11

Computational Nosology and Precision Psychiatry

A Proof of Concept

Karl J. Friston

Abstract

This chapter provides an illustrative treatment of psychiatric morbidity that offers an alternative to the standard nosological model in psychiatry. It considers what would happen if we treated diagnostic categories not as putative causes of signs and symptoms, but as diagnostic consequences of psychopathology and pathophysiology. This reconstitution (of the standard model) opens the door to a more natural formulation of how patients present and their likely response to therapeutic interventions. The chapter describes a model that generates symptoms, signs, and diagnostic outcomes from latent psychopathological states. In turn, psychopathology is caused by pathophysiological processes that are perturbed by (etiological) causes, such as predisposing factors, life events, and therapeutic interventions. The key advantages of this nosological formulation include: (a) the formal integration of diagnostic (e.g., DSM) categories and latent psychopathological constructs (e.g., the dimensions of RDoC); (b) the provision of a hypothesis or model space that accommodates formal evidence-based hypothesis testing or model selection (using Bayesian model comparison); (c) the ability to predict therapeutic responses (using a posterior predictive density), as in precision medicine; and (d) a framework that allows one to test hypotheses about the interactions between pharmacological and psychotherapeutic interventions. These and other advantages are largely promissory at present: the purpose of this chapter is to show what might be possible, through the use of idealized simulations. These simulations can be regarded as a (conceptual) prospectus that motivates a computational nosology for psychiatry.

Introduction

One of the key issues addressed by our working group at this Forum (Flagel et al., this volume) was the status of psychiatric nosology and how it might

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.

be informed by advances in computational neurobiology (Redish and Johnson 2007; Montague et al. 2012; Wang and Krystal 2014). In brief, our starting point was the realization that diagnostic categories are not the causes of psychopathology—they are (diagnostic) consequences. Although rather obvious in hindsight, this was something of a revelation, largely because it disclosed the missing link between putative causes of psychiatric illness (e.g., genetic predisposition, environmental stressors, iatrogenic) and the consequences, as observed by clinicians (e.g., symptoms, signs and, crucially, diagnostic outcome). In what follows, I briefly rehearse the ideas—borrowed from computational neurobiology—that we hope might close this gap (for full discussion, see Flagel et al., this volume). This treatment provides a technical summary of our conclusions, using an illustrative (simulated) case study of nosology, diagnosis, and prognosis.

The principal contribution of a formal or computational approach to nosology rests on the notion of a *generative model*. A generative model generates consequences from causes—in our case, symptoms, signs, and diagnoses—from underlying psychopathology and pathophysiology. Generally, these models are state-space models that describe dynamics and trajectories in the space of latent (e.g., pathophysiological) states. These states are latent or hidden from direct observation and are only expressed in terms of measurable consequences, such as symptoms and signs. The utility of a generative model lies in the ability to infer latent states from observed outcomes and, possibly more importantly, assess the evidence for one model relative to others, given a set of measurements. This is known as (Bayesian) model comparison, where the evidence is simply the probability of any sequence of observations under a particular model (Stephan et al. 2009b). To assess the evidence for a particular model, it is necessary to fit the model to observed data, a procedure known as *model inversion*. This is because the mapping from causes to consequences is inverted, to infer from consequences to causes (e.g., inferring pathophysiology from symptoms). Furthermore, having inverted a model—by optimizing its parameters to maximize model evidence—one can then simulate or predict new outcomes in the future using something called the *posterior predictive density*. Heuristically, this is the technology behind weather forecasts, where the generative model is a detailed state-space model of meteorological dynamics (Young 2002). So can we conceive of an equivalent meteorology for psychiatry?

