17

From Psychiatry to Computation and Back Again

A. David Redish and Joshua A. Gordon

In the opening chapters of this volume, we outlined a series of challenges facing psychiatry, as well as a description of its various promises, and suggested that taking a computational perspective could potentially illuminate a way forward. In this concluding chapter, we revisit these challenges and promises, in the context of what transpired at this Ernst Strüngmann Forum, to highlight the connections between the various themes raised. In particular, we will bring out the points of agreement and disagreement between the discussion groups and the chapters that arose from those discussions. We conclude with a description of the efforts, current and ongoing, to bring the potential synergy between psychiatry and computational neuroscience emphasized in this volume to a reality in the scientific and clinical arenas.

The Challenges of Psychiatry

The principal task for psychiatry is clear: using what we know of how the mind arises from interactions between the physical brain and its environmental and social milieu, how do we define and treat psychiatric disorders? In the opening chapter to this book, we identified three challenges that psychiatry currently faces which, if we could address, would go a long way toward that goal of defining and treating psychiatric disorders:

1. We need a *diagnostic nosology* that is better suited to access the current knowledge in psychology and neuroscience.
2. We need new *biomarkers* capable of assisting with diagnosis and prediction.
3. We need to develop improved *treatments*.

Furthermore, we emphasized that the path from genetics to behavior is complex and nonlinear, and that it depends on neural circuits. Finally, we pointed out that the end goal was personalized medicine, uniquely identified to be what was best for a specific patient.

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

Many of these challenges were echoed and expanded throughout this volume. However, several complexities were also delineated which make these challenges particularly difficult. For example, psychiatric disorders are *heterogeneous*, both at the *observational level*—each person has a very individual reaction to their neuropsychiatric, environmental, and social situation (Totah et al., this volume)—and the *etiological level*—many of the current psychiatric diagnoses are actually observations built around symptoms, not causal entities in themselves (Totah et al. and Flagel et al., this volume). As discussed by MacDonald and colleagues in Chapter 9, psychiatric dysfunction is *multisourced* (also known as *multifinal*, i.e., a single cause can have divergent outcomes) and *multipotential* (also known as *equifinal*, i.e., multiple causes can lead to observationally similar outcomes). Thus, an important fourth challenge relates to how one can develop a nosology of a system that is marked by a kaleidoscope of causes, each of which can be expressed as a pleiotropy of symptoms.

The *dynamics* of psychiatric disease add an additional layer of complexity. As noted by Totah et al. (see also Flagel et al., Barch, and Krystal et al., this volume), dysfunction proceeds through phases. Thus, patients can present quite differently at the various phases. There are several elements to this complexity. The first and most obvious is that diseases have a time course (episodic, chronic, or progressive) which helps define them. Although psychiatric syndromes often have canonical time courses associated with them, here too there is tremendous heterogeneity within the categories (Totah et al., Barch, and Krystal et al., this volume). Next, one must consider dynamic differences that occur because the patient is developing: a dysfunction in childhood may well manifest differently in adolescence or adulthood (Rutter et al. 2006; Totah et al., this volume). Finally, dysfunction itself can evolve over time, because the dysfunction itself may be progressing, because processes in the brain are attempting to compensate for the dysfunction (Krystal et al., this volume), or as a result of interactions with the environment (Totah et al., this volume; Borsboom et al. 2011; Borsboom and Cramer 2013). This reveals a fifth challenge: It is necessary to take into account the dynamics of illness while remembering that the key to treatment is to change the patient's trajectory (Flagel et al. and Friston, this volume).

The Promise of Computation

The computational perspective uses formal methods to relate how processes at one level can explain processing at other levels (e.g., how information processing in neural circuits can drive mental processes and behavior) and, in particular, how changes at one level can explain effects at other levels (e.g., how genetic differences can change the function of neural circuits) (Chapter 2). Computational neuroscience provides a diverse toolkit of models and theories

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

that can be used to provide that explanatory power (Kurth-Nelson et al., this volume). The most appropriate theory or model to use depends on the levels involved in one's questions.

