CellPress

Neuron
Perspective

Computational Psychiatry

Xiao-Jing Wang[1,2,3,*] and John H. Krystal[3,4,5,6]
[1]NYU-ECNU Institute of Brain and Cognitive Science, NYU-Shanghai, Shanghai, China
[2]Center for Neural Science, New York University, 4 Washington Place, New York, NY 10003, USA
[3]Department of Neurobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA
[4]Department of Psychiatry, Yale University School of Medicine, 300 George Street, Suite #901, New Haven, CT 06520, USA
[5]Psychiatry Service, Yale-New Haven Hospital, New Haven, CT 06510, USA
[6]Clinical Neuroscience Division, VA National Center for PTSD, VA Connecticut Healthcare System, West Haven, CT 06516, USA
*Correspondence: xjwang@nyu.edu
http://dx.doi.org/10.1016/j.neuron.2014.10.018

Psychiatric disorders such as autism and schizophrenia, arise from abnormalities in brain systems that underlie cognitive, emotional, and social functions. The brain is enormously complex and its abundant feedback loops on multiple scales preclude intuitive explication of circuit functions. In close interplay with experiments, theory and computational modeling are essential for understanding how, precisely, neural circuits generate flexible behaviors and their impairments give rise to psychiatric symptoms. This Perspective highlights recent progress in applying computational neuroscience to the study of mental disorders. We outline basic approaches, including identification of core deficits that cut across disease categories, biologically realistic modeling bridging cellular and synaptic mechanisms with behavior, and model-aided diagnosis. The need for new research strategies in psychiatry is urgent. Computational psychiatry potentially provides powerful tools for elucidating pathophysiology that may inform both diagnosis and treatment. To achieve this promise will require investment in cross-disciplinary training and research in this nascent field.

Introduction

In 1988, a computational neuroscience “manifesto” (Sejnowski et al., 1988) mentioned three reasons for the emergence of this new research field: advances in neuroscience had generated a large body of neurophysiologic data, new computers possessed sufficient power to conduct neural model simulations, and simplified brain models were introduced that provided insights into complex neural circuit functions. Since then, dramatic advances made on all three fronts fundamentally changed the computational neuroscience landscape (Abbott, 2008). Notably, computational neuroscience initially focused on the early stages of sensory processing (Sejnowski et al., 1988), because studies of the neural bases of higher cognitive functions were beyond empirical neuroscience of that era. Indeed, only in recent years has the confluence of single-unit physiology, human functional brain imaging, and advances in computational modeling made significant strides in tackling executive functions (such as working memory and decision making) that underlie cognitively controlled flexible behavior. These higher functions critically depend on the prefrontal cortex (PFC) (Fuster, 2008; Miller and Cohen, 2001; Wang, 2013; Szczepanski and Knight, 2014). Because impairments of the PFC and related circuits are implicated in major psychiatric disorders, such as schizophrenia and autism (Goldman-Rakic, 1994; Insel, 2010; Courchesne et al., 2011; Anticevic et al., 2013a), the newly acquired insights and computational models offer an opportunity to elucidate how cellular and circuit level pathologies give rise to cognitive deficits observed in mental illness, advances in this direction could inform studies of psychiatric diagnosis, pathophysiology and treatment.

Therefore, the time is ripe for computational psychiatry to emerge as a field at the interface between basic and clinical neuroscience (Montague et al., 2012; Friston et al., 2014). In this Perspective, we review recent work demonstrating that computational psychiatry introduces novel approaches and tools to investigate neural circuit mechanisms underlying the cognitive and behavioral features of neuropsychiatric disorders. First, we will spell out the rationale of a computational approach to psychiatry, i.e., “why computational psychiatry? What theories and models are relevant to this field?” Second, we will discuss how theories and models have been applied to the investigation of behavioral impairments in terms of transdiagnostic endophenotypes. Third, we will summarize recent work that advocates for a model-aided framework of diagnosis and treatment. The fourth part will be devoted to biophysically based neural circuit modeling that we argue represents the optimal approach for cross-level understanding from cellular processes to collective and emergent circuit dynamics and ultimately to behavior. Fifth and finally, we will end with practical recommendations related to the training and funding needed to foster this nascent field.

Why Computational Psychiatry?

It is widely acknowledged that current psychiatric diagnostic schema and the treatments for psychiatric disorders lack a firm biological foundation. The complexity of the brain presents unique challenges to the development of highly specific mechanistic hypotheses to guide research in psychiatry. Advances in genetics, and molecular and cellular neurosciences are providing, at long last, clues to the etiology of human cognitive, emotional, and behavioral problems. For example, candidate-gene studies have revealed gene variations (such as DISC1; Brandon et al., 2009) associated with psychiatric disorders. However, many in the field think that attempts to seek single genes underlying complex psychiatric phenotypes have been largely disappointing, and that efforts to link genes to more basic cognitive and behavioral functions and functional impairments could be more promising. The progress in these areas has yet to provide a firm basis for a diagnostic system or a single pharmacotherapy for common psychiatric disorders (Krystal and State, 2014).

A major hindrance in our capacity to develop novel pharmacotherapies for psychiatric disorders is the still superficial nature of our understanding of how circuits produce behavior. In this regard, synaptic and systems physiology are producing remarkable advances in our specific understanding of the functional properties of microcircuits and the beginnings of connecting these insights into behavioral processes including basic visual perception (Parker and Newsome, 1998), fear conditioning and extinction (Johansen et al., 2011), and mental representations in working memory (Arnsten et al., 2010). There are even examples where aspects of the neural representation of distinct fear memories can be ascribed to the functional integrity of a few distinct sets of cells in the amygdala (Josselyn, 2010). Yet, perhaps as a consequence of the limitations of our animal models combined with the limited spatial and temporal resolution of current neuroimaging technologies (MRI, magnetoencephalograpy, positron emission tomography), there is not a single symptom of a single psychiatric disorder for which we fully understand its physiologic basis at a molecular, cellular, and microcircuit level. In other words, we have only a somewhat vague idea of how the brain generates the cognitive, emotional, and behavioral problems that lead people to seek treatment by psychiatrists and other mental health clinicians.

As a consequence of our limited understanding of how circuits represent information, there are a plethora of attempts to explain circuit dysfunction in psychiatric disorders in superficial ways, giving rise to an equally large number of relatively risky potential pharmacologic strategies to address the unmet need for more effective treatments. The implications of this knowledge gap are profound for the field of psychiatry and for society. For example, psychiatric diagnoses have categorical qualities as exemplified by the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (DSM-5). Although this new version of the DSM takes into consideration the recent explosions in the genetics of disorders, such as autism and schizophrenia (Krystal and State, 2014), it is widely criticized for lack of a solid biological foundation based on either etiology or pathophysiology. Categorizing patients by symptom checklists results in enormous clinical heterogeneity within diagnostic categories, surprisingly poor interrater reliability for many common psychiatric diagnoses (Freedman et al., 2013), and very likely, poorer clinical outcomes.

An alternative schema has emerged from the recognition that behavioral impairments are traits that may be shared across psychiatric disorders (Krueger, 1999). The shift from a categorical diagnostic focus to a dimensional transdiagnostic approach emerged in the form of the Research Domain Criteria (RDoC, http://www.nimh.nih.gov/research-priorities/rdoc/index.shtml) (Insel et al., 2010; Insel 2014). The RDoC program aims at identifying core cognitive, emotional, and social dysfunctions, then elucidating their brain mechanisms bridging different levels (from molecules, cells, circuits to functions). Yet, the next step in this process is to determine whether the circuits are dysfunctional in the same way across disorders or whether, when characterized in increasingly accurate molecular and physiological ways, categorical features of psychiatric diagnoses reemerge. Furthermore, diagnoses may have both categorical and dimensional features. For example, schizophrenia appears to be a more severe form of circuit dysfunction than bipolar disorder with respect to the thalamo-cortical functional connectivity (Anticevic et al., 2013b), but a completely distinct type of disorder than bipolar disorder with respect to the variance or ''noise'' level of cortical activity (Yang et al., 2014). Neither DSM nor RDoC in its current form provides guidance as to how to integrate the dimensional and categorical features of psychiatric pathophysiology. A second consequence is the lack of precision with which one can predict whether a particular treatment mechanism will work for psychiatric disorders. It is not just that biomarkers of illness are lacking, but rather the biomarkers that we have are not sufficiently mechanistically precise as to specify a particular treatment. In addition, even when aspects of molecular pathology are characterized, the impact on micro-and macrocircuit functions and the paths to correct that circuit dysfunction are not clear. As a result, in the case of schizophrenia, it is not clear that GABA signaling deficits (Lewis et al., 2005, Lewis and Gonzalez-Burgos, 2006) should be treated by GABA$_A$ receptor agonists nor deficits in NMDA receptor (NMDAR) signaling should be treated with drugs that increase the stimulation of the glycine coagonist site of the NMDAR (Buchanan et al., 2011; Goff, 2014).

The gap between genetic, molecular, and cellular studies, on the one hand, and systems and behavioral neuroscience studies, on the other, currently cannot be bridged purely through experimentation. Take, again, the example of the PFC. Its crucial role in a wide range of executive functions (Fuster, 2008; Miller and Cohen, 2001; Wang, 2013) begs the question: what are the key properties that enable the PFC to subserve cognitive processes, in contrast to primary sensory or motor systems? This question is difficult to address by laboratory experiments alone, partly because PFC circuitry is endowed with powerful positive and negative feedback loops and the behavior of any such dynamical system is not predictable by intuition alone. While physiological studies in animals and humans yield data on the correlation of particular measurements to specific cognitive operations, theory and modeling are usually needed, together with experimentation, to investigate the ''follow-up'' questions: what circuit mechanisms give rise to the observed neuronal and other brain signals? What are the computational algorithms and generalizable principles that are reflected in the observed biological signals and sufficient to explain behavior?

Computational modeling offers a suitable approach to quantitatively explore the properties of complex systems across levels of investigation. Therefore, by incorporating computational neuroscience modeling within translational neuroscience research programs, it may be possible to develop more specific hypotheses related to circuit dysfunction in model systems and psychiatric disorders. There are many forms of computational models; we will present two types. Models of Mathematical Psychology or algorithmic models from Computer Science are enormously useful for quantifying behavioral data and relating their fitted parameters to neural computations (Maia and Frank, 2011; Montague et al., 2012). On the other hand, biophysically informed computational modeling, that are constrained by the