In the past decade, there have been considerable advances in using state-space models of distributed neuronal processes to understand (context-sensitive) connectivity and functional architectures in the brain. This is known as *dynamic causal modeling*, which accounts for a significant number of papers in the imaging neuroscience literature (Friston et al. 2003; Daunizeau et al. 2011a). In what follows, I apply exactly the same (computational neuroscience) principles to the problem of modeling the causes of nosological outcomes in psychiatry. Indeed, all the examples below use standard Bayesian model

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.

inversion schemes that are available in freely available academic software: the simulations described below can be reproduced by downloading the SPM software (http://www.fil.ion.ucl.ac.uk/spm/) and invoking the Matlab script *DEM_demo_ontology.m*. The dynamic causal modeling of psychopathology can, in principle, offer a number of advantages over the standard nosological model. As noted in the abstract and by Flagel et al. (this volume), these include: (a) the formal integration of diagnostic (e.g., DSM) categories and latent psychopathological constructs (e.g., the dimensions of RDoC) (Stephan and Mathys 2014), (b) the provision of a hypothesis or model space that accommodates formal evidence-based hypothesis testing (Krystal and State 2014), and (c) the ability to predict therapeutic responses (using a posterior predictive density). Crucially, by adopting a dynamic modeling approach, one can properly accommodate the personal history and trajectory of individual patients in determining (and predicting) the course of their illness.

Our overall approach to nosology (and its promissory advantages) may seem rather abstract and perhaps even grandiose. This chapter should therefore be taken as a prospectus for future discussions about nosology and the potential for individualized or precision psychiatry in the future. Its purpose is to illustrate what could be possible, in an idealized world, if we were able to use clinical data to optimize generative models of psychopathology. Whether or not this is possible with current data is an outstanding question. In short, this chapter offers a (mathematical) sketch of what a computational nosology could look like.

I begin by describing a formal model that generates symptomatic and diagnostic outcomes from latent (pathophysiological and psychopathological) causes. This model should not be taken too seriously; it is just used to illustrate the promise of such modeling initiatives and to show how a formal approach to nosology forces one to think carefully about the known and unknown variables in psychiatric processes and how they influence each other. In the next section, I consider the use of ratings of symptoms and signs (and diagnosis) to estimate or infer their latent causes. This is a necessary prelude for model comparison and is discussed briefly. Finally, prognosis and prediction are considered by using the generative model to predict the outcome of a (simulated) schizoaffective process and its response to treatment.

Generative Models for Psychiatric Morbidity

This section introduces the general form of generative (dynamic causal) models for psychiatric morbidity and a particular example that will be used to illustrate model inversion, model selection, and prediction in subsequent sections. As noted above, a generative model generates consequences from causes. The basic form assumed here starts with the (observable) causes of psychiatric illness, such as genetic or environmental predispositions, and therapeutic

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.

interventions. These factors induce pathophysiological states, such as aberrant dopamine receptor availability or glucocorticoid receptor function, to alter their trajectories (i.e., time courses or fluctuations) over a course of weeks to months. These latent pathophysiological states then determine psychopathology, cast in terms of latent cognitive, emotional, or behavioral function (e.g., low mood, psychomotor poverty, thought disorder). Psychopathological states correspond to the constructs underlying things like Research Domain Criteria (RDoC) (Kaufman et al. 2015) and clinical brain profiling (Peled 2009). Finally, psychopathological states generate measured symptoms using, for example, standardized instruments (e.g., PANSS [Kay 1990], Beck depression inventory, mini mental state) or diagnostic outcomes (e.g., schizophrenia, major affective disorder, schizoaffective disorder).

Note that in this setup, a diagnosis represents an outcome, provided by a clinician. In other words, symptoms, signs, and diagnosis have a common cause, where the diagnostic categorization provides a useful summary outcome that integrates aspects of psychopathology which may not be covered explicitly by standardized symptom ratings or particular signs (e.g., psychomotor poverty, EEG abnormalities, abnormal dexamethasone suppression).

This formulation of psychiatric nosology is largely common sense and reiterates what most people would understand about psychiatric disorders. However, can this understanding be articulated formally in a way that can be used to make quantitative predictions and test competing etiological hypotheses? This is where a formal nosology or generative model comes into play. The first step is to construct a *graphical model* of dependencies among the variables generating measurable outcomes. Figure 11.1 (left panel) shows the graphical model that summarizes the probabilistic dependencies among etiological causes $u(t)$, pathophysiological states $x(t)$, psychopathology $v(t)$, and, finally, symptoms and diagnosis $s(t)$, $\Delta(t)$. In this format, the variables in white circles correspond to latent states that are hidden from direct observation, while the observable outcomes are in the cyan circle.