The computational perspective has three effects which we suggest hold promise to address these challenges. First, the necessary *formalism* of computation forces one to be more complete and often reveals obscure consequences. All four discussion groups found themselves independently using a similar formalism to link causal dysfunction with observational effects that will have large implications for nosology (see Totah et al., Kurth-Nelson et al., Moran et al., and Flagel et al., this volume, and discussion below).

Second, because computation in neural circuits is fundamentally about *information processing* and how that information processing drives behavior, the questions one begins to ask about psychiatric dysfunction changes. For instance, when addressing surprising actions—such as continued use of addictive drugs (Redish 2004; Redish et al. 2008; Moran et al., this volume), misstated logic in semantic dementia (McClelland and Rogers 2003; Moran et al., this volume), unreasonable actions taken in schizophrenia (see chapters by Barch and Krystal et al., this volume), or a lack of action in depression (Huys, this volume)—the computational perspective changes the question from merely identifying *what* the subject is doing differently to identifying *how* the subject is recognizing information and processing that information differently. In particular, physical changes in neural circuits can have profound effects on how information is stored and processed in that neural circuit.

Third, this suggests that *treatment* can be aimed at (a) repairing the physical dysfunction in the neural circuit, (b) changing the environmental or social milieu in which that dysfunction occurs, or (c) changing other neural circuits to accommodate or replace the function of the dysfunctional circuit.

Synthesis from the Four Groups

As noted in the Chapters 1 and 2, and can be seen in this book's structure, Forum participants were divided into four working groups, each of which was tasked with a key topic relating to the question of how the computational perspective changes psychiatry, both in theory and practice.

Totah et al.: The Complexity and Heterogeneity of Psychiatry Disorders

This discussion group was tasked with examining what it was that computation needed to address. A central theme of this chapter is that complexity and heterogeneity are not noise to be abstracted away, but rather critical factors that need to be included in any computational theory. Using three case studies, they point out the variability from patient to patient and exhort us not to forget that patients are individuals with specific life stories. They recommend embracing

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

this heterogeneity. Moreover, they remind us that psychiatric dysfunctions change over time, whether through development, through progression of the dysfunction itself, or through interactions with the environment.

An issue raised by the accompanying chapter (Barch, this volume) is that a single dysfunction ("something is going on with dopamine in schizophrenia") can have many consequences throughout neural function, making the relationship between biological dysfunction and psychiatric observations even more complex. An important factor raised in this section is that we need to find a way to connect the biological and environmental factors that are dysfunctional to the observed psychiatric categories. It is suggested that this connection is going to be both multisourced and multipotential.

Kurth-Nelson et al.: Computational Approaches

The next group was tasked with examining what computational models were available and how they could be applied. Their discussion raised the very important issue that there is not a single computational model. Computational models and the computational perspective are very broad and wide ranging; however, they have in common the ability to link levels, particularly in non-intuitive and complex ways. Importantly, not all computational models have to be mechanistic; they can formally describe interactions between processes without specifying mechanism. Furthermore, the group pointed out that our goal in judging the success of computational approaches should not be to replicate current diagnoses of psychiatry. Success needs to be measured against more fundamental outcomes, reminding us that the goal is treatment and improvement in patient prognoses.

In an accompanying chapter, Frank (this volume) reviewed computational neuroscience approaches across specific levels and suggested that the computational framework characterizes mental illness in terms of difficulties in balancing trade-offs. Here again, we need to think of psychiatry in terms of how an individual with a specific biological brain interacts with the environmental and social milieu.

Mathys (this volume) reviewed a specific computational framework that was explicitly designed to handle multisourced and multipotential connections: Bayesian inference (Pearl 1988, 2009b; Jaynes 2003). This framework can formalize the relationship between a dimensional model of underlying potential causes and the pleiotropy of observed behaviors. Importantly, this framework allows one to formalize the inverse logic that allows reasoning from observations to potential dimensional causes. The key to this framework is to see both causes and observations as probabilistic rather than certainties. In setting out the principles of Bayesian inference, Mathys introduced a key method to develop frameworks for integrating computation and psychiatry in novel ways that was used by both Flagel et al. and Moran et al. (this volume).