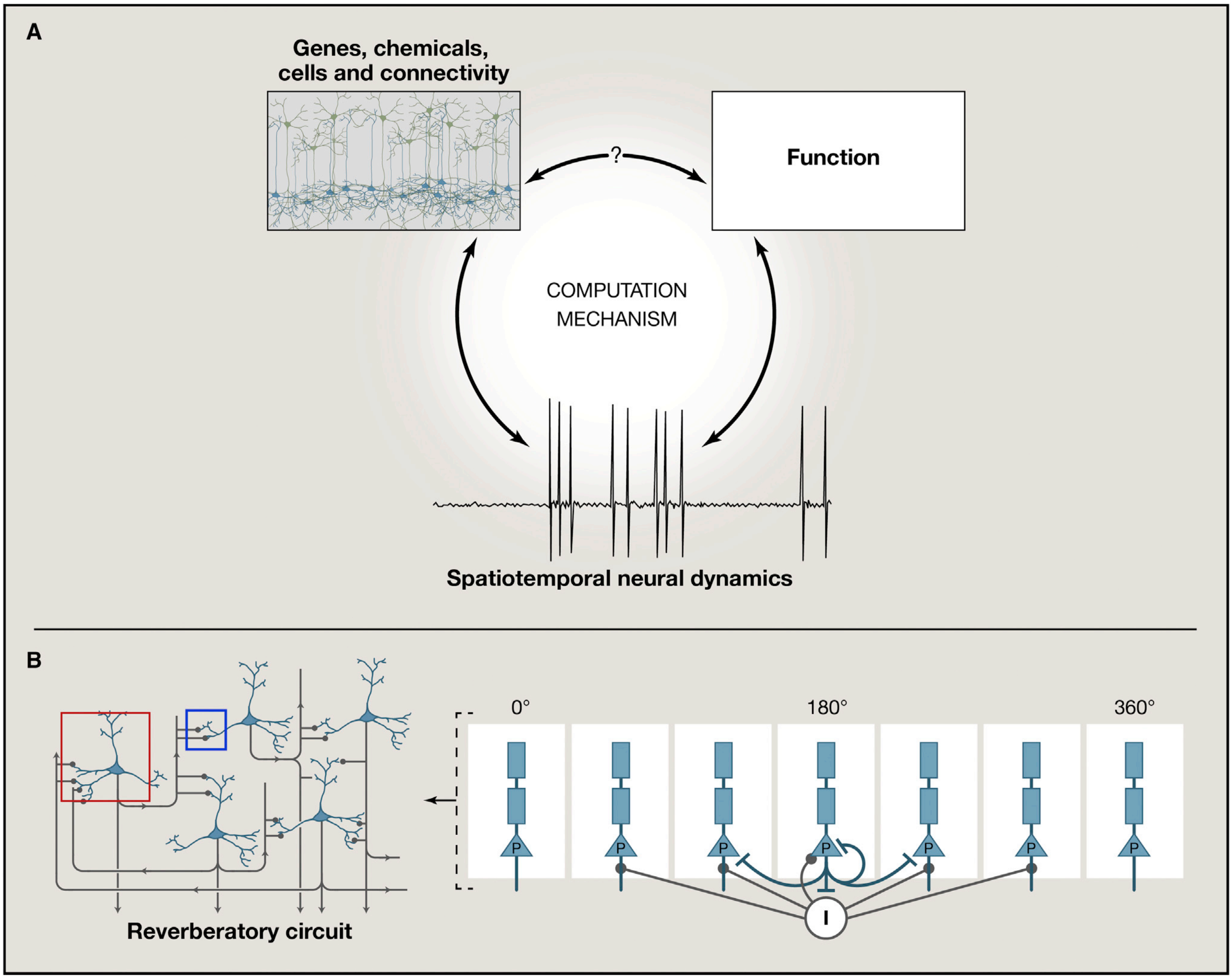
A
Genes, chemicals, cells and connectivity
?
Function
COMPUTATION
MECHANISM
Spatiotemporal neural dynamics
B
Reverberatory circuit
0°
180°
360°
P
P
P
P
P
P
P
I

Figure 1. Mechanistic Understanding of Brain Functions Must Relate Structure and Dynamics with Behavior

(A) Brain measures probe spatiotemporal neural activity patterns that are correlated with specific aspects of behavior. Theory and modeling provide a powerful tool to elucidate how such a pattern is produced by its biological substrate, on one hand, and give rise to computations necessary to account for brain function, on the other hand.

(B) Biologically based neural circuit modeling is calibrated by physiology of single neurons and synapses (red, blue box in the left panel, respectively), and constrained by quantitative network connectivity data. This approach is arguably necessary for the three-way understanding among function, neural dynamics and computation, and biological mechanism.

biophysical properties of identified synaptic signaling mechanisms and other properties of microcircuits, has proven to be an effective approach to understanding the neurobiology underlying cortical functions and psychiatric disorders (Wang, 2006; Anticevic et al., 2013a).

Biologically Based Neural Circuit Models

What is biologically based neural circuit modeling? Simply put, it is a computational framework that is constrained by neurobiology and designed to achieve a cross-level understanding of brain functions in terms of neural dynamics, computation, and biological mechanisms (Figure 1). One may question whether such models are too complex to be useful in cognitive science or psychiatry (Carandini, 2012; Montague et al., 2012). Three points are worth noting on this regard. First, biologically based modeling is a broad term that embraces a diversity of models with varying degrees of complexity. A model does not necessarily improve when more biological details are included. There is always a tradeoff between incorporating important details in order for the model to be suitable (given a scientific question) on one hand and simplicity and generalizability on the other hand. It is also tremendously useful to be able to go back and forth between models differing in their levels of abstraction, for instance between a spiking network model and its reduced "mean-field" firing-rate model for population-level dynamics. Second, neuronal modeling is most appropriate for those functions for which we have some knowledge about the underlying neural processes, such as dopamine neural signaling of reward-prediction error, persistent activity subserving the internal representation of working memory, and neural integrators in perceptual decision-making. In contrast, modeling at the neuronal level would seem premature for other behavioral phenomena such as hallucinations, in the absence of neurophysiological characterization. Finally, to the extent that biophysically based neural circuit modeling begins by incorporating the simplest and most fundamental features of synaptic connectivity, it is arguably the simplest possible framework that permits us to elucidate the interrelationship among biological mechanism, neural dynamics and computations, and circuit functional output.

In a spiking network model, single neurons are often described by either the leaky integrate-and-fire model or the Hodgkin-Huxley model. These models are calibrated by physiological measurements, such as the membrane time constant and the input-output function (the spike firing rate as a function of the synaptic input), which can be different for excitatory pyramidal cells and inhibitory interneurons. Furthermore, it is worth emphasizing that in biophysically based models, synapses must be modeled accurately. Unlike connectionist models in which coupling between neurons is typically an instantaneous function of firing activity, synapses have their own rise-times and decay time constants, and they exhibit summation properties. Synaptic dynamics are crucial factors in determining the integration time of a neural circuit and the stability of a strongly recurrent network (Wang, 1999). Finally, networks endowed with a biologically plausible architecture need to be constructed based on quantitative anatomy (Douglas and Martin, 2004). For example, a commonly assumed circuit organization is local excitation between neurons of similar selectivity combined with a more global inhibition. Dynamic balance between synaptic excitation and inhibition is another feature of cortical microcircuits that has been increasingly recognized experimentally and incorporated in cortical network models (http://www.scholarpedia.org/article/Balance_of_excitation_and_inhibition).

Consider decision making, the process of reaching a particular choice among several alternative options, such as rendering a judgment out of multiple possibilities given incomplete information or choosing one of actions expected to yield different outcomes (Glimcher, 2003; Gold and Shadlen, 2007; Wang, 2008; Glimcher and Fehr, 2013). Broadly speaking, there are two types of computational models of decision making: behavioral models and neural circuit models. In behavioral psychology, decision making is commonly modeled by the drift diffusion model (Ratcliff, 1978; Smith and Ratcliff, 2004). In this model, an activity variable X represents the difference between the respective amounts of accumulated information about the two alternatives, say X_A and X_B, $X = X_A - X_B$. The dynamics of X is given by the drift diffusion equation, $dX/dt = \mu + w(t)$, where μ is the drift rate, w(t) represents noise. The drift rate μ represents the bias (net difference in the evidence) in favor of one of the two choices (and is zero if there is no net bias). For instance, in a random-dot motion direction discrimination task, μ is proportional to the strength of motion signal. This system is a perfect integrator of the input. The integration process is terminated and the decision time is read out, whenever $X(t)$ reaches a positive threshold θ (choice A) or a negative threshold $-\theta$ (choice B). If the drift rate μ is positive, then choice A is correct, whereas choice B is an error. Therefore, this type of models is commonly referred to as ramping-to-threshold model, with the average ramping slope given by μ.

A biophysically based neural circuit model has been proposed for decision making (Wang, 2002). This model reproduces not only behavioral observations, but also single neural activity associated with decision making observed in a monkey experiment (Roitman and Shadlen, 2002). Moreover, it suggests a specific biological basis for temporal accumulation of evidence in decision-making. The drift diffusion model is an ideal perfect integrator (with an infinite time constant), whereas neurons and synapses are leaky with short time constants of tens of milliseconds. The neural circuit model suggests that a long integration time can be realized in a decision network through recurrent excitation. Reverberating excitation represents a salient characteristic of cortical local circuits (Douglas et al., 1995; Douglas and Martin, 2004). When this positive feedback is sufficiently strong, recurrent excitation in interplay with synaptic inhibition can create multiple stable states (‘‘attractors’’). Such models have been initially proposed for working memory. The same models, provided that excitatory reverberation is slow (i.e., mediated by the NMDARs), has been shown to be capable of decision-making computations (Wang, 2002, 2008; Machens et al., 2005; Miller and Wang, 2006; Wong and Wang, 2006; Soltani and Wang, 2006; Deco et al., 2007, 2009; Furman and Wang, 2008; Engel and Wang, 2011; Hunt et al., 2012). Interestingly, physiological studies in behaving nonhuman primates often reported neural activity correlated with decision making in cortical areas such as the prefrontal cortex or the parietal cortex, that also exhibit mnemonic persistent activity during working memory. Hence, this model and supporting experimental data suggest a common, ‘‘cognitive-type’’ circuit mechanism for decision making and working memory in the brain (Wang, 2013).

Behavioral modeling is often powerful in describing computations that solve a problem normatively or algorithmically. On the other hand, neural circuit models may be more suited for enabling us to investigate the underlying neural mechanisms and potentially pharmacologic or genetic manipulations of the circuits. Importantly, neural circuit models are not merely implementations of abstract mathematical models. For instance, the two types of models of perceptual decision making have distinct predictions at the behavioral level (Wang, 2008). These approaches are usually developed independently, but we are witnessing some convergence of the two in recent years. For example, spiking network models have been shown to have the capability of fitting quantitatively with behavioral performance (accuracy and reaction time) data (Lo et al., 2009), whereas such data fitting and model comparisons are commonly done with more abstract models due to their lower computational cost. Spiking network models can also be reduced to population rate models (Wong and Wang, 2006), that have features of abstract connectionist models. On the other hand, connectionist neural network models have increasingly taken biological information (with identified brain structures, receptors, etc) into account (O’Reilly and Frank, 2006). Thus, to bridge gaps in the current knowledge base and to facilitate research, there are advantages to move back and forth across several models that

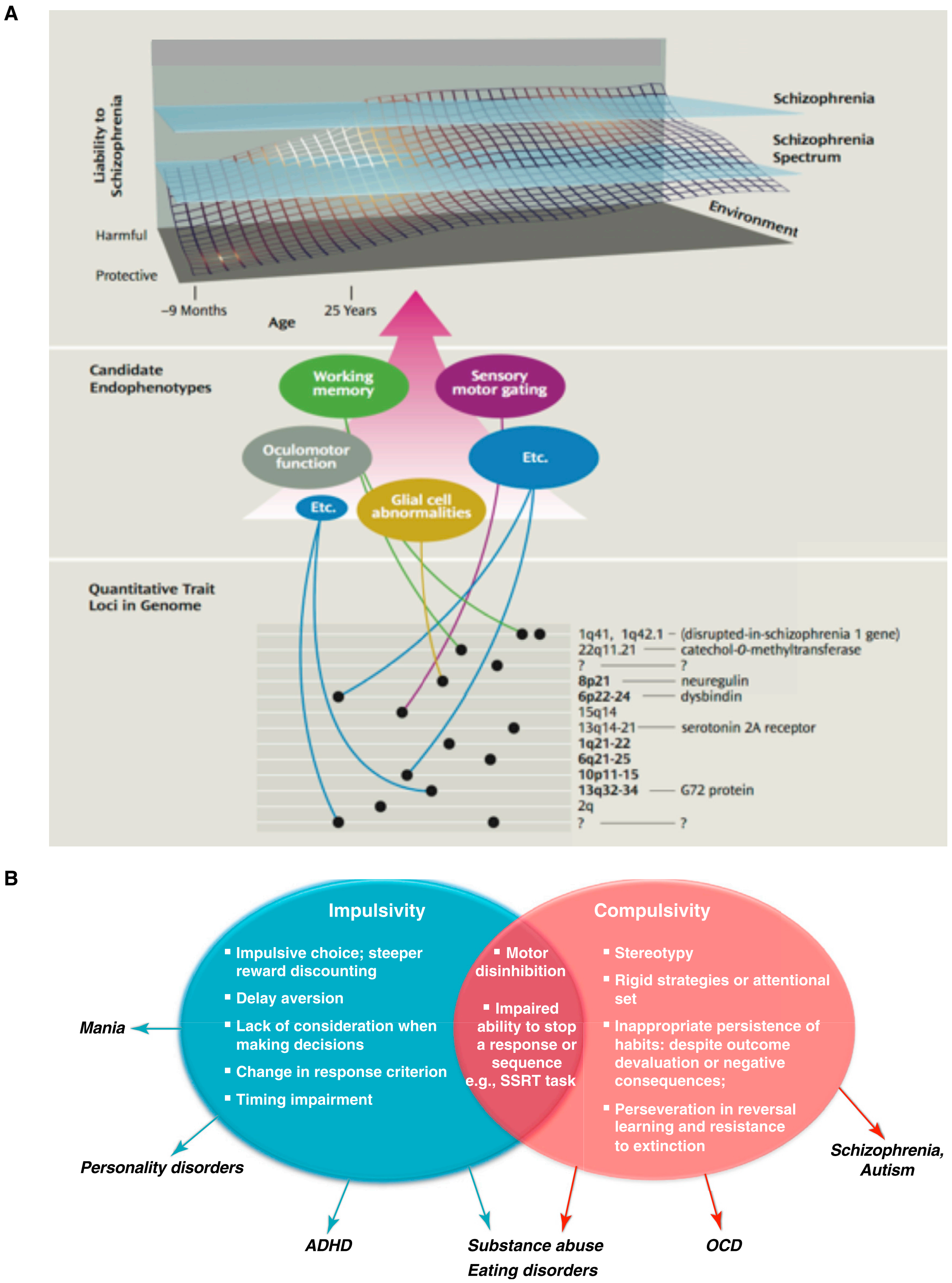
A
Liability to Schizophrenia
Schizophrenia
Schizophrenia Spectrum
Environment
Harmful
Protective
–9 Months
25 Years
Age
Candidate Endophenotypes
Working memory
Sensory motor gating
Oculomotor function
Etc.
Etc.
Glial cell abnormalities
Quantitative Trait Loci in Genome
1q41, 1q42.1 – (disrupted-in-schizophrenia 1 gene)
22q11.21 — catechol-O-methyltransferase
? — ?
8p21 — neuregulin
6p22-24 — dysbindin
15q14
13q14-21 — serotonin 2A receptor
1q21-22
6q21-25
10p11-15
13q32-34 — G72 protein
2q
? — ?
B
Impulsivity
Impulsive choice; steeper reward discounting
Delay aversion
Lack of consideration when making decisions
Change in response criterion
Timing impairment
Compulsivity
Motor disinhibition
Impaired ability to stop a response or sequence e.g., SSRT task
Stereotypy
Rigid strategies or attentional set
Inappropriate persistence of habits: despite outcome devaluation or negative consequences;
Perseveration in reversal learning and resistance to extinction
Mania
Personality disorders
ADHD
Substance abuse
Eating disorders
OCD
Schizophrenia, Autism