Probabilistic dependencies are denoted by arrows that entail (time-invariant) parameters ($\theta = \theta^s, \theta^n, \theta^p$). This formulation clarifies the roles of different quantities and makes their interdependencies explicit. For example, a diagnostic classification at a particular time would be an outcome variable, whereas a patient's drug history would be an etiological cause that influenced pathophysiology. Having established the form of the graphical model, it is now necessary to specify the nature of the dependencies within and among latent variables. An example is provided in the right panel and can be described as follows.

A Generative (Dynamic Causal) Model

This example of a generative model is deliberately very simple and restricts itself to modeling a limited differential diagnosis that includes (a simulation

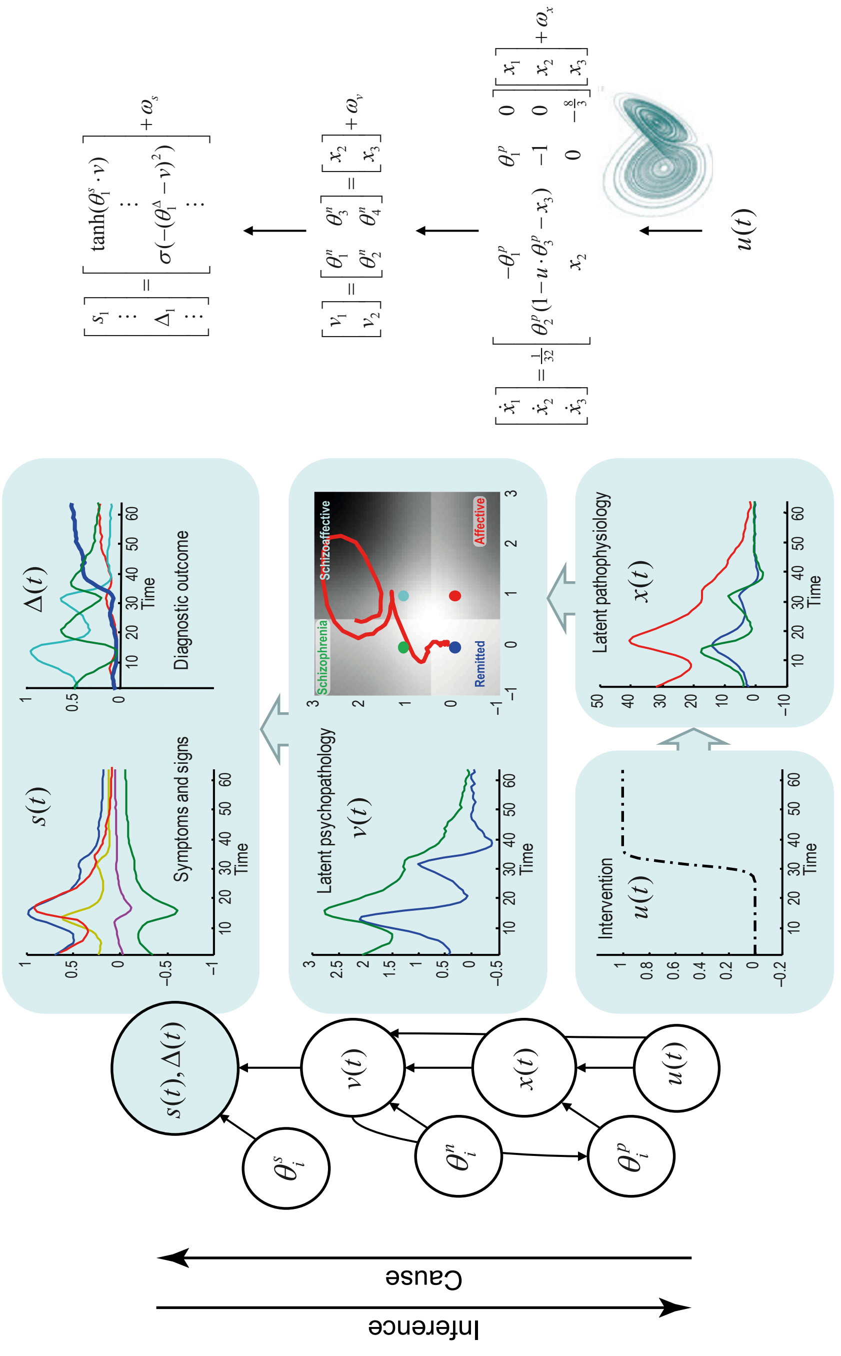
From "Computational Psychiatry: New Perspectives on Mental Illness,"
A. David Redish and Joshua A. Gordon, eds. 2016. Strüngmann Forum Reports, vol. 20,
series ed. J. Lupp. Cambridge, MA: MIT Press. ISBN 978-0-262-03542-2.