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

Flagel et al.: A New Framework

This group was tasked with identifying how the computational perspective can be used to improve nosology. Supporting chapters by First, MacDonald et al. and Mathys (this volume) framed the issues for the group. First (this volume) reminds us of the importance of nosology, noting that the original intention of the Diagnostic and Statistical Manual of Mental Disorders (DSM) and International Classification of Diseases (ICD) was to maintain an atheoretical categorical perspective. Of course, no perspective is atheoretical, and the categories in the DSM and ICD have become the definitions of named syndromes. The new, highly theoretical Research Domain Criteria (RDoC) project, implemented by the U.S. National Institute of Mental Health (Cuthbert and Insel 2010; Insel et al. 2010), has taken a wholly new dimensional approach. It has been extremely difficult, however, to meld the dimensional descriptions in RDoC with the categorical descriptions of DSM and ICD (First, this volume). This is partially due to issues raised by MacDonald et al. (this volume)—the relationship between causes and consequences are both multifinal (multisourced) and equifinal (multipotential). They suggest that causal networks such as those used in reliability engineering could be a way to link these multisourced and multipotential relationships. These relationships are directly formalizable in the Bayesian inference perspective raised by Mathys (this volume).

This discussion group came to the conclusion that if one takes the novel perspective that *psychiatric diagnoses are observations*, not causes, then a natural nosology appears in which there are biological and environmental causes that probabilistically lead to observations (Flagel et al., this volume). The biological and environmental causes can be highly dimensional and highly theoretical (like RDoC). Psychiatric diagnoses remain an important part of clinical practice, but they become cues to the underlying hypothesized causes rather than syndromes themselves. Importantly, observations can also include other measurements, such as biological measurements, clinical instruments, or cognitive tasks. Furthermore, because Bayesian inference allows reasoning in both directions (from observations to hypothesized causes and from probability distributions over causes to predicted observations), prognoses can be thought of as yet another observation: one that unfolds over future time.

In his accompanying chapter, Friston (this volume) uses a computational simulation to show how this logic might work. Importantly, as noted by Totah et al. (this volume), a patient's experience is a trajectory, and the goal must be to shift that trajectory. This means that the prognosis needs to be seen as a probability distribution across trajectories, and that the goal of treatment is to change that probability distribution. Therefore, this framework naturally includes the dynamic nature of psychiatric illness as well as the heterogeneity inherent in these dynamics, but with the capacity to formally analyze and quantify these dynamics, as well as the effects of treatment on disease course.

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

Moran et al.: Candidate Examples

The fourth group was tasked with finding targets, dysfunctions, syndromes, and disorders that could be used as canonical examples of the computational perspective in action. Moran et al. sketched out several examples of specific models that have had an impact on the field, including dopamine models in schizophrenia and addiction, computational analyses of neuroimaging and EEG measurements, and a model of treatment in changing allostasis along the amygdala-HPA axis. They began, however, with a generative Bayesian inference model in which biological parameters are linked to symptoms, biomarkers, and diagnoses through computational parameters. This model is identical to the new nosology suggested by Flagel et al. (compare Figure 12.1 with Figure 10.3). In the general framework by Moran et al. (identified as a "generative model"), the biological parameters are the underlying structure (putative causes and hidden physiological states of Flagel et al.), the computational parameters are the theoretical structure (the latent constructs of Flagel et al.), and the symptoms, biomarkers, and diagnoses are the observations. Importantly, Moran et al. view this generative model as a trajectory that needs to be shifted by treatment. In the companion chapters, Montague, Paulus et al., Huys, and Krystal et al. (all this volume) lay out examples of these relationships, looking specifically at the relationship between computation and dysfunction in risk-sensitivity, value and addiction, depression, and schizophrenia.

Common Themes and New Breakthroughs

Remarkably, all four groups converged independently on a similar breakthrough: that the key to connecting the fundamental science with clinical practice is a multipotential and multisourced computational perspective which links psychiatric observations with underlying dysfunction, but which breaks the one-to-one assumptions currently underlying clinical practice. This key breakthrough can be summarized by a simple statement: DSM diagnoses are symptoms, not syndromes. This realization led us to propose a new system in which fundamental science identifies *neuropsychological processes* and *failure modes* within those processes. These processes can be computational theoretic models (e.g., reinforcement-learning algorithms; Montague, this volume) or psychological constructs, such as are used to define the RDoC matrix. These processes probabilistically produce *outcomes*, which can be either observations (e.g., scales on a psychiatric instrumental questionnaire), measurements (e.g., scores on a task), or DSM diagnoses. This new perspective unifies the fundamental science processes, such as RDoC, with psychiatric categorizations, such as used in the DSM or ICD-10.

Early discussions of computational models talked in terms of "vulnerabilities" or "failure modes" (Redish 2004; Huys 2007; Rangel et al. 2008; Redish

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

et al. 2008; Huys et al. 2015a), in which one derived observed dysfunction from specific errors in hypothesized underlying processes. However, as became clear in the discussions, the multisourced and multipotential nature of psychiatric dysfunction is going to depend on a more complex path from dysfunction to symptom (see Totah et al., MacDonald et al., Flagel et al., Krystal et al., this volume), taking into account genetic and environmental causes, heterogeneity, and dynamics.

Models of fundamental dysfunctions are inherently dimensional, based on errors in specific parameters and processes (e.g., RDoC), whereas models of clinical practice are inherently categorical (e.g., DSM, ICD). The relationship between fundamental science and clinical practice is more than a simple threshold on dimensionality. To accommodate this complexity, we propose a Bayesian causal model of failure modes in underlying fundamental neuropsychological processes leading to observations (diagnoses) with specific probabilities. This new system (exemplified by the figures in Flagel et al., this volume) opens up an entirely new perspective on psychiatry, providing a way to connect clinical observation with underlying neuropsychological causes. Moreover, this probabilistic framework has heterogeneity built into it: a given dysfunction in a fundamental process might lead to psychosis with some (non-zero) probability or mood instability with a different (non-zero) probability. Thus the same failure in the same neuropsychological process could lead to two different diagnoses from the perspective of the clinician. The framework also explains comorbidity: some fraction of people with that dysfunction may find themselves with both diagnoses.

Importantly, these Bayesian inferences can be applied to trajectories, both through the past and into the future. Bayesian inferences, taking into account past trajectories, allow the hypothesized dysfunction to explain pathophysiological processes and symptoms that change and develop over time. Bayesian inferences about future trajectories are *prognoses*. As pointed out by Flagel et al. (see also Friston, this volume), a prognosis is simply a predicted observation about the future trajectory.

This new nosology solves the heterogeneity and comorbidity problems completely, while unifying RDoC (and other neuropsychological hypothesis-based schemes) with clinical categorizations (such as are used in DSM and ICD). This proposal unifies observations, clinical categorizations (DSM syndromes), measurements (e.g., clinical questionnaire instruments, performance on cognitive tasks), and prognoses as observations linked through Bayesian inference to a scientific hypothesis of latent constructs.

An important open question that we still must face is whether the hypothesized latent constructs are the correct taxonomy. Of course, science is always progressing by refining and replacing theoretical constructs. Aristotle's theory of gravity was replaced by Newton's theory, and Newton's equations were replaced by Einstein's. Current observations of galactic motion are hinting at the possibility of further refinement in Einstein's equations. Nevertheless, at

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

each stage, scientific theory was able to provide practical consequences. Do we have the right fundamental science dimensions and the right experimental tests to identify patient treatments? One of the major advantages of the Bayesian perspective is that it provides an *ongoing process* rather than a direct answer to this open question. Latent constructs which do not inform prognoses will be excluded mathematically because they are unpredictive, whereas those that provide for better description of the observations and better prognoses will be preserved.

The Way Forward

Establishing a Computational Nosology

Our breakthrough is a proposal: it describes a methodology for translational psychiatry, for a way to connect fundamental (basic) science research to clinical practice. Actually implementing this methodology, however, is going to require buy-in from the current stakeholders in these processes.

Clinicians might be daunted by the complexity of Bayesian inference, or the notion of carrying out such analyses for every patient mathematically. However, it would not be hard to implement the algorithm in a computer app or online website. Essentially, a clinician would enter a series of observations (e.g., diagnoses, clinical instrument test scores, cognitive task results, perhaps even genetic tests) and would receive a probability distribution over potential treatments, explanations, and outcomes.

This is actually not that different from what current practice is supposed to be. The clinician determines the presence or absence of symptoms through a clinical interview; consults the diagnostic "algorithms" in the DSM (which define categories from lists of symptoms), and determines the appropriate diagnoses. A treatment is then selected based on the likelihood of response for that diagnosis. In reality, of course, both diagnoses and treatments are selected based much more on experience than on an algorithm. This reality, however, is based on a lack of specific information, which if present would enable decisions to be made with greater confidence. In this sense, what we are proposing is an improved DSM: one that might be complicated enough to require a computer, but one which would be much more useful than current systems. It supplements the clinician's judgment by providing explicit probabilities over latent causes, treatments, diagnoses, and prognoses that the clinician can use to communicate with the patient and to help make decisions with the patient about treatment. This requires communicating probabilities to patients, but medical fields are already communicating probabilities for other complex diseases (e.g., cancer), in terms of the probability distribution of survival curves and future prognosis trajectories based on different treatments.

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

A final consideration for the clinician is the notion that the algorithmic recommendations are personalized and personalizeable. One inputs patient-specific information and the outputs are probabilities that a given treatment will work in a given way. Thus, there is not only room but rather a requirement for clinical judgment in adopting these recommendations.

Acceptance of the model will require cooperation from scientists as well. Fundamental (basic) scientists will determine the latent constructs (scientific hypotheses) that underlie neuropsychological function, including how those processes interact with each other as well as with environmental and sociological factors. Translational scientists will determine how observations and measurements predict (and are predicted by) those underlying constructs. A team of experts will convene and codify those relationships. Presumably that team of experts will meet regularly to update that codification. There is no reason not to expect that team of experts to have computational help to run complex calculations measuring specific relationships. Once the experts have codified the relationships, they can be compiled into an algorithm that can be accessed by any clinician anywhere, even without the level of expertise needed to codify the relationships. This algorithm would be updated periodically, not unlike the iterative editions of the DSM, but more frequently (and with mathematical rigor applied to any changes, which would have to have documented efficacy).

Of course, as with any scientific endeavor, this is a cyclical process by which new fundamental science hypotheses are derived from clinical observations, which will lead to new discoveries, requiring new codifications, and (hopefully) improved patient outcomes (Figure 17.1).

Challenges and Promises

We began with a list of challenges (a new diagnostic nosology, new biomarkers, and improved treatments), complexities (the heterogeneity of patients, the dynamics through which psychiatric dysfunction progresses, that the path from genetics to behavior goes through neural circuits), and promises (the hope of personalized medicine). We believe that this Forum succeeded in addressing these issues far beyond what we could have hoped. The new perspective integrating dimensional science (RDoC, fundamental/basic hypotheses) with observations (DSM/ICD syndromes and categories, clinical instruments, tests, etc.) is explicitly a new diagnostic nosology that builds on the progress made over the last half-century in both the clinical and fundamental sciences. It provides a direct way to incorporate new biomarkers as they are discovered and provides a new way to identify the best treatments (as probabilities over prognoses). It directly addresses the complexities, including taking into account nonlinearities and the heterogeneity of patients. Because our proposed diagnostic nosology reflects trajectories in the past (history) and future (prognoses), it takes into account the dynamics of psychiatric dysfunction. Because it depends on latent constructs, one does not have to go directly from genetics to

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

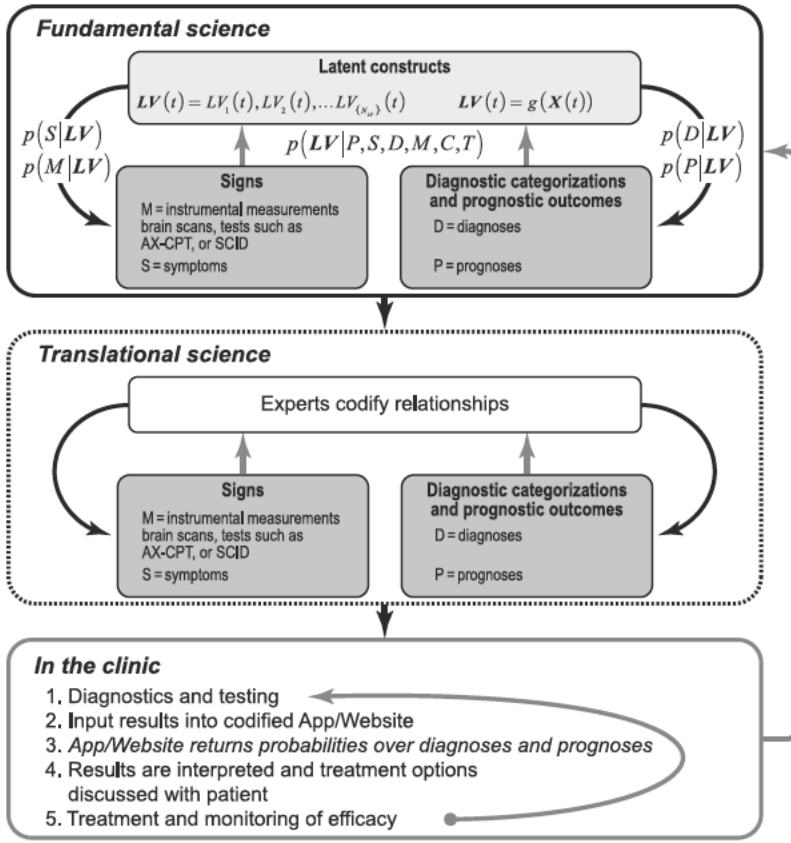

Figure 17.1 The proposed process to translate fundamental science to clinical practice. Fundamental (basic) scientists determine the most appropriate latent constructs and the relationship between those latent constructs and signs and outcomes. A committee of experts codifies these relationships so that one can integrate the two steps of Bayesian inference from signs and observations to diagnostic categorizations and prognostic outcomes, maintaining the probabilities. A clinician then uses this codified relationship to translate from signs (observations) and diagnoses to a probability distribution over diagnostic categorizations and prognoses. Treatments can be presented in how they would change the prognoses (probabilistically). Because continued observations, including the consequences of treatments, provide additional information to the probabilities of diagnoses and prognoses, the clinician will keep returning to the codification for refinement of diagnosis, treatment, and prognosis. Of course, this path from fundamental science to the clinic is a continual cycle of scientific progress and refinement, with fundamental science learning from continued clinical observations.

behavior or psychiatric categories; instead, one can go (probabilistically) from genetics to latent constructs and then (probabilistically) from latent constructs to behavior. Finally, each person will have a unique set of genetics, history, and symptoms, which probabilistically interact to produce a unique diagnosis and prognosis, directly giving us personalized medicine.

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

New Structures

The field of "computational psychiatry" is exploding with a new ongoing workshop group that has been formed, new departments forming, a new journal, and several books being developed, including this one. A community bringing together experienced psychiatrists, experienced computational neuroscientists, and newly trained students with expertise in both fields is appearing. This community is developing experimental tests which can be used to identify parameters in underlying dysfunctions, such as the two-step decision task capable of differentiating model-based and model-free decision-making processes, which predicts a broad spectrum of obsessive behaviors (Gillan et al. 2016). New studies suggest that these multidimensional (genetic and task-based) categorizations produce more reliable categorizations than traditional DSM categories (Clementz et al. 2016). More work is clearly needed to translate the tasks developed for measuring fundamental science dimensions (in both human and nonhuman models) into clinically relevant measures. We believe that the field is ready to have a direct effect on the practice of psychiatry.

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