Figure 2. Research on Endophenotypes Cuts across Traditionally Defined Psychiatric Categories
(A) Gene regions, genes, and putative endophenotypes implicated in a biological systems approach to schizophrenia research. The dynamic developmental interplay among genetic, environmental, and epigenetic factors that produce cumulative liability to developing schizophrenia. Endophenotypes as schizophrenia discriminators involve sensory motor gating, oculomotor function, working memory, and glial cell abnormalities. Many more gene loci, genes, and candidate endophenotypes remain to be discovered (represented by question marks). The figure is not to scale.

(legend continued on next page)

vary in their degree of abstraction, biological realism, and their level of analysis (circuits, computational operations, behaviors).

Endophenotypes across Brain Disorder Categories

Inasmuch as features of the pathophysiology of psychiatric disorders are shared across diagnostic boundaries (Krueger 1999), a promising research direction is to search for transdiagnostic endophenotypes, i.e., quantitative heritable traits that are intermediate between risk genotypes and the psychiatric disorder syndrome itself (Figure 2A; Gottesman and Gould, 2003). While it has yet to be demonstrated that endophenotypes have a more simple genetics than psychiatric diagnoses, there remains a hope that endophenotypes may be more precisely defined, measured, and related to the underlying biology and to animal models. For instance, impulsivity and compulsivity are behavioral endophenotypes that cut across a range of diagnostic categories including obsessive-compulsive disorders, substance dependence, and attention-deficit hyperactivity disorder. Neither impulsivity nor compulsivity may be unitary constructs, but they may derive from a set of psychological processes which themselves are candidate endophenotypes (Figure 2B; Robbins et al., 2012). Thus, one could show impulsive choice behavior because of an aversion to delayed gratification, or impulsive response due to motor disinhibition or timing impairment. While this dimensional approach has not supplanted the prevailing psychiatric diagnostic schema, it has powerfully stimulated psychiatry research.

It is a major challenge to accurately and reliably identify endophenotypes. To make progress, it is beneficiary to complement consideration of symptoms (how people feel) with attention to what people do (choices and actions). By using behavioral paradigms that are designed to probe a specific cognitive function or functional domain, one can quantify the abnormalities of a particular function that are shared by multiple mental disorders. Those carefully designed tasks should be doable by both human subjects and nonhuman animals, thereby enabling more productive translational research (Carter et al., 2008; Wang, 2013; Insel, 2014). Theories can be developed and applied to both normal subjects and patients, providing insights into the core of a brain dysfunction.

Consider the case of disturbances in decision making. Many people who meet current diagnostic criteria for a number of neuropsychiatric disorders repeatedly make bad choices in the social, vocational, and recreational domains that compromise the quality of their lives. There is increasing evidence that specific impairments in decision making may represent cognitive endophenotypes across diagnostic boundaries (Robbins et al., 2012; Montague et al., 2012). A number of studies have dealt with the valuation process in reward-based decision making. The computations that enable one to learn to evaluate alternative options through experience are fundamental for adaptive choice behavior, i.e., to make a choice, assess its outcome, and to use this experience to guide the next choice. Reinforcement learning (RL) theory (Rescorla and Wagner, 1972; Sutton and Barto, 1998; Rangel et al., 2008) offers a framework for this adaptive process and impairments associated with psychiatric conditions (Montague et al., 2012; Maia and Frank, 2011; Lee, 2013). This field, which lies at the interface behavior and neurobiological mechanisms, was galvanized by the discovery that phasic activity of dopamine neurons in the ventral tegmental area signals reward prediction error (RPE) (Montague et al., 1996; Schultz et al., 1997). Specifically, dopamine phasic firing has been shown to confirm with RPE according to temporal-difference RL (TDRL) (Sutton and Barto, 1998; Dayan and Abbott, 2001). TDRL computes the reward expectation in terms of all anticipated reward events in the future, and learns to predict reward by driving RPE to zero. For the sake of simplicity, here we describe a simplified notion of RPE, $\delta_t = r_t - V_t$, where r_t is the actual reward and V_t is the expected reward, at time *t*. The idea is that the mismatch between the actual reward and the expected reward generates an "error signal" that informs learning. RL is hypothesized to be driven by $\alpha\delta_t$, with the rate α controlling the speed of learning. Therefore, there is a solid foundation for bridging reward-related learning with a specific underlying brain circuit (the dopamine system). Empirical evidence for impaired RL has been documented for Parkinson disease, schizophrenia, Tourette syndrome, attention-deficit disorder, drug addiction, and depression (Maia and Frank, 2011; Lee, 2013; Huys et al., 2013), demonstrating powerfully the importance of function-based, transdiagnostic, approach in psychiatry.

For instance, addiction can be viewed as RL gone awry. Indeed, a pioneering application of RL to psychiatry (Redish, 2004; Redish et al., 2007) was inspired by TDRL. It was proposed that addiction accesses the same RL system as in the normal brain, but drug-induced positive prediction errors could produce unbounded increases in the value of drug receipt. A merit of such quantitative models is that they are precise enough to be falsifiable by new experiments, a hallmark of scientific inquiry. Redish's model predicts that a behavioral trait called blocking does not occur when drugs are used as unconditional reinforcers. Blocking refers to the observation that after a subject learns to associate a stimulus A with a reward, later pairing A with another stimulus B should not lead to learning to associate B with the reward. If, however, drugs (as stimuli A and B) lead to unlimited value increase, blocking should not be observed. Behavioral experiments using cocaine as unconditional stimulus showed that this is not the case, i.e., blocking does occur (Panlilio et al., 2007). One possible interpretation of this result is that blocking is not due to the specific form TDRL of RL. Indeed, blocking is accounted for in an alternative model of addiction that assumes the expected reward V_t to be computed by a weighted average over past reward events (Dezfouli et al., 2009). Another possibility is that RL involves multiple competing systems (Redish et al., 2007).

The RL approach has also been applied to depression. Huys et al. (2013) set out to test the hypothesis that depression is

(B) The impulsivity and compulsivity constructs. The diagram describes possible psychological component mechanisms underlying the two constructs. It would appear that these different measures likely do not intercorrelate well, which would argue against a unitary construct for either impulsivity or compulsivity, but this issue is still actively being researched. Both impulsivity and compulsivity involve motor/response disinhibition, but at different stages of the response process. (A) Reproduced from Gottesman and Gould (2003) and (B) Robbins et al. (2012) with permission.

associated with an altered sensitivity to reward; specifically, the RPE becomes $\delta_t = \rho r_t - V_t$, where the parameter ρ represents reward sensitivity. Meta-analysis of experiments with about 50 healthy subjects and 50 subjects with major depression disorder has been carried out by fitting behavioral data with a RL model. It was found that compared to the control group, the patient group shows a significantly reduced reward sensitivity (a smaller value of ρ), but no change in the learning rate α, consistent with the anhedonia and lack of motivation found in patients with depression. Similar findings were also reported by Strauss et al. (2011). This work illustrates how computational modeling enables us to dissect distinct aspects (reward sensitivity but not learning rate) of a maladaptive behavior.

The RL theory is currently been extended beyond single-factor considerations. In particular, it has been recognized that RL involves two separate neural systems (Balleine and Dickinson, 1998; Daw et al., 2005, 2011; Kahneman, 2011; Dolan and Dayan, 2013). One of these systems subserves habits and related behaviors. It is referred to as "model-free" because these behaviors are elicited in an automatized way by cues. The second, model-based, system is endowed with an internal representation of the causal structure of the environment and underlies goal-oriented behaviors. The model-free and model-based systems must be balanced. A dual-system learning model (Daw et al., 2011) has been combined with human brain imaging to examine specific ways an imbalance of these two systems might lead to maladaptive choice behavior in mental illness. Using this framework, it was found that repeated exposure to addictive drugs shifts behavior from model-based to model-free emphasis (Kurth-Nelson and Redish, 2011; Lucantonio et al., 2012). Likewise, data fitting by the dual-system model revealed that subjects diagnosed with obsessive-compulsive disorder display a bias toward model-free habit acquisition (Voon et al., 2014). The central control mechanisms governing the balance maintenance and shifts between model-based and model-free systems represent an area of intense ongoing research (Simon and Daw, 2011).

Whereas the model-free system relies on RPE, the model-based system presumably depends on a more abstract "state prediction error" that might implicate lateral prefrontal cortex, giving rise to "dual system" RL models (Gläscher et al., 2010). RL approaches have advanced translational neuroscience research on such phenomena as delusions that have been previously extremely challenging to study from this perspective. The focus on prediction error, a mismatch between expectation and experience, has inspired neurobiological studies of psychosis (Corlett et al., 2010). Delusions are false beliefs about the world that persist tenaciously despite repeated encounters with contradicting evidence. Corlett et al. (2007) found that violations of causal associations activate the right lateral prefrontal cortex (rPFC) during fMRI, a putative prediction error signal, and, deficits in this fMRI prediction error signal among subjects with first-episode psychosis strongly correlated with the severity of delusions across subjects (Corlett et al., 2007). Thus, false beliefs may be generated through compromised prediction error and sustained as aberrant learning (Corlett et al., 2010).

RL has also been extended to hierarchically organized behaviors (Botvinick et al., 2009). These studies focused on RL illustrate well how theory and computational modeling, in conjunction with experimentation, can help dissect distinct component processes (such as reward sensitivity, learning rate, balance between model-free and model-based systems, etc.) that may be abnormal in multiple mental disorders but in different ways. This opens up the possibility that each cognitive endophenotype (such as impulsivity) could be defined in terms of a specific combination of quantitative impairments of these component processes. If so, future progress in this direction could yield a promising new framework to guide translational neuroscience studies of neuropsychiatric disorders.

Big Data and Model-Aided Diagnosis

Typically, the process of building from a behavioral experiment to a computational model follows several steps: (1) a cognitive task is strategically designed to probe a particular function (e.g., reward-related learning in decision making), (2) an appropriate computational model (e.g., reinforcement learning) is chosen to simulate the process (e.g., valuation) under consideration, and (3) model-fitting of data yields estimation of model parameters (e.g., reward sensitivity and learning rate). Many of these studies compare people deemed to be free of a psychiatric diagnosis to people who have been recruited specifically for the presence of a specific psychiatric diagnosis (e.g., according to DSM or international classification of diseases criteria). Significant differences between the healthy group and patient group in some model parameters (e.g., reward sensitivity but not learning rate) provide the basis for characterizing the presumed "abnormality" in the patient group. However, computational psychiatry is not limited to existing diagnostic schema. Its focus on relating mechanisms to cognitive operations and behavioral processes promotes a transdiagnostic perspective. For instance, a similar bias toward model-free versus model-based learning has been found in disorders involving both natural (binge eating) and artificial (methamphetamine) reward, as well as obsessive-compulsive disorder (Voon et al., 2014).

Recently, Frank and collaborators (Wiecki et al., 2014) proposed to extend this approach from subject groups to individuals. This requires a fourth step, i.e., to use sophisticated statistical analysis algorithms to investigate whether model parameter values extracted from individual subjects are clustered into distinct groups (Figure 3A). This step is crucial for this paradigm to potentially serve as a clinical tool, because diagnosis must obviously be done for single individuals. A similar approach has been advocated by Stephan and his colleagues (Figure 3B) (Brodersen et al., 2014). These authors proposed a cross-disciplinary approach that combines behavior, brain measures (fMRI), and computation (dynamical causal modeling, DCM; Friston et al., 2003; Stephan et al., 2007). In a working memory study of schizophrenic patients, they focused on DCM-based estimates of effective connectivity between visual, parietal and prefrontal cortex, since these three cortical areas were critically involved in their visual working memory task. An unsupervised clustering procedure operating on the individual connectivity patterns yielded three distinct patient subgroups (Figure 3C): those with greater fronto-parietal

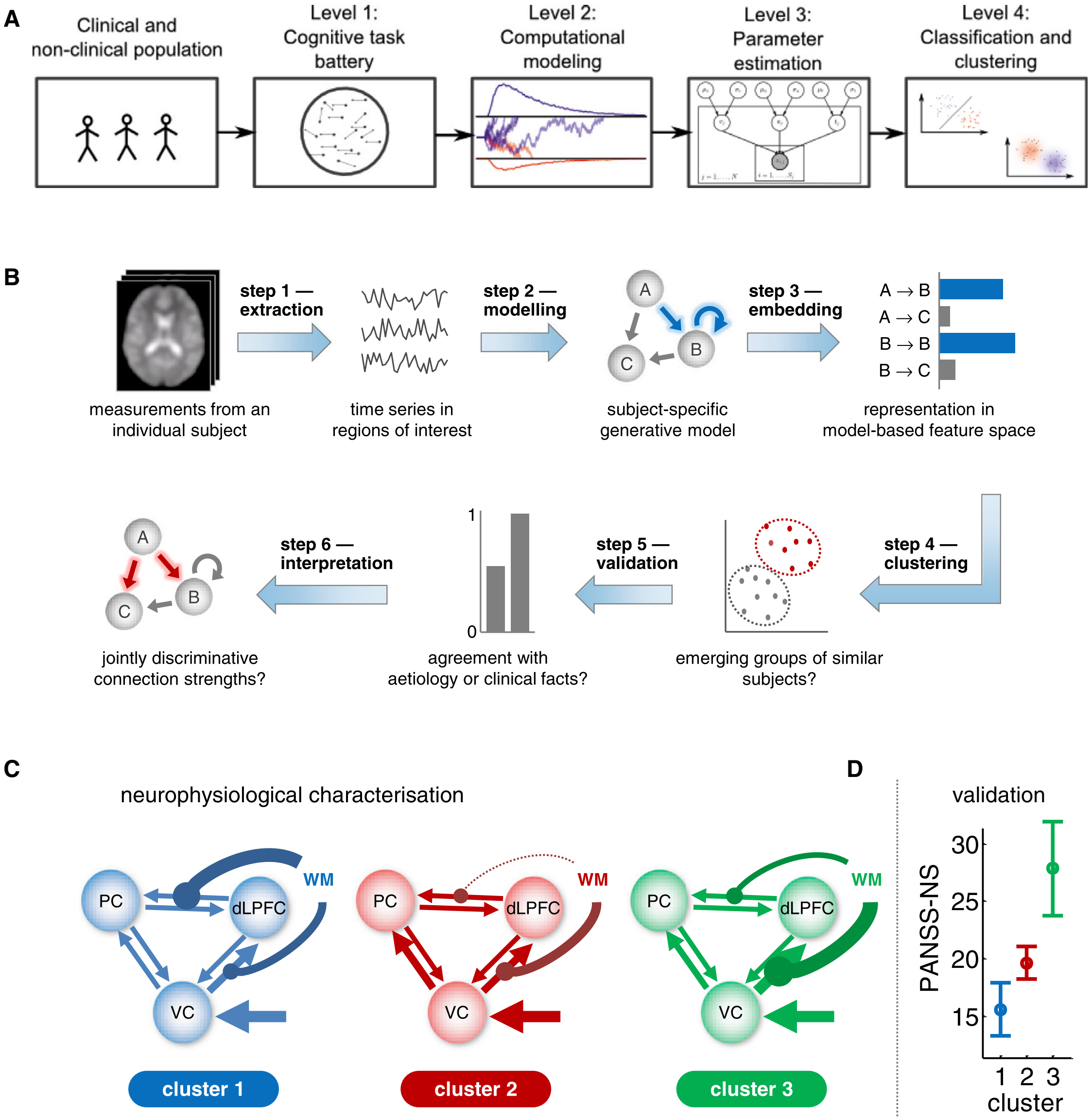

Figure 3. Model-Aided Subject Clustering as a Potential Diagnosis Tool

(A) Illustration of the four levels of computational psychiatry. Clinical and nonclinical populations are tested on a battery of cognitive tasks. Computational models can relate raw task performance (e.g., RT and accuracy) to psychological and/or neurocognitive processes. These models can be estimated via various methods. Finally, based on resulting computational multidimensional profile, training using learning algorithms can either uncover groups and subgroups in clinical and healthy populations, or relate model parameters to clinical symptom severity.

(B) Conceptual overview of model-aided clustering of fMRI data. First, separately for each subject, BOLD time series are extracted from a number of regions of interest. Second, subject-specific time series are used to estimate the parameters of a model. Third, subjects are embedded in a score space in which each dimension represents a specific model parameter. This space implies a similarity metric under which any two subjects can be compared. Fourth, a clustering algorithm is used to identify salient substructures in the data. Fifth, the resulting clusters are validated against known external (clinical) variables. Sixth, once validated, a clustering solution can be interpreted mechanistically in the context of the underlying model.

(C and D) Model-based clustering of fMRI data from patients with schizophrenia in a working memory task. (C) An unsupervised clustering analysis of the patient group only, using Gaussian mixture models operating on dynamical causal model (DCM) parameter estimates, yield the average posterior parameter estimates (in terms of maximum a posteriori estimates) for each coupling and input parameter in the model. This is displayed graphically by the thickness of the respective arrows. (D) The three subgroups, which are defined based on connection strengths, also differ in terms of negative clinical symptoms as operationalized by the negative symptoms (NS) subscale of the PANSS score.

(A) Reproduced from Wiecki et al. (2014) and (B–D) Brodersen et al. (2014) with permission.

connectivity, those with weaker fronto-parietal connectivity, and those with greater visuo-frontal connectivity. The authors further pushed the approach by including two more steps (Figure 3B): (5) assessment of whether clusters of subjects obtained by model-fitting are correlated with different severity of behavioral impairment (indeed they found that subjects in the three clusters display a different degree of negative symptom severity; Figure 3D), and (6) interpretation of the results from step (5) that attributes the behavioral deficit (negative symptom) to a possible underlying brain substrate (visual-parietal-prefrontal circuitry connectivity), generating new hypotheses to be tested in future research.

This line of work raises the question of whether it might be possible to use brain imaging data (or models of such data) rather than symptoms as the substrate for diagnostic classification schema. A related line of thinking is to view psychiatric illness from the perspective of brain connectome (Rubinov and Bullmore, 2013), according to which the analysis of functional connectivity patterns inferred from brain imaging offers a window to pathoconnectomics associated with mental disorders. It would be interesting to know the impact of attempting to, on a very large scale, identify model parameters that cluster patients in new ways. Would this approach yield a classification schema different from DSM? Would this classification schema be replicable and generalizable? Would it suggest new directions for research and treatment? This type of strategy might address a conundrum in psychiatry, which is the absence of biomarkers. It may be impossible to develop meaningful illness biomarkers within a diagnostic framework that is not based in biology. However, if the diagnostic framework were, itself, built around an imaging biomarker, then it would seem highly likely that this biomarker would have predictive power in relation to diagnosis and treatment.

A number of factors will determine the success of this framework: very large samples of subjects, efficient and statistically reliable analysis methods, and judicious choices as well as in-depth understanding of computational models. With the advance of big data science, and computational modeling, a radical modern paradigm shift may be on the horizon.

Biophysically Based Neural Circuit Modeling: Understanding across Levels

In contrast to more abstract models, biophysically realistic neural circuit modeling has the potential to be rigorously calibrated by quantitative neurophysiology and anatomy. Ultimately, this is necessary to elucidate deficits at the molecular, cellular, and circuit levels that underlie cognitive and behavioral disorders in mental illness.

Among hierarchically interrelated cognitive dysfunctions associated with schizophrenia (Millan et al., 2012), perhaps the best studied is working memory (Park and Holzman, 1992; Lee and Park, 2005; Lewis and Gonzalez-Burgos, 2006; Barch and Ceaser, 2012). Working memory, the brain's ability to encode and sustain the neural representation of information in the absence of direct sensory stimulation and to manipulate this information in the service of future action, is a core cognitive function that depends on the PFC (Fuster, 2008; Goldman-Rakic, 1995; D'Esposito, 2007; Baddeley, 2012). Fortunately, working memory has been particularly amenable to biophysically based neural circuit modeling, because of the richness of experimental data at multiple levels of study.

A well-known working memory paradigm is the delayed oculomotor response task, in which a subject is required to remember a visual cue (a directional angle) across a delay period to perform a memory-guided saccadic eye movement (Funahashi et al., 1989; Constantinidis and Wang, 2004). A biologically-based network model of spiking neurons has been developed for this spatial working memory experiment (Figure 4A) (Compte et al., 2000; Renart et al., 2003; Wang et al., 2004; Carter and Wang, 2007; Wei et al., 2012; Kilpatrick et al., 2013; Hansel and Mato, 2013; Pereira and Wang, 2014). Figure 4B shows a model simulation of the delayed oculomotor task. Initially, the network is in a resting state in which all cells fire spontaneously at low rates. A transient input drives a subpopulation of cells to fire at high rates. As a result, they send recruited excitation to each other via horizontal connections. This internal excitation is large enough to sustain elevated activity, so that the firing pattern persists after the stimulus is withdrawn. Synaptic inhibition ensures that the activity does not spread to the rest of the network, and persistent activity has a localized, bell shape (''bump attractor''). At the end of a mnemonic delay period, the cue information can be retrieved by reading out the peak location of the persistent activity pattern; and the network is reset back to the resting state. This type of spatial working memory network is endowed with a continuous family of bump attractors, each encoding a specific potential location.

In this model, a mnemonic persistent activity pattern is sustained internally by strong recurrent excitation, which the model predicts to be slow and dependent on the NMDAR-mediated synaptic transmission at local synapses (Wang, 1999, 2001; Wang et al., 2008) (Figure 4C). In a recent experiment with monkeys performing a working memory task (Wang et al., 2013), iontophoresis of drugs that blocked the NMDARs suppressed delay-period persistent activity of PFC (Figure 4D), in support of an important role of the NMDARs in PFC processes. Another monkey experiment showed that ketamine (an NMDAR antagonist) reduces task selectivity of PFC neurons in parallel with behavioral impairment (Skoblenick and Everling, 2012). These findings are directly relevant to psychiatry. Indeed, it has been hypothesized that NMDA hypofunction underlies working memory deficits in schizophrenia (Coyle et al., 2003; Moghaddam and Krystal, 2012), and subanesthetic dose of ketamine produces working memory impairment in healthy human subjects, similar to that seen in schizophrenia (Krystal et al., 1994). The finding that NMDARs are critical for mnemonic persistent activity and its selectivity offers a possible mechanistic explanation as to why NMDA signaling pathway is essential for working memory function.

Like yin and yang in ancient Chinese philosophy, the dynamic balance between synaptic excitation and inhibition within local and distributed networks is a fundamental property of cortical function. This balance is important for normal functions within a biophysically based PFC neural circuit model because it defines many emergent properties of the network including: dynamic network stability (if unchecked by inhibition, strong recurrent excitation would lead to runaway positive feedback), fast coherent oscillations (generated by the interplay between fast AMPA receptor-mediated excitation and slower GABA$_A$ receptor-mediated inhibition), stimulus selectivity (synaptic inhibition is critical for neural tuning), and resistance to distractors (reduced responsiveness to distracting stimuli by neurons not involved in memory storage) (Compte et al., 2000; Brunel and Wang, 2001; Wang, 2013).

These results have functional implications for the observed pathology of inhibitory circuits associated with schizophrenia (Lewis et al., 2005, 2012). In particular, enhanced distractibility

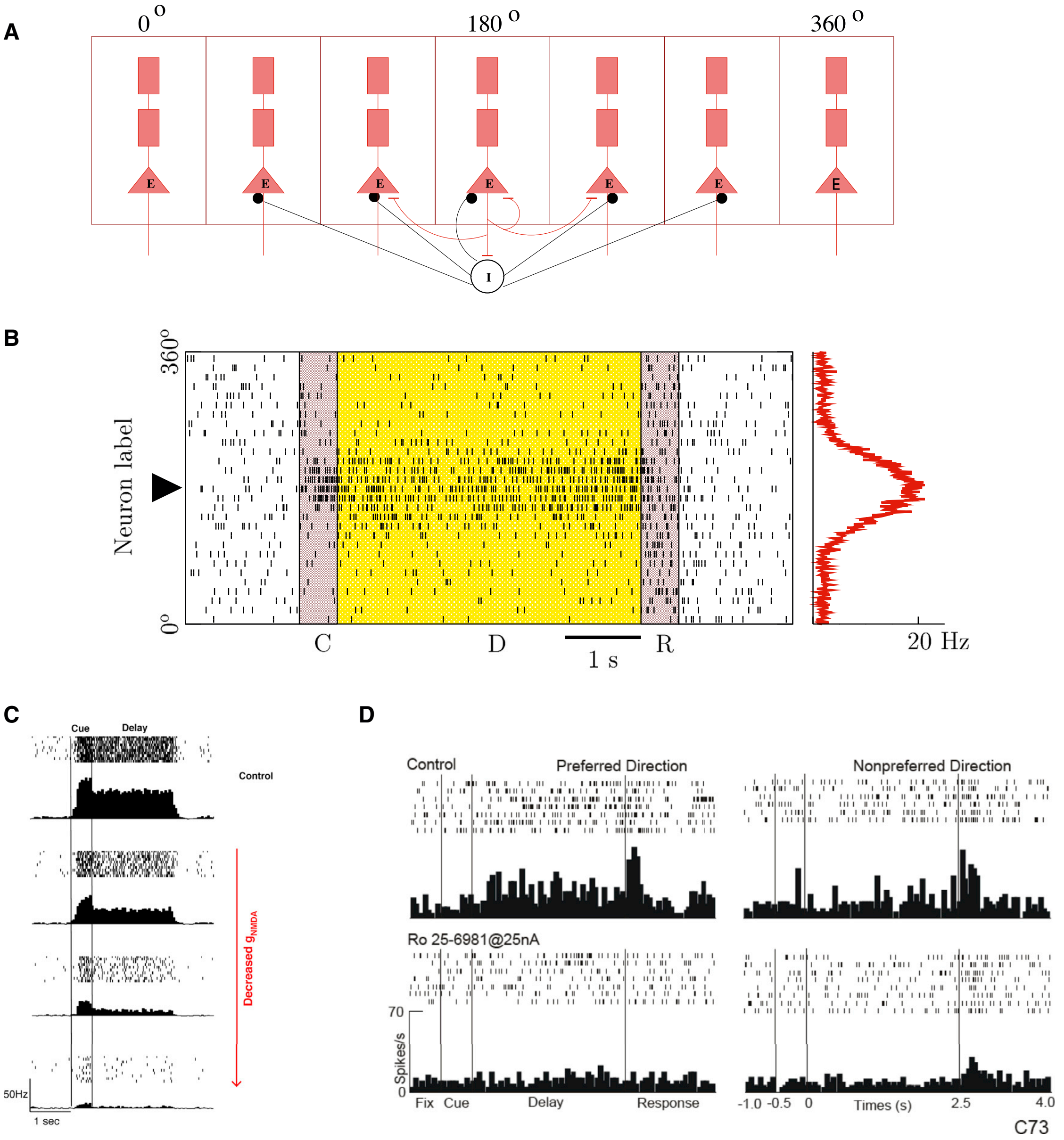

Figure 4. Spatial Working Memory Modeling and the Role of NMDA Receptors in Mnemonic Persistent Activity

(A and B) Spiking network model of working memory. (A) Model architecture. Excitatory pyramidal cells are labeled by their preferred cues (0° to 360°). Pyramidal cells of similar preferred cues are connected through local excitatory-to-excitatory connections. Inhibitory interneurons receive inputs from excitatory cells and send feedback inhibition by broad projections. (B) A stimulus is encoded and actively maintained by a self-sustained network persistent activity pattern (a "bump attractor") in a simulation of the delayed oculomotor experiment. C, cue period; D, delay period; R, response period. Pyramidal neurons are labeled along the y axis according to their preferred cues. The x axis represents time. A dot in the rastergram indicates a spike of a neuron whose preferred location is at y, at time x. An elevated and localized neural activity is triggered by a transient cue stimulus and persists during the delay period.

(C) The effects of iontophoretic NMDA blockade on working memory activity in a computational model of working memory. Under control conditions, a stimulus cue selectively activates a group of neurons, leading to persistent activity sustained by NMDAR-dependent recurrent excitation. NMDA conductance is reduced from control to 90%, 80%, and 70% (to bottom) of a reference level in a few pyramidal neurons in the network model. Stimulus-selective persistent activity gradually decreases with more NMDAR blockade and eventually disappears in these affected cells.

(D) An example of an individual dorsolateral PFC cell recorded from behaving monkey in a delayed oculomotor response task. Upper: control condition; lower: after iontophoresis of Ro 25-6981 (25 nA), a blocker of NR2B-containing NMDARs. The rasters and histograms show firing patterns for the neuron's preferred direction and the nonpreferred direction (opposite to the preferred direction). Iontophoresis of Ro 25-6981 markedly reduced mnemonic delay period firing to baseline.

(B) Adapted from Compte et al. (2000) and (C and D) Wang et al. (2013) with permission.

represents a common behavioral deficit in schizophrenic patients (Goldman-Rakic, 1987; Mesulam, 2000; Luck and Gold, 2008). A recent computational study examined how a reduced inhibition might lead to PFC's deficient ability to filter out distracting stimuli during working memory (Murray et al., 2014). Disinhibition induced a broadening of the neural representation for the memorandum maintained in working memory through persistent activity (Figure 5A). Importantly, this feature of the circuit was a function of the overall balance between excitation and inhibition (Figure 5B). Neural broadening, in turn, induced specific behavioral deficits, making working memory more vulnerable to intervening distractors. In the model, distractibility depends on the similarity between the distractor and the mnemonic representation, and therefore broadening the mnemonic representation increases the range of distractors that can disrupt behavior. The authors tested this model prediction by analyzing behavior from healthy humans administered ketamine, a pharmacological model of schizophrenia, during a spatial delayed match-to-sample task. Matching the model prediction, ketamine increased the rate of errors specifically for distractors that would overlap with a broadened mnemonic representation (Figure 5C). Just as the biophysical basis of the model allows instantiation of potential pathologies, it can also readily explore pharmacological treatments to compensate for these deficits. In particular, in this model it was demonstrated as proof-of-principle that glutamatergic or GABAergic manipulations could restore excitation-inhibition balance, reversing the broadened mnemonic representations and corresponding distractibility induced by disinhibition (Figure 5D). An open question is concerned with the brain mechanisms for deciding which information should be considered task-relevant versus distracting and how this may or may not be related to reward value processing of potentially relevant or distracting stimuli. Impairments of this decision process could be relatively independent from those of working memory circuit's ability to resist distractors as described above, which would suggest an orthogonality between these deficits. Future research is needed to assess whether this is indeed the case.

In the model, the network's ability to filter out distractors is impaired by a reduced excitation in inhibitory neurons. The main insight is that predominant behavioral disturbance due to modest disinhibition may not be so much the inability of memory storage per se as the difficulty of ignoring behaviorally irrelevant inputs during memory maintenance. The observation that ketamine in human subjects leads to impaired resistance against near distractors, as predicted by the model, suggests that disinhibition involves NMDA receptors (NMDAR). Intuitively, this could be caused by a reduced NMDAR-mediated excitation in inhibitory neurons. In support of this view, there is evidence that, in rodents, acute ketamine administration led to a decreased activity of putative fast-spiking interneurons, and increased activity of putative pyramidal cells (Homayoun and Moghaddam, 2007). Moreover, because fast-spiking inhibitory neurons are critically involved in the generation of fast γ oscillations (Buzsáki and Wang, 2012; Wang, 2010), a reduced excitation of those neurons could explain abnormal γ synchrony observed in schizophrenic patients (Spencer et al., 2004; Lisman et al., 2008).

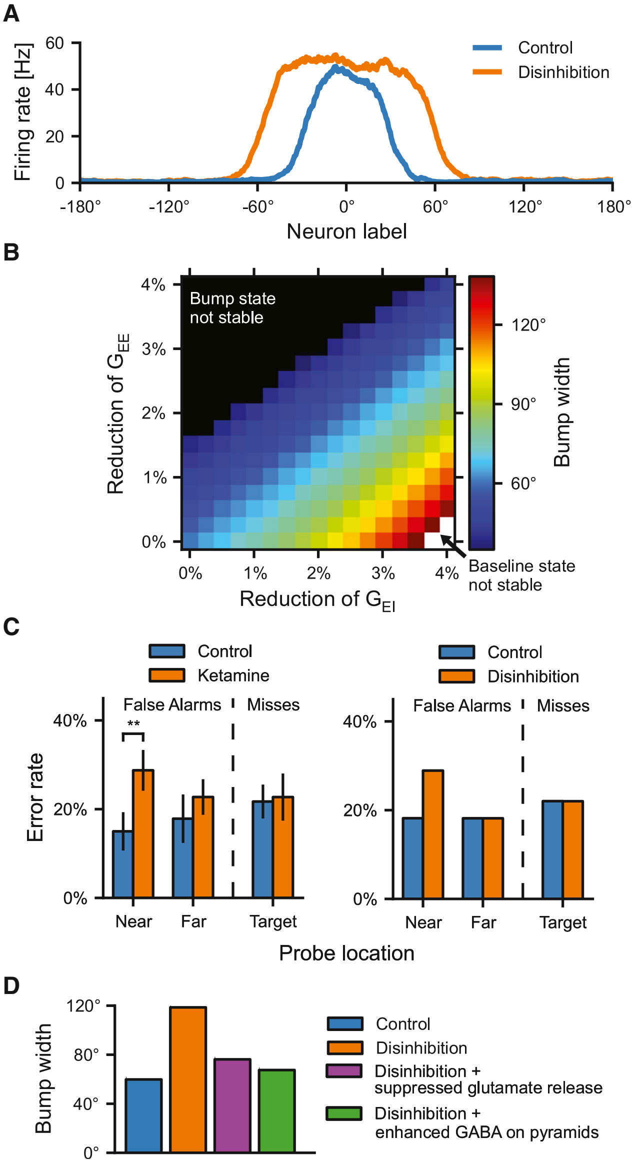

Figure 5. Computational Modeling of Excitation-Inhibition Balance in Working Memory Circuits

(A) A spatial working-memory model can generate a bump-shaped stimulus-selective persistent activity pattern following stimulus withdrawal. Disinhibition, mediated by NMDAR hypofunction on interneurons, broadens working-memory representations at the neural level.

(B) The parameter space of NMDAR hypofunction highlights the importance of E/I balance for working memory function. If the E/I ratio is elevated as in disinhibition, the width of the representation increases. In contrast, if the E/I ratio is reduced too much through weakened recurrent excitation between pyramidal cells, the circuit cannot support memory-related persistent activity (upper left corner).