of) schizophrenia, affective disorder, schizoaffective disorder, and a remitted state. These diagnostic outcomes are accompanied by six symptom scores that have been normalized to lie between plus and minus one. Furthermore, only a limited number of exogenous causes and latent variables are considered: specifically, a single therapeutic cause perturbs the evolution of three physiological states, from which two psychopathological states are derived. At any one time, the location in the psychopathological state-space determines a symptom profile (through a linear mixture of psychopathological states that is passed through a sigmoid function). Diagnostic outcomes are, here, encoded by a probability profile over the differential diagnosis (i.e., the relative confidence a clinician places in the differential diagnoses). In the model, the probability of any one diagnosis corresponds to the relative proximity to a particular point in psychopathological state-space. In other words, as the disorder progresses, a trajectory is traced out in a two-dimensional psychopathological state-space, where, at any one time, the prevalent diagnosis is determined by the location to which the current state is closest. Technically, this has been modeled by a softmax function of diagnostic potential, where diagnostic potential is the (negative) Euclidean distance between the current state and the locations associated with each diagnostic category (encoded by θ_i^{Δ}, the color dots in Figure 11.1).

The trajectory of psychopathology is determined by the corresponding trajectory through a pathophysiological state-space which has its own dynamics. These are encoded by *equations of motion* or flow based, in this example, on a Lorenz attractor (Lorenz 1963). The Lorenz form is an arbitrary choice and could easily be replaced by other plausible equations: Moran et al. (this volume) provide an example based on normal form stochastic dynamics and indeed optimized on the basis of Bayesian model evidence (see below). Having said this, the Lorenz attractor provides a simple model of chaotic dynamics in the physical (Poland 1993) and biological (de Boer and Perelson 1991) sciences. Interestingly, it arose in the modeling of convection dynamics, which speaks to the analogy between the current modeling proposal and weather forecasting. In this setting, the ensuing dynamics can be regarded as a canonical form for nonlinear coupled processes that might underlie pathophysiology in psychosis. It has a canonical form because, as seen in Figure 11.1 (on the right), one can regard the parameters as specifying coupling coefficients or connections that mediate the influence of one physiological state on the others. Crucially, some of these connections are state dependent. This is important because it means one can model fluctuations in pathophysiology in terms of self-organized (chaotic) dynamics that have an underlying attracting set. In other words, we have a rough model of pathophysiological dynamics that summarize slow fluctuations in neuronal (or hormonal) states that show homoeostatic or allostatic tendencies (e.g., Leyton and Vezina 2014; Misiak et al. 2014; Oglodek et al. 2014; Pettorruso et al. 2014; see also Krystal et al., this volume).

The equations in Figure 11.1 all include random fluctuations. These fluctuations render the generative model a probabilistic statement about how various

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.



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 11.1 Schematic of a generative model for psychiatric morbidity. The model is shown in terms of a (probabilistic) graphical model on the left. In this format, the quantities in white circles correspond to random variables that include (unknown) parameters on the left and hidden or latent states on the right: etiological causes $u(t)$, pathophysiological states $x(t)$, psychopathology $v(t)$. The arrows denote conditional dependencies and describe the influences among latent variables that generate observations or outcomes in the cyan circle at the top. Here, the outcomes are clinical symptom scores and a differential diagnosis, measured as a probability distribution over diagnostic classifications. The only difference between these outcomes is that the diagnostic probabilities are constrained to the nonnegative and sum to one. The outcomes are generated as functions of psychopathological states which, themselves, are mixtures of pathophysiological states. Finally, the pathophysiological states are perturbed by inputs (like therapeutic interventions). An example of the form of the conditional dependencies is provided on the right, in terms of functions and random (Gaussian) fluctuations. In this example, a single therapeutic intervention enters the dynamics (or equations of motion) governing the evolution of pathophysiological states. Here, the therapeutic intervention changes the influence of the second physiological state on the first, where this coupling is itself state-dependent (and changes with the third state). These dynamics are based upon a Lorenz attractor (illustrated with the insert). The psychopathological states are generated as a linear mixture of the last two physiological states. In turn, the psychopathological variables are mapped to clinical outcomes through sigmoid functions (to generate symptom scores) and a softmax function of diagnostic potential (to generate a differential diagnosis). The diagnostic potential is based on the proximity of the psychopathological state to locations representing diagnostic categories. The middle panels illustrate a particular realization of this generative model over 64 time bins (i.e., weekly assessments). The lower panels show the therapeutic input (starting at 32 weeks) and the dynamic response of the three pathophysiological states. The ensuing psychopathology is shown as a function of time and as a trajectory in state-space in the left and right middle panels respectively. The state-space of psychopathology contains the locations associated with diagnostic categories (colored dots), which determine the diagnostic classification that tiles state-space (shaded gray regions). In this example, there are four diagnostic categories: schizophrenia, schizoaffective and affective disorder, and a state of remission. The symptoms and differential diagnosis, $s(t)$, $\Delta(t)$, generated by this trajectory are shown in the upper panels on the left and right respectively. In this and subsequent simulations, the initial (physiological) states were [8,10,32], the parameters for the symptom scores are sampled from a unit Gaussian distribution, while the remaining parameters are shown in Figure 11.4.

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.

variables influence each other. Here, the random fluctuations can be regarded as observation noise (when assessing symptoms) and random fluctuations or perturbations to psychological or physiological processing (e.g., life events or drug misuse). In this chapter, these random fluctuations are smooth processes with a Gaussian correlation function with a correlation length of half an assessment interval (i.e., a few days).

There is no pretence that any of these states map in a simple way onto physiological variables; rather, they stand in for mixtures of physiological variables that have relatively simple dynamics. The existence of mixtures is assured by technical theorems such as the center manifold theorem and the slaving principle in physics (Carr 1981; Haken 1983; Frank 2004; Davis 2006b). Basically, these theorems say that any set of coupled dynamical systems can always be described in terms of a small number of patterns (known variously as order parameters or eigenmodes), which change slowly relative to fast and noisy fluctuations about these patterns.

Crucially, this generative model has been constructed such that the therapeutic intervention changes the state-dependent coupling between the first and second pathophysiological states (in fluid dynamics, this control parameter is known as a *Rayleigh number* and reflects the degree of turbulent flow). This means we could regard this intervention as pharmacotherapy that changes the coupling between different neuronal (or hormonal) systems, e.g., an influence of an atypical antipsychotic (Hrdlicka and Dudova 2015) on dopaminergic and serotonin receptor function responsible for monoaminergic tone in the ventral striatum and serotoninergic projections from the amygdala to the paraventricular nucleus (Wieland et al. 2015; Muzerelle et al. 2016). Furthermore, I have introduced a parameter θ_3^p that determines the sensitivity to the intervention that may be important in determining a patient's responsiveness to therapy (Brennan 2014).

The middle panel of Figure 11.1 provides an illustration of how a patient might present over time under this particular model. Imagine we wanted to model (six) symptom scores and a probabilistic differential diagnosis over four diagnoses (schizophrenia, schizoaffective, affective, and remitted), when assessing an outpatient on a weekly basis for 64 weeks. A therapeutic intervention, say an atypical antipsychotic, is introduced at 32 weeks and we want to model the response. This therapeutic input is shown in the lower left panel as a dotted line and affects the evolution of physiological states according to equations of motion on the right. These equations generate chaotic fluctuations in (three) pathological states shown on the lower right. Two of these states are then mixed to produce a trajectory in a psychopathological state-space. This trajectory is shown in the middle panel as a function of time (middle left) and as a trajectory in state-space (red line in the middle right panel). In turn, the psychopathology generates symptom scores (shown as colored lines on the upper left) and diagnostic probabilities (shown on the upper right). The relationship between the continuous (dimensional) latent space of psychopathology

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.

and the (categorical) differential diagnosis is determined by diagnostic parameters θ_i^{Δ} defining the characteristic location of the i^{th} diagnosis. These locations are shown as dots in the state-space: the blue dot corresponds to a diagnosis of remission, while the green, cyan, and red locations correspond to a diagnosis of schizophrenia, schizoaffective, