuroimaging or electrophysiological; here, EEG responses) to infer

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

values of synaptic parameters, at a circuit-level of description, that are empirically known to be sensitive to changes in NMDA receptor and muscarinic receptor status.

Concretely, the DCM considered here concerns an auditory-prefrontal circuit (with bilateral primary and secondary auditory cortex, as well as right inferior frontal gyrus) known to be involved in mismatch negativity (MMN), an event-related potential in response to unexpected or surprising auditory events (for review, see Garrido et al. 2009). This particular combination of model and task is of special interest for three reasons:

1. MMN is significantly reduced in schizophrenic patients: a meta-analysis including more than thirty studies has indicated a robust effect at the group level (Umbricht and Krljes 2005).
2. Pharmacological studies in both animals and humans indicate that this reduction can be mimicked by administering antagonists of NMDA and cholinergic receptors (e.g., Javitt et al. 1996; Umbricht et al. 2000; Pekkonen et al. 2001; Schmidt et al. 2012).
3. Several previous studies have applied different DCMs to MMN data acquired under pharmacological manipulation and have demonstrated that appropriate physiological parameters of the DCMs are sensitive to selective pharmacological interventions. For example, parameters encoding the strength of glutamatergic connections from primary to secondary auditory cortex are sensitive to ketamine, an NMDA receptor antagonist (Schmidt et al. 2012). Furthermore, parameters controlling the postsynaptic gain of supragranular pyramidal cells in primary auditory cortex reflect the level of acetylcholine under manipulation by the acetylcholinesterase inhibitor galantamine (Moran et al. 2013).

In the proposed application, different types of DCMs could be used. The simplest variant would be a current-based neural mass model predicated on the formulation of Jansen and Rit (1995). This model has been used successfully in the ketamine DCM study of MMN (Schmidt et al. 2012) but offers a relatively limited representation of physiological mechanisms. A more sophisticated alternative is a conductance-based DCM, which represents a circuit of interacting cortical modules, each characterized by a mean-field model, where the neuronal state equations describe the change in average membrane potential as a function of conductance changes in ionotropic receptors (AMPA, NMDA, GABA) with sufficiently distinct time constants (Marreiros et al. 2009; Moran et al. 2011).

These equations of hidden neuronal dynamics can be coupled to an observation model which predicts sensor-level EEG measurements as a linear superposition of sources (Kiebel et al. 2006). Under Gaussian assumptions about the observation noise and Gaussian priors on the parameters, the model can be inverted using a variety of techniques (e.g., variational Bayes or Markov chain Monte Carlo), yielding posterior parameter estimates. In other words,

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

the model described here in brevity allows one to obtain probabilistic estimates, from conventional EEG measurements, of synaptic parameters within a circuit of interest, some of which have been previously found empirically to be sensitive to pharmacological perturbations of NMDA receptor function and acetylcholine levels (Schmidt et al. 2012; Moran et al. 2013).

In the proposed application, this model is applied to MMN data obtained from EEG measurements in schizophrenic patients who have not adequately responded to a first course treatment with the antipsychotic drug risperidone, and for whom the treating physician recommends a switch to another drug, olanzapine. We consider this particular constellation for two reasons. First, the sequence of treatments considered here represents a common clinical algorithm for schizophrenia treatment and has been investigated by other recent studies, which examined the success of a treatment switch from risperidone to olanzapine (e.g., Agid et al. 2013) and tried to predict this from initial clinical data (Kinon et al. 2010). Second, although both drugs possess antagonistic effects at various dopamine receptors, they differ strongly along the cholinergic dimension in that risperidone has no affinity to cholinergic receptors, while olanzapine is a strong muscarinic antagonist (PDSP K_i Database[1]). This means that any potential individual difference in treatment response might be attributable to individual differences in cholinergic function, which in turn may be detectable using model-based inference and used for predictions about treatment response.

Following EEG measurements and treatment switch, patients would require clinical follow-up examinations, with clinical assessment obtained at fixed intervals (e.g., two and eight weeks after treatment) according to the positive and negative syndrome scale (PANSS). These clinical symptom scores would represent the target of prediction by model parameters. From individual parameter estimates of the circuit described above, one would select parameters with empirically demonstrated sensitivity to pharmacological manipulations of NMDA receptors (e.g., parameters encoding the plasticity of glutamatergic connections from primary to secondary auditory cortex; Schmidt et al. 2012) and cholinergic receptors (e.g., parameters representing the postsynaptic gain of pyramidal cells in primary auditory cortex; Moran et al. 2013). Thereafter, one could test whether subject-specific parameter estimates predict clinical symptom scores following treatment switch.[2] Evaluating this putative predictive power could proceed in at least two ways. First, the individual parameter estimates of interest could serve as features for unsupervised learning (clustering), with the goal

1 http://kidbdev.med.unc.edu/databases/kidb.php (accessed July 10, 2016).

2 As a caveat, Brodersen et al. (2011) used galantamine (an acetylcholineesterase inhibitor) which also has allosteric action at nicotinic receptors. Thus it is not clear to what degree the empirically demonstrated sensitivity of DCM parameters to galantamine partitions into muscarinic and nicotinic effects. Generally, however, previous studies in humans (Pekkonen et al. 2001) and unpublished data from rats (based on selective muscarinic receptor manipulations) demonstrate sensitivity of the MMN to muscarinic receptor alterations.

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

of detecting patient subgroups delineated by differences in the parameters of interest (see Brodersen et al. 2014). Under this perspective, one could then try to validate the proposed subgroups by testing for significant differences in response to a treatment switch across patient subgroups. Alternatively, one could directly predict the change in symptom scores as a function of model parameter estimates, using conventional multiple linear regression. If successful, the former option would provide the clinician with a tool that would enable assignment of individual patients to a particular subgroup, and hence predict treatment response in a categorical fashion. The latter option, by constrast, would enable the clinician to predict the change in (continuous) symptom scores for an individual patient, following a treatment switch.

The above (so far hypothetical) application of a computational model to individual patient data represents an example of how a relevant clinical question could be addressed through existing modeling frameworks that are supported by pharmacological validation studies in humans and animals. We have spelled out this fictitious case study in some detail to showcase the motivation, potential, and limitations of a computational psychiatry approach.

An Allostatic Recovery Model

Recovery models in psychiatry and mental health refer to the personal restabilization of a patient's life through their participation in treatment, the development of coping strategies, and renewal of a sense of self. In other words, this process guides the patient to return to a new, nonharmful "normal" state of being (Ramon et al. 2007).

Selfhood has physiological representations through internal bodily states (Critchley and Seth 2012) that are sensed by the brain in the insular cortex (Simmons et al. 2004; Gu et al. 2013; Kirk et al. 2014). Dysregulation of this process is associated with anxiety disorders and depression (for a review, see (Paulus and Stein 2010). Mounting computational literature (Seth et al. 2011) casts these dysregulated interoceptive signals as errors of prediction, whereby "individuals who are prone to anxiety show an altered interoceptive prediction signal, i.e., manifest augmented detection of the difference between the observed and expected body state" (Paulus and Stein 2006:383). In this novel example, sketched purely for this chapter, we develop the idea of dysregulated stress responses in the hypothalamic-pituitary-adrenal (HPA) axis and its potential behavioral control through amygdala–hypothalamic projections. Specifically, given the ubiquity of serotonergic drugs in treating depression, we aim to model the dynamic balance of a positive feedback loop between excitatory serotonergic effects on HPA mediated by the amygdala (Weidenfeld et al. 2005), amygdala → HPA, and the amygdala's excitability dependency from corticosertone in the HPA (Stutzmann et al. 1998), HPA → amygdala. We aim to illustrate how a simple model could capture this network and, in

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

doing so, quantify a biological marker of recovery where the patient's goal is to cycle through amygdala and HPA activity under stress. This example thus serves as a conceptual outline to a computational recovery model of anxiety and depression.

To build a generative model of this process, we utilize dynamical systems theory, employing a simple description which has state variables that represent the dynamic processes of homeostasis and allostasis (Figure 12.3). In this setting, we refer to the general definition of allostasis: "stability through change" (Sterling 2014), specifically stabilization to a new state that is not harmful, but is also not the "normal" state that patients were used to prior to their illness.

Homeostasis is a process by which system perturbations are reactively damped, ensuring return to a stable equilibrium. Allostasis is the process of reaching equilibrium through change and offers anticipatory control over homeostatic mechanisms (Schulkin 2010). Whereas homeostatic mechanisms have a fixed-point equilibrium, allostatic processes can yield more complex dynamics, including limit cycles and even chaos (Rodrigues et al. 2007) through, for example, positive feedback loops (Spiga et al. 2008). The dynamics of allostatic control over homeostatic dynamics begins in our model with a single state variable, x, a signal that returns to homeostasis through a mean reverting Ornstein–Uhlenbeck process (i.e., a nondelayed self-correcting random walk). In turn, this process is modulated by a second signal, y, which embodies a prediction of x. The introduction of a (negative) time lag between

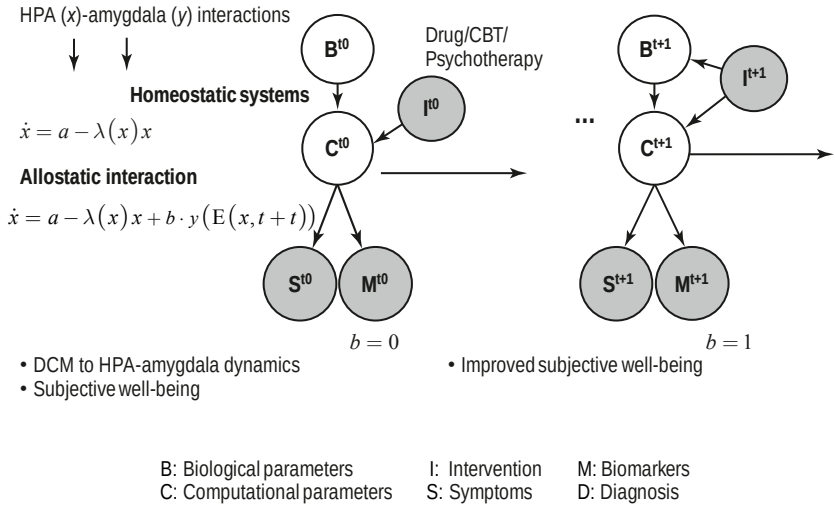

Figure 12.3 A dynamical systems model of allostatic equilibrium in the HPA-amygdala axis to test the effectiveness of mental health recovery. This app is dedicated to the memory of Xavier, the random walking spider.

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

prediction, measurement, and action corresponds to the extension from homeostasis to allostasis and allows for more complex behaviors. The goal of our hypothetical treatment intervention is to ensure that the homeostatic signal is receiving allostatic system inputs and that allostatic systems are exhibiting complex dynamic behaviors, such as periodicity and even multistability, hence achieving stability through change.

To make this more concrete, let us consider the case of depression. In this disorder, stress can act as a precipitating factor leading to elevated cortisol levels and a disruption of its rhythmicity (Johnson et al. 2006). This has been hypothesized to result in abnormal homeostasis in HPA and could arise from reduced excitatory amygdala inputs (Herman and Cullinan 1997). In our model, this downregulation of allostatic drive will break its rhythmicity without sufficient predictions and render the patient enslaved to homeostatic effects, as reflected in hyperactivity of the HPA axis (Pace et al. 2006). The model is designed to monitor the return of a patient's HPA axis from homeostatic reaction to allostatic control. The types of data which could be used to monitor these dynamics include metabolic and neurophysiological assays together with neuroimaging time series where connectivity assessments (e.g., DCM) could potentially be used to monitor rebalance. The advantage of this approach is its ability to quantify directly the return to allostasis. It would thus provide a metric for suitable change in a recovery model, which has received some criticism for its lack of an evidence base (Davidson et al. 2005).

Developing Community-Wide Standards for Model Development

Impacting standard clinical practice will require committed efforts from stakeholders, including psychiatrists and allied mental healthcare professionals as well as patients, insurance companies, and policy makers. The goal is to enable the type of computational prototypes developed here to be interactive, so that clinicians can access and probe precise simulations to better predict patient outcomes. For these systems to be realized, communities must adopt standards to ensure consistent reporting in terms of patient tests, specific models, and measurements. The important consideration is how, in practice, to best advance the agenda of applying computational approaches to psychiatry, so as to maximize the probability of the field having a tangible impact on psychiatric theory and practice. For this to happen, close collaborations must be forged between clinicians, theoreticians, and neurobiologists. Each area of expertise brings something essential to the table. For example, a theoretician may be able to build an elegant model of a computational process that may be altered in a particular disease or symptom. However, if such models are based on idealized characterizations of a disorder and fail to make contact with the real complexity and/or heterogeneity of a disease, then the model will most likely not have predictive or even face validity, and is therefore unlikely to be useful

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

to the field as a whole. Conversely, to interpret findings about the meaning of specific effects observable at the molecular, cellular, circuit, or systems level in the brain, and to translate those findings effectively to psychiatry, neurobiologists need to interact closely with theoreticians and psychiatrists. One important overall strategic objective for advocates of this field will be to find ways to promote such close collaborations.

How can this be achieved? One obvious avenue is to bring psychiatrists, theoretical neuroscientists, and neurobiologists together in regular conferences, summer schools, and workshops focused on computational psychiatry. This would help facilitate a common language and provide a fertile environment for new collaborations to form. Because of the difficulty in understanding each other's terminology, such meetings should take the form of dialogue meetings, in the spirit of the Strüngmann Forum, and utilize a structured approach to address specific topics in computational psychiatry. Furthermore, the development of interdisciplinary training programs would serve a critical role in training the next generation of clinicians and basic scientists. Their goal should be (a) to train psychiatrists in computational techniques so that they are familiar with the computational language without necessarily becoming theoreticians in their own right and (b) to introduce neurobiologists and theoretical neuroscientists to the complexity inherent in the diagnosis and treatment in psychiatry. Efforts to facilitate the emergence of genuine hybrids who are both theoretical neuroscientists and practicing clinicians (psychiatrists or clinical psychologists) could have a major impact on the future of the field. Finally, we need to institutionalize computational psychiatry by establishing units where scientists with computational and biomedical backgrounds can work together, literally under the same roof and ideally in shared offices. Physical coexistence is perhaps one of the most powerful ways to facilitate the creation of a common language and mutual understanding.

Other practical considerations include working to convince funding agencies throughout the world as to the potential impact of the field, with the goal of encouraging these agencies to set up specific funding programs or mechanisms tailored to computational psychiatry. This could happen, for instance, by explicitly requiring collaborations between theoreticians, neurobiologists, and psychiatrists as an eligibility criterion in applications for a particular funding mechanism focusing on topics in computational psychiatry.

Finally, we need to consider how the field as a whole can work to facilitate progress in research in this domain. One possible avenue is to develop frameworks in support of data sharing, or the development of large-scale collaborations so that we can reach a critical mass in terms of number of participants, diversity of theoretical constructs, tasks and measures to test computational hypotheses in psychiatric populations, in a manner that ensures sufficient statistical power and robustness to variation across psychiatric populations.

One practical suggestion for how to go forward with all of these proposals is to convene a committee charged with developing strategies for making

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

progress in each of these domains over the next several years. Below, we outline a major component of community-wide models; namely, the stimulus paradigms from which we elicit our signals and to which we apply our models. Thereafter we consider existing efforts, highlight the CNTRICS battery developed for schizophrenia researchers, and conclude by suggesting a new international consortium for resource and knowledge sharing.

Standard Tasks and Paradigms

Developing standardized experimental paradigms that provide reliable assessments of cognitive function, such as working memory or decision making, will be crucial to get data for precise modeling results. The downside of these general assessments may be twofold: First, they are potentially ignorant about the current emotional state or symptom expression of the individual patient. For example, specific aspects of decision making in an individual patient with addiction can be fully intact in multiple domains and yet be highly impaired in the context of drug-associated environments. Paradigms that fail to elicit relevant contextual states may suffer from small effect sizes and yield uninformative experimental outcomes. Second, using fixed stimulus sequences might not account for the heterogeneity across individuals, particularly in patient populations. To address these issues, experimental setups could be tailored to individual patients by probing state- rather than trait-dependent aspects of cognitive function and/or adjusting data acquisition online to optimize the feasibility of model inversion (parameter estimation) and model comparison.

One concrete example of state-specific paradigms is symptom-provoking stimuli in obsessive-compulsive disorder, where patients are exposed to stimuli that relate to their individually expressed obsessions and compulsions (Adler et al. 2000). Emotional stimuli (movies) or subject-specific biographical events are similarly effective tools in the context of mood disorders. In addition to choosing stimuli with higher face validity, the design of the experiment itself can be optimized in further aspects. Game-inspired paradigms can provide more naturalistic environments to provoke domain-specific behavioral patterns: the use of slot machines to probe decision making in a gambling context (Clark 2010) or virtual environments to simulate real-world interactions (Parsons and Rizzo 2008).

Optimized data acquisition can also be accomplished through adaptive experimental designs, whereby stimulus presentation is dynamically adjusted throughout the experiment. This may involve online adjustment of stimulus presentation, based on the past history of individual responses, to yield optimal data for parameter estimation and model comparison. A simple example of this type of adaptive data acquisition is the staircase paradigm for delivering stimuli according to individual (and potentially time-varying), perceptual, and performance-related properties (e.g., perceptual thresholds and task accuracy). More sophisticated approaches derive from probability theory and

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

would enable an optimal experimental design to be found for testing a specific hypothesis embodied by a particular model's structure. This could be done *a priori* (before the experiment) or online (Daunizeau et al. 2011b).

In summary, individually tailored paradigm designs have the potential of markedly improving model-based inference by targeting relevant state-dependent behavior and optimizing data acquisition in particularly heterogeneous patient groups. This would result in increased effect sizes and more sensitive statistical tests.

Example Task Battery: The Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia (CNTRICS) Initiative

During the 1990s, a growing awareness of the disabling nature and treatment refractoriness of cognitive impairment in schizophrenia highlighted the need to develop new treatments for this aspect of the illness. In developing a pathway to drug registration, a set of tools was developed with the support of the U.S. National Institute of Mental Health (NIMH). One initiative, the Measurement and Treatment Research to Improve Cognition in Schizophrenia or MATRICS, developed a battery of tests that consisted primarily of clinical neuropsychological tests already in use in drug development trials. During this process, it was proposed that experimental measures from cognitive neuroscience, as opposed to these clinical tests, would offer the advantage of targeting more specific cognitive systems that were linked more directly to discrete neural systems. Concerns included that there was no general consensus in the field as to what cognitive systems should be targeted, no standard, easy to administer tasks to be used for measurement, and no information about the psychometric properties (reliability, presence of ceiling, and floor effects). To address these concerns, and propel the field toward a neuroscience-based approach to measuring cognition and the impact of treatment on cognition in schizophrenia, the CNTRICS Initiative was launched in 2007, supported by an R13 conference grant from NIMH and led by Cameron Carter and Deanna Barch.

In all, seven conferences were held over a period of four years at a variety of locations across the United States. Each was informed by pre-meeting surveys of the larger field and brought together an international group of basic cognitive neuroscientists, clinical researchers, and those involved in treatment development, using a semi-structured consensus-based process. The initial three meetings developed a set of theoretical cognitive domains to be targeted and a set of experimental cognitive tasks with strong construct validity as measures of these domains. In all, twenty-three tasks across seven domains were recommended for development. The next four meetings focused on developing imaging and ERP biomarkers with strong construct validity for measuring the cognitive and neural systems associated with each domain, in addition to two meetings which focused on the development of more integrated animal model systems for use in the drug discovery process.

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

At about the midway through the CNTRICS process, funding was obtained to begin developing tasks (selected because of their construct validity, clinical importance, and other factors) into tools that could be used for standardized measurement of cognition in clinical and treatment research. This new project—Cognitive Neuroscience Test Reliability And Clinical Applications for Schizophrenia, or CNTRACS—is currently ongoing and involves five sites across the United States. Tasks have been adapted to ensure that specific deficits in cognitive mechanisms could be measured independently of generalized deficits (e.g., attention lapses, poor motivation), optimized (on factors such as numbers of trials, length of administration and maximizing effect sizes), and studied for their psychometric properties and relationship to symptoms and to measures of functioning. The first round of data collection has been completed, including a supplementary study using fMRI for three of the four tasks initially studied, and a number of publications have resulted. Importantly, brief, well-tolerated versions of measures of cognitive control (AX-CPT), episodic memory (Relational and Item Specific Encoding, or RISE task), perceptual integration (Jitter Orientation Visual Integration Task, or JOVI task), and early visual perception (Contrast-Contrast Effect, or CCE task) have been developed with acceptable test-retest reliability and predictable relationships with different sets of symptoms and functioning in the patients with schizophrenia. A number of the theoretical constructs and recommendations were subsequently incorporated into the Research Domain Criteria (RDoC) framework by leaders at NIMH. The presentations from each of the CNTRICS meetings, along with the papers documenting the results of each meeting are available on line (cntrics.ucdavis.edu). In addition, publications resulting from the cognitive neuroscience test reliability and clinical applications for schizophrenia (CNTRACS) consortium and scripts for running the four tasks studied in the initial round of this project are programmed in Eprime and available for free download from the site.

Despite this valuable activity and our optimism toward it, we end with a note of caution: The field of computational psychiatry is still very much in its infancy and the problem it aims to address is immense. Therefore, we need to take a long-term view and exercise patience, for progress may proceed inconsistently and irregularly. As with genetics, the original promise inherent in the field of computational psychiatry may take decades to be fully realized.

From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.