(C) Broadening of working-memory representations was tested using behavioral data from human subjects performing a spatial working-memory task combined with ketamine infusion, a pharmacological model of schizophrenia. Consistent with broadening, ketamine induced errors specifically for near distractor probes (left), as predicted by the model (right).

(D) Compensations can restore E/I balance and ameliorate behavioral deficits in the model. We paired the disinhibition mechanism with either reduced excitation (purple) or increased inhibition (green), following proposed pharmacological treatments.

Adapted with permission from Murray et al. (2014).

However, in fast-spiking interneurons of the mice frontal cortex, NMDAR-mediated excitation is small and insensitive to NMDAR blocker AP5 (Rotaru et al., 2011). In adult rats, the majority of fast-spiking interneurons are devoid of NMDARs, whereas NMDAR-dependent synaptic excitation is more significant in other subclasses of regular-spiking and low-threshold spiking inhibitory cells (Wang and Gao, 2009). The latter mediate dendritic inhibition, thereby gating synaptic inputs onto pyramidal cells. Further, the dendrite-targeting interneurons function in an input-specific manner, enabling pyramidal neurons to be selectively activated by task-relevant inputs. This has been incorporated in an extended working memory microcircuit model endowed with three subtypes of inhibitory neurons: soma-targeting interneurons that express parvalbumin and control pyramidal firing output, interneurons that express calbindin or somatostatin and gate dendritic inputs to pyramidal cells, and interneurons that express calrintinin or vasoactive intestinal peptide and preferentially target dendrite-targeting interneurons (thereby providing a new disinhibition mechanism) (Wang et al., 2004; Wang, 2013). It was found that dendritic inhibition controls the network's ability to resist irrelevant distractors more effectively than perisomatic inhibition that controls the spiking output of pyramidal neurons. Taken together, one plausible scenario consistent with currently available evidence is that disinhibition induced by ketamine results from a reduction of NMDAR-dependent excitation of dendrite-targeting interneurons. This prediction can be tested using cell-type specific genetic tools (Kepecs and Fishell, 2014; Higley, 2014) in future animal experiments.

What happens when the excitation-inhibition balance is tilted in a way that synaptic excitation becomes excessively strong? Model simulations showed that one consequence of such an imbalance could lead to behavioral inflexibility: attractor states encoding memory items become so robust that it becomes difficult to switch off from one memory attractor state either to rest (memory erasure) or another memory state (Rolls et al., 2008; Durstewitz and Seamans, 2008; Gruber et al., 2010). This idea is interesting especially in the light of the fact that working memory is not limited to sensory stimuli but also more abstract information such as behavioral task sets or rules (Miller and Cohen, 2001; Wallis et al., 2001; Sakai, 2008; Buckley et al., 2009; Lapish et al., 2008; Sigala et al., 2008), and attractor network models have been extended to internal representation of behavioral rule or context in flexible behavior (Rigotti et al., 2010, 2013). Thus, behavioral inflexibility may be reflected in the difficulty to make a transition from a behavioral context to another one, which is a hallmark of abnormal cognition in schizophrenia.

This framework is also useful for analyzing abnormal neuromodulation in mental illness. The dopamine system represents an example par excellence. It is well known that working memory performance exhibits an inverted U-shaped dependence on dopamine modulation: too little dopamine, you lose working memory; too much dopamine, you are inflexible with switching on and off in a working memory system. Dopamine modulation acts on targets such as NMDAR-mediated excitatory synaptic excitation and GABA-mediated inhibitory synaptic inhibition (Brunel and Wang, 2001; Seamans et al., 2001; Durstewitz et al., 2000), or the gain of single-neuron input-output relationship (Cohen and Servan-Schreiber, 1992). Computational modeling showed that an inverted-U shape of dopamine modulation can be readily explained if dopamine modulation has a differential sensitivity to the NMDA conductance and GABA conductance (Brunel and Wang, 2001). Furthermore, interestingly, the network's ability to ignore distractors is sensitive to modulation by dopamine of recurrent excitation and inhibition. Therefore, even a mild impairment of dopaminergic signaling in the prefrontal cortex could be very detrimental to robust working memory maintenance in spite of ongoing sensory flow.

These studies on working memory demonstrate how biophysically based modeling in interplay with experimentation can play a powerful role in making discoveries and producing new hypotheses about the brain mechanisms of core cognitive processes implicated in psychiatric disorders.

Looking Forward: Building a New Cross-Disciplinary Field

The economic cost of mental illness represents an enormous burden on the society (Wittchen et al., 2011; Olesen et al., 2012; Vos et al., 2012). The critical nature of our knowledge gap for the clinical neuroscience fields, including neurology, neurosurgery, psychiatry, and psychology, is well known. In the United States, NIH initiatives, including the Human Connectome Project (http://www.humanconnectomeproject.org) and the BRAIN Initiative (http://www.braininitiative.nih.gov/index.htm), are designed to advance current approaches and to develop new technologies to characterize brain circuit function. Parallel initiatives are underway in Europe and Asia.

In this Perspective, we marshaled findings from recent work on reinforcement learning and working memory to argue for a computational psychiatry approach to brain disorders. This perspective emphasizes an integration of experimentation, data analysis, and theory in concerted efforts to understand neural circuits involved in mental illness. Although we have focused on local circuit mechanisms, computational psychiatry must also be developed for large brain systems. We need to develop large-scale brain circuit models to investigate how the PFC controls and interacts with many other brain regions in a highly interconnected complex system. A notable line of research in this regard is concerned with the interplay between the PFC and basal ganglia, which is important for both working memory and decision making (O'Reilly and Frank, 2006; Lo and Wang, 2006; Ding and Gold, 2013). In fact, behavioral evidence from a cleverly designed experiment suggests that impaired RL in schizophrenia is attributable, largely, to working memory deficits rather than valuation process (Collins et al., 2014). Another interplay involves cortex and thalamus (Vukadinovic, 2011; Anticevic et al., 2013b). More broadly, new approaches applied to the study of the connectivity properties of large-scale brain systems are exciting advances (Sporns, 2009; Bullmore and Sporns, 2009; Markov et al., 2013) with important implications for psychiatric disorders (Anticevic et al., 2013a; Rubinov and Bullmore, 2013; Yang et al., 2014).

Unprecedented ongoing progress in neuroscience offers extraordinary opportunities as well as challenges. First, progress in genomics, massive neuroimaging, and other advances are creating enormous data sets that, in turn, require new mathematical/statistical tools. Second, a pressing need is to develop new ideas for cross-level investigations. For instance, genome-wide analysis revealed that genes encoding L-type voltage-gated calcium channels are associated with several psychiatric disorders including schizophrenia (Smoller et al., 2013). How would alternations of L-type calcium channels give rise to abnormal circuit formation, ultimately explain specific mental and behavioral disturbances? This question could be addressed using modeling that enables us to go back and forth between different levels (from molecules and cells to circuits and behavior). Third, major mental disorders like schizophrenia, autism, and attention-deficit hyperactivity disorder are neurodevelopmental diseases (Moore et al., 2006; Belujon and Grace, 2008; Insel, 2010; Fair et al., 2012). Thus, it is critical to build computational models for investigating developmental changes in synaptic and circuit function in disease-related models. For instance, the human neural representation of working memory assessed with fMRI changes during adolescence (Satterthwaite et al., 2013). Similarly, synaptic mechanisms evolve during adolescence. In rodents, for example, NMDARs are abundant on parvalbumin-expressing interneurons early in life, but they are present more sparsely in adults (Belforte et al., 2010). In these circuits, reducing NMDAR expression early in life, but not in adulthood, impairs cognitive function in adulthood. There is a dearth of computational modeling dedicated to understanding critical periods in neurodevelopment and the impact of even ‘‘transient’’ disruption on circuit development and cognitive function in adulthood. Progress along these lines will require sophisticated neural circuit modeling in conjunction with genetic, physiological, and imaging experimentation. Fourth and finally, can one quantitatively capture specific features of dysfunctional flow of thought associated with mental illness? A recent work took the view that language could be used ‘‘as a privileged measuring lens into thought,’’ and showed that quantitative analysis of speech could yield accurate sorting of schizophrenia versus mania with high sensitivity and specificity (Mota et al., 2012). Language is a human cognitive ability implicated in mental disorders, thus elucidation of brain’s language circuit represents another neuroscientific theme relevant to psychiatry.

It is our belief that these challenges cannot be overcome without theory and computational modeling. To advance the field, we need new infrastructure, resources, and training of cross-disciplinary young talents who are well versed both in mathematical modeling and experimentation. First, it would be important to develop training programs whereby graduate students and postdoctoral fellows trained in the physical and mathematical sciences could more easily be introduced to and engage in psychiatric research. Second, it will be important to develop a cadre of young psychiatrists who learn computational modeling, lest computational psychiatry develop without the input of physician-scientists. Third, government funding agencies and nonprofit organizations and foundations should offer new programs to promote highly cross-disciplinary education and research in computational psychiatry. Through these concerted efforts, we are optimistic that computational psychiatry could play an indispensable role in addressing the great challenges of mental health in the 21st century.

ACKNOWLEDGMENTS

We thank John Murray, Alan Anticevic, and Philip Corlett for discussions related to computational psychiatry and John Murray for making Figure 5; Michael Frank, Nathaniel Daw, Ray Dolan, and Klaas Enno Stephan for their comments on an earlier version of the manuscript. This work was supported by NIH grant R01MH062349 and ONR grant N00014-13-1-0297 (to X.-J.W.), National Institute on Alcohol Abuse and Alcoholism grant (2P50AA012870-11 to J.H.K.), National Center for Advancing Translational Science grants (UL1RR024139, principal investigator: R. Sherwin; UH2TR000960 to J.H.K.), and the U.S. Department of Veterans Affairs through their support of the VA National Center for PTSD and the Consortium to Alleviate PTSD (to J.H.K). None of the authors of this manuscript have a financial interest related to this work.

REFERENCES

Abbott, L.F. (2008). Theoretical neuroscience rising. Neuron *60*, 489–495.

Anticevic, A., Cole, M.W., Repovs, G., Savic, A., Driesen, N.R., Yang, G., Cho, Y.T., Murray, J.D., Glahn, D.C., Wang, X.J., and Krystal, J.H. (2013a). Connectivity, pharmacology, and computation: toward a mechanistic understanding of neural system dysfunction in schizophrenia. Front Psychiatry *4*, 169.

Anticevic, A., Cole, M.W., Repovs, G., Murray, J.D., Brumbaugh, M.S., Winkler, A.M., Savic, A., Krystal, J.H., Pearlson, G.D., and Glahn, D.C. (2013b). Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. Cereb. Cortex. Published online July 3, 2013. http://dx.doi.org/10.1093/cercor/bht165.

Arnsten, A.F., Paspalas, C.D., Gamo, N.J., Yang, Y., and Wang, M. (2010). Dynamic Network Connectivity: A new form of neuroplasticity. Trends Cogn. Sci. *14*, 365–375.

Baddeley, A. (2012). Working memory: theories, models, and controversies. Annu. Rev. Psychol. *63*, 1–29.

Balleine, B.W., and Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology *37*, 407–419.

Barch, D.M., and Ceaser, A. (2012). Cognition in schizophrenia: core psychological and neural mechanisms. Trends Cogn. Sci. *16*, 27–34.

Belforte, J.E., Zsiros, V., Sklar, E.R., Jiang, Z., Yu, G., Li, Y., Quinlan, E.M., and Nakazawa, K. (2010). Postnatal NMDA receptor ablation in corticolimbic interneurons confers schizophrenia-like phenotypes. Nat. Neurosci. *13*, 76–83.

Belujon, P., and Grace, A.A. (2008). Critical role of the prefrontal cortex in the regulation of hippocampus-accumbens information flow. J. Neurosci. *28*, 9797–9805.

Botvinick, M.M., Niv, Y., and Barto, A.C. (2009). Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. Cognition *113*, 262–280.

Brandon, N.J., Millar, J.K., Korth, C., Sive, H., Singh, K.K., and Sawa, A. (2009). Understanding the role of DISC1 in psychiatric disease and during normal development. J. Neurosci. *29*, 12768–12775.

Brodersen, K.H., Deserno, L., Schlagenhauf, F., Lin, Z., Penny, W.D., Buhmann, J.M., and Stephan, K.E. (2014). Dissecting psychiatric spectrum disorders by generative embedding. Neuroimage Clin *4*, 98–111.

Brunel, N., and Wang, X.-J. (2001). Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. J. Comput. Neurosci. *11*, 63–85.

Buchanan, R.W., Keefe, R.S., Lieberman, J.A., Barch, D.M., Csernansky, J.G., Goff, D.C., Gold, J.M., Green, M.F., Jarskog, L.F., Javitt, D.C., et al. (2011). A randomized clinical trial of MK-0777 for the treatment of cognitive impairments in people with schizophrenia. Biol. Psychiatry *69*, 442–449.

Buckley, M.J., Mansouri, F.A., Hoda, H., Mahboubi, M., Browning, P.G., Kwok, S.C., Phillips, A., and Tanaka, K. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science *325*, 52–58.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. *10*, 186–198.

Buzsáki, G., and Wang, X.-J. (2012). Mechanisms of gamma oscillations. Annu. Rev. Neurosci. *35*, 203–225.

Carandini, M. (2012). From circuits to behavior: a bridge too far? Nat. Neurosci. *15*, 507–509.

Carter, E., and Wang, X.-J. (2007). Cannabinoid-mediated disinhibition and working memory: dynamical interplay of multiple feedback mechanisms in a continuous attractor model of prefrontal cortex. Cereb. Cortex *17* (*Suppl 1*), i16–i26.

Carter, C.S., Barch, D.M., Buchanan, R.W., Bullmore, E., Krystal, J.H., Cohen, J., Geyer, M., Green, M., Nuechterlein, K.H., Robbins, T., et al. (2008). Identifying cognitive mechanisms targeted for treatment development in schizophrenia: an overview of the first meeting of the Cognitive Neuroscience Treatment Research to Improve Cognition in Schizophrenia Initiative. Biol. Psychiatry *64*, 4–10.

Cohen, J.D., and Servan-Schreiber, D. (1992). Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol. Rev. *99*, 45–77.

Collins, A., Brown, J., Gold, J., Waltz, J., and Frank, M. (2014). Working memory contributions to reinforcement learning impairments in schizophrenia. J. Neurosci. *34*, 13747–13756.

Compte, A., Brunel, N., Goldman-Rakic, P.S., and Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. Cereb. Cortex *10*, 910–923.

Constantinidis, C., and Wang, X.-J. (2004). A neural circuit basis for spatial working memory. Neuroscientist *10*, 553–565.

Corlett, P.R., Murray, G.K., Honey, G.D., Aitken, M.R., Shanks, D.R., Robbins, T.W., Bullmore, E.T., Dickinson, A., and Fletcher, P.C. (2007). Disrupted prediction-error signal in psychosis: evidence for an associative account of delusions. Brain *130*, 2387–2400.

Corlett, P.R., Taylor, J.R., Wang, X.J., Fletcher, P.C., and Krystal, J.H. (2010). Toward a neurobiology of delusions. Prog. Neurobiol. *92*, 345–369.

Courchesne, E., Mouton, P.R., Calhoun, M.E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M.J., Barnes, C.C., and Pierce, K. (2011). Neuron number and size in prefrontal cortex of children with autism. JAMA *306*, 2001–2010.

Coyle, J.T., Tsai, G., and Goff, D. (2003). Converging evidence of NMDA receptor hypofunction in the pathophysiology of schizophrenia. Ann. N Y Acad. Sci. *1003*, 318–327.

D'Esposito, M. (2007). From cognitive to neural models of working memory. Philos. Trans. R. Soc. Lond. B Biol. Sci. *362*, 761–772.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron *69*, 1204–1215.

Dayan, P., and Abbott, L.F. (2001). Theoretical Neuroscience. (Cambridge, MA: MIT Press).

Deco, G., Scarano, L., and Soto-Faraco, S. (2007). Weber's law in decision making: integrating behavioral data in humans with a neurophysiological model. J. Neurosci. *27*, 11192–11200.

Deco, G., Rolls, E.T., and Romo, R. (2009). Stochastic dynamics as a principle of brain function. Prog. Neurobiol. *88*, 1–16.

Dezfouli, A., Piray, P., Keramati, M.M., Ekhtiari, H., Lucas, C., and Mokri, A. (2009). A neurocomputational model for cocaine addiction. Neural Comput. *21*, 2869–2893.

Ding, L., and Gold, J.I. (2013). The basal ganglia's contributions to perceptual decision making. Neuron *79*, 640–649.

Dolan, R.J., and Dayan, P. (2013). Goals and habits in the brain. Neuron *80*, 312–325.

Douglas, R.J., and Martin, K.A.C. (2004). Neuronal circuits of the neocortex. Annu. Rev. Neurosci. *27*, 419–451.

Douglas, R.J., Koch, C., Mahowald, M., Martin, K.A., and Suarez, H.H. (1995). Recurrent excitation in neocortical circuits. Science *269*, 981–985.

Durstewitz, D., and Seamans, J.K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. Biol. Psychiatry *64*, 739–749.

Durstewitz, D., Seamans, J.K., and Sejnowski, T.J. (2000). Dopamine-mediated stabilization of delay-period activity in a network model of prefrontal cortex. J. Neurophysiol. *83*, 1733–1750.

Engel, T.A., and Wang, X.-J. (2011). Same or different? A neural circuit mechanism of similarity-based pattern match decision making. J. Neurosci. *31*, 6982–6996.

Fair, D.A., Bathula, D., Nikolas, M.A., and Nigg, J.T. (2012). Distinct neuropsychological subgroups in typically developing youth inform heterogeneity in children with ADHD. Proc. Natl. Acad. Sci. USA *109*, 6769–6774.

Freedman, R., Lewis, D.A., Michels, R., Pine, D.S., Schultz, S.K., Tamminga, C.A., Gabbard, G.O., Gau, S.S., Javitt, D.C., Oquendo, M.A., et al. (2013). The initial field trials of DSM-5: new blooms and old thorns. Am. J. Psychiatry *170*, 1–5.

Friston, K.J., Harrison, L., and Penny, W. (2003). Dynamic causal modelling. Neuroimage *19*, 1273–1302.

Friston, K.J., Stephan, K.E., Montague, R., and Dolan, R.J. (2014). Computational psychiatry: the brain as a phantastic organ. Lancet Psychiatry *1*, 148–158.

Funahashi, S., Bruce, C.J., and Goldman-Rakic, P.S. (1989). Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. *61*, 331–349.

Furman, M., and Wang, X.-J. (2008). Similarity effect and optimal control of multiple-choice decision making. Neuron *60*, 1153–1168.

Fuster, J.M. (2008). The Prefrontal Cortex, Fourth Edition. (New York: Academic Press).

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron *66*, 585–595.

Glimcher, P.W. (2003). Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics. (Cambridge, MA: MIT Press).

Glimcher, P.W., and Fehr, C.F. (2013). Neuroeconomics: Decision Making and the Brain, Second Edition. (London: Academic Press).

Goff, D.C. (2014). Bitopertin: the good news and bad news. JAMA Psychiatry *71*, 621–622.

Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. Annu. Rev. Neurosci. *30*, 535–574.

Goldman-Rakic, P.S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In Handbook of Physiology – The Nervous System V, F. Plum and V. Mountcastle, eds. (Bethesda, Maryland: American Physiological Society), pp. 373–417.

Goldman-Rakic, P.S. (1994). Working memory dysfunction in schizophrenia. J. Neuropsychiatry Clin. Neurosci. *6*, 348–357.

Goldman-Rakic, P.S. (1995). Cellular basis of working memory. Neuron *14*, 477–485.

Gottesman, I.I., and Gould, T.D. (2003). The endophenotype concept in psychiatry: etymology and strategic intentions. Am. J. Psychiatry *160*, 636–645.

Gruber, A.J., Calhoon, G.G., Shusterman, I., Schoenbaum, G., Roesch, M.R., and O'Donnell, P. (2010). More is less: a disinhibited prefrontal cortex impairs cognitive flexibility. J. Neurosci. *30*, 17102–17110.

Hansel, D., and Mato, G. (2013). Short-term plasticity explains irregular persistent activity in working memory tasks. J. Neurosci. *33*, 133–149.

Higley, M.J. (2014). Localized GABAergic inhibition of dendritic Ca(2+) signalling. Nat. Rev. Neurosci. *15*, 567–572.

Homayoun, H., and Moghaddam, B. (2007). NMDA receptor hypofunction produces opposite effects on prefrontal cortex interneurons and pyramidal neurons. J. Neurosci. *27*, 11496–11500.

Hunt, L.T., Kolling, N., Soltani, A., Woolrich, M.W., Rushworth, M.F., and Behrens, T.E. (2012). Mechanisms underlying cortical activity during value-guided choice. Nat. Neurosci. *15*, 470–476, S1–S3.

Huys, Q.J., Pizzagalli, D.A., Bogdan, R., and Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. Biol Mood Anxiety Disord *3*, 12.

Insel, T.R. (2010). Rethinking schizophrenia. Nature *468*, 187–193.

Insel, T.R. (2014). The NIMH Research Domain Criteria (RDoC) Project: precision medicine for psychiatry. Am. J. Psychiatry *171*, 395–397.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry *167*, 748–751.

Johansen, J.P., Cain, C.K., Ostroff, L.E., and LeDoux, J.E. (2011). Molecular mechanisms of fear learning and memory. Cell *147*, 509–524.

Josselyn, S.A. (2010). Continuing the search for the engram: examining the mechanism of fear memories. J. Psychiatry Neurosci. *35*, 221–228.

Kahneman, D. (2011). Thinking, Fast and Slow. (New York: Farrar, Straus and Giroux).

Kepecs, A., and Fishell, G. (2014). Interneuron cell types are fit to function. Nature *505*, 318–326.

Kilpatrick, Z.P., Ermentrout, B., and Doiron, B. (2013). Optimizing working memory with heterogeneity of recurrent cortical excitation. J. Neurosci. *33*, 18999–19011.

Krueger, R.F. (1999). The structure of common mental disorders. Arch. Gen. Psychiatry *56*, 921–926.

Krystal, J.H., and State, M.W. (2014). Psychiatric disorders: diagnosis to therapy. Cell *157*, 201–214.

Krystal, J.H., Karper, L.P., Seibyl, J.P., Freeman, G.K., Delaney, R., Bremner, J.D., Heninger, G.R., Bowers, M.B., Jr., and Charney, D.S. (1994). Subanesthetic effects of the noncompetitive NMDA antagonist, ketamine, in humans. Psychotomimetic, perceptual, cognitive, and neuroendocrine responses. Arch. Gen. Psychiatry *51*, 199–214.

Kurth-Nelson, Z., and Redish, A.D. (2011). Modeling decision-making systems in addiction. In Computational Neuroscience of Drug Addiction, B. Gutkin and S.H. Ahmed, eds. (Springer Publishing), pp. 163–188.

Lapish, C.C., Durstewitz, D., Chandler, L.J., and Seamans, J.K. (2008). Successful choice behavior is associated with distinct and coherent network states in anterior cingulate cortex. Proc. Natl. Acad. Sci. USA *105*, 11963–11968.

Lee, D. (2013). Decision making: from neuroscience to psychiatry. Neuron *78*, 233–248.

Lee, J., and Park, S. (2005). Working memory impairments in schizophrenia: a meta-analysis. J. Abnorm. Psychol. *114*, 599–611.

Lewis, D.A., and Gonzalez-Burgos, G. (2006). Pathophysiologically based treatment interventions in schizophrenia. Nat. Med. *12*, 1016–1022.

Lewis, D.A., Hashimoto, T., and Volk, D.W. (2005). Cortical inhibitory neurons and schizophrenia. Nat. Rev. Neurosci. *6*, 312–324.

Lewis, D.A., Curley, A.A., Glausier, J.R., and Volk, D.W. (2012). Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. Trends Neurosci. *35*, 57–67.

Lisman, J.E., Coyle, J.T., Green, R.W., Javitt, D.C., Benes, F.M., Heckers, S., and Grace, A.A. (2008). Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. Trends Neurosci. *31*, 234–242.

Lo, C.C., and Wang, X.-J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nat. Neurosci. *9*, 956–963.

Lo, C.C., Boucher, L., Paré, M., Schall, J.D., and Wang, X.-J. (2009). Proactive inhibitory control and attractor dynamics in countermanding action: a spiking neural circuit model. J. Neurosci. *29*, 9059–9071.

Lucantonio, F., Stalnaker, T.A., Shaham, Y., Niv, Y., and Schoenbaum, G. (2012). The impact of orbitofrontal dysfunction on cocaine addiction. Nat. Neurosci. *15*, 358–366.

Luck, S.J., and Gold, J.M. (2008). The construct of attention in schizophrenia. Biol. Psychiatry *64*, 34–39.

Machens, C.K., Romo, R., and Brody, C.D. (2005). Flexible control of mutual inhibition: a neural model of two-interval discrimination. Science *307*, 1121–1124.

Maia, T.V., and Frank, M.J. (2011). From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. *14*, 154–162.

Markov, N.T., Ercsey-Ravasz, M., Van Essen, D.C., Knoblauch, K., Toroczkai, Z., and Kennedy, H. (2013). Cortical high-density counterstream architectures. Science *342*, 1238406.

Mesulam, M.-M. (2000). Principles of Behavioral and Cognitive Neurology, Second Edition. (New York: Oxford University Press).

Millan, M.J., Agid, Y., Brüne, M., Bullmore, E.T., Carter, C.S., Clayton, N.S., Connor, R., Davis, S., Deakin, B., DeRubeis, R.J., et al. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. Nat. Rev. Drug Discov. *11*, 141–168.

Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. *24*, 167–202.

Miller, P., and Wang, X.-J. (2006). Inhibitory control by an integral feedback signal in prefrontal cortex: a model of discrimination between sequential stimuli. Proc. Natl. Acad. Sci. USA *103*, 201–206.

Moghaddam, B., and Krystal, J.H. (2012). Capturing the angel in ‘‘angel dust’’: twenty years of translational neuroscience studies of NMDA receptor antagonists in animals and humans. Schizophr. Bull. *38*, 942–949.

Montague, P.R., Dayan, P., and Sejnowski, T.J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. *16*, 1936–1947.

Montague, P.R., Dolan, R.J., Friston, K.J., and Dayan, P. (2012). Computational psychiatry. Trends Cogn. Sci. *16*, 72–80.

Moore, H., Jentsch, J.D., Ghajarnia, M., Geyer, M.A., and Grace, A.A. (2006). A neurobehavioral systems analysis of adult rats exposed to methylazoxymethanol acetate on E17: implications for the neuropathology of schizophrenia. Biol. Psychiatry *60*, 253–264.

Mota, N.B., Vasconcelos, N.A., Lemos, N., Pieretti, A.C., Kinouchi, O., Cecchi, G.A., Copelli, M., and Ribeiro, S. (2012). Speech graphs provide a quantitative measure of thought disorder in psychosis. PLoS ONE *7*, e34928.

Murray, J.D., Anticevic, A., Gancsos, M., Ichinose, M., Corlett, P.R., Krystal, J.H., and Wang, X.J. (2014). Linking microcircuit dysfunction to cognitive impairment: effects of disinhibition associated with schizophrenia in a cortical working memory model. Cereb. Cortex *24*, 859–872.

O’Reilly, R.C., and Frank, M.J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput. *18*, 283–328.

Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H.U., Jönsson, B., Jordanova, A., Musayev, A., Gustavsson, A., Gabilondo, A., and Maercker, A.CDBE2010 study group; European Brain Council (2012). The economic cost of brain disorders in Europe. Eur. J. Neurol. *19*, 155–162.

Panlilio, L.V., Thorndike, E.B., and Schindler, C.W. (2007). Blocking of conditioning to a cocaine-paired stimulus: testing the hypothesis that cocaine perpetually produces a signal of larger-than-expected reward. Pharmacol. Biochem. Behav. *86*, 774–777.

Park, S., and Holzman, P.S. (1992). Schizophrenics show spatial working memory deficits. Arch. Gen. Psychiatry *49*, 975–982.

Parker, A.J., and Newsome, W.T. (1998). Sense and the single neuron: probing the physiology of perception. Annu. Rev. Neurosci. *21*, 227–277.

Pereira, J., and Wang, X.-J. (2014). A trade-off between accuracy and flexibility in a working memory circuit endowed with slow feedback mechanisms. Cereb. Cortex in press.

Rangel, A., Camerer, C., and Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. *9*, 545–556.

Ratcliff, R. (1978). A theory of memory retrieval. Psychol. Rev. *85*, 59–108.

Redish, A.D. (2004). Addiction as a computational process gone awry. Science *306*, 1944–1947.

Redish, A.D., Jensen, S., Johnson, A., and Kurth-Nelson, Z. (2007). Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. Psychol. Rev. *114*, 784–805.

Renart, A., Song, P., and Wang, X.-J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. Neuron *38*, 473–485.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement. In Classical Conditioning II, A.H. Black and W.F. Prokasy, eds. (New York: Appleton-Century-Crofts), pp. 64–69.

Rigotti, M., Ben Dayan Rubin, D., Wang, X.-J., and Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: the importance of the diversity of neural responses. Front. Comput. Neurosci. *4*, 24.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585–590.

Robbins, T.W., Gillan, C.M., Smith, D.G., de Wit, S., and Ersche, K.D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. Trends Cogn. Sci. *16*, 81–91.

Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. *22*, 9475–9489.

Rolls, E.T., Loh, M., Deco, G., and Winterer, G. (2008). Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. Nat. Rev. Neurosci. *9*, 696–709.

Rotaru, D.C., Yoshino, H., Lewis, D.A., Ermentrout, G.B., and Gonzalez-Burgos, G. (2011). Glutamate receptor subtypes mediating synaptic activation of prefrontal cortex neurons: relevance for schizophrenia. J. Neurosci. *31*, 142–156.

Rubinov, M., and Bullmore, E. (2013). Fledgling pathoconnectomics of psychiatric disorders. Trends Cogn. Sci. *17*, 641–647.

Sakai, K. (2008). Task set and prefrontal cortex. Annu. Rev. Neurosci. *31*, 219–245.

Satterthwaite, T.D., Wolf, D.H., Erus, G., Ruparel, K., Elliott, M.A., Gennatas, E.D., Hopson, R., Jackson, C., Prabhakaran, K., Bilker, W.B., et al. (2013). Functional maturation of the executive system during adolescence. J. Neurosci. *33*, 16249–16261.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.

Seamans, J.K., Durstewitz, D., Christie, B.R., Stevens, C.F., and Sejnowski, T.J. (2001). Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. Proc. Natl. Acad. Sci. USA *98*, 301–306.

Sejnowski, T.J., Koch, C., and Churchland, P.S. (1988). Computational neuroscience. Science *241*, 1299–1306.

Sigala, N., Kusunoki, M., Nimmo-Smith, I., Gaffan, D., and Duncan, J. (2008). Hierarchical coding for sequential task events in the monkey prefrontal cortex. Proc. Natl. Acad. Sci. USA *105*, 11969–11974.

Simon, D.A., and Daw, N.D. (2011). Dual-system learning models and drugs of abuse. In Computational Neuroscience of Drug Addiction, B. Gutkin and S.H. Ahmed, eds. (New York: Springer Publishing), pp. 145–161.

Skoblenick, K., and Everling, S. (2012). NMDA antagonist ketamine reduces task selectivity in macaque dorsolateral prefrontal neurons and impairs performance of randomly interleaved prosaccades and antisaccades. J. Neurosci. *32*, 12018–12027.

Smith, P.L., and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. Trends Neurosci. *27*, 161–168.

Smoller, J.W., Ripke, S., Lee, P.H., Neale, B., Nurnberger, J.I., Santangelo, S., Sullivan, P.F., Perlis, R.H., Purcell, S.M., Fanous, A., et al.; Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379.

Soltani, A., and Wang, X.-J. (2006). A biophysically based neural model of matching law behavior: melioration by stochastic synapses. J. Neurosci. *26*, 3731–3744.

Spencer, K.M., Nestor, P.G., Perlmutter, R., Niznikiewicz, M.A., Klump, M.C., Frumin, M., Shenton, M.E., and McCarley, R.W. (2004). Neural synchrony indexes disordered perception and cognition in schizophrenia. Proc. Natl. Acad. Sci. USA *101*, 17288–17293.

Sporns, O. (2009). Networks of the Brain. (Cambridge, MA: MIT Press).

Stephan, K.E., Harrison, L.M., Kiebel, S.J., David, O., Penny, W.D., and Friston, K.J. (2007). Dynamic causal models of neural system dynamics:current state and future extensions. J. Biosci. *32*, 129–144.

Strauss, G.P., Frank, M.J., Waltz, J.A., Kasanova, Z., Herbener, E.S., and Gold, J.M. (2011). Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in schizophrenia. Biol. Psychiatry *69*, 424–431.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement Learning: An Introduction. (Cambridge, MA: MIT Press).

Szczepanski, S.M., and Knight, R.T. (2014). Insights into human behavior from lesions to the prefrontal cortex. Neuron *83*, 1002–1018.

Voon, V., Derbyshire, K., Rück, C., Irvine, M.A., Worbe, Y., Enander, J., Schreiber, L.R., Gillan, C., Fineberg, N.A., Sahakian, B.J., et al. (2014). Disorders of compulsivity: a common bias towards learning habits. Mol. Psychiatry in press.

Vos, T., Flaxman, A.D., Naghavi, M., Lozano, R., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J.A., Abdalla, S., Aboyans, V., et al. (2012). Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet *380*, 2163–2196.

Vukadinovic, Z. (2011). Sleep abnormalities in schizophrenia may suggest impaired trans-thalamic cortico-cortical communication: towards a dynamic model of the illness. Eur. J. Neurosci. *34*, 1031–1039.

Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956.

Wang, X.-J. (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. J. Neurosci. *19*, 9587–9603.

Wang, X.-J. (2001). Synaptic reverberation underlying mnemonic persistent activity. Trends Neurosci. *24*, 455–463.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron *36*, 955–968.

Wang, X.-J. (2006). Toward a prefrontal microcircuit model for cognitive deficits in schizophrenia. Pharmacopsychiatry *39*, S80–S87.

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. Neuron *60*, 215–234.

Wang, X.-J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. Physiol. Rev. *90*, 1195–1268.

Wang, X.-J. (2013). The prefrontal cortex as a quintessential 'cognitive-type' neural circuit: working memory and decision making. In Principles of Frontal Lobe Function, Second Edition, D.T. Stuss and R.T. Knight, eds. (New York: Cambridge University Press), pp. 226–248.

Wang, H.X., and Gao, W.J. (2009). Cell type-specific development of NMDA receptors in the interneurons of rat prefrontal cortex. Neuropsychopharmacology *34*, 2028–2040.

Wang, X.-J., Tegnér, J., Constantinidis, C., and Goldman-Rakic, P.S. (2004). Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. Proc. Natl. Acad. Sci. USA *101*, 1368–1373.

Wang, H., Stradtman, G.G., 3rd, Wang, X.-J., and Gao, W.J. (2008). A specialized NMDA receptor function in layer 5 recurrent microcircuitry of the adult rat prefrontal cortex. Proc. Natl. Acad. Sci. USA *105*, 16791–16796.

Wang, M., Yang, Y., Wang, C.J., Gamo, N.J., Jin, L.E., Mazer, J.A., Morrison, J.H., Wang, X.-J., and Arnsten, A.F. (2013). NMDA receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. Neuron *77*, 736–749.

Wei, Z., Wang, X.-J., and Wang, D.H. (2012). From distributed resources to limited slots in multiple-item working memory: a spiking network model with normalization. J. Neurosci. *32*, 11228–11240.

Wiecki, T.V., Poland, J.S., and Frank, M.J. (2014). Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. Clinical Psychological Sciences.. http://ski.cog.brown.edu/papers/wieckietal_comp_psych_rev_appendix.pdf

Wittchen, H.U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., et al. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. Eur. Neuropsychopharmacol. *21*, 655–679.

Wong, K.F., and Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. *26*, 1314–1328.

Yang, G.J., Murray, J.D., Repovs, G., Cole, M.W., Savic, A., Glasser, M.F., Pittenger, C., Krystal, J.H., Wang, X.J., Pearlson, G.D., et al. (2014). Altered global brain signal in schizophrenia. Proc. Natl. Acad. Sci. USA *111*, 7438–7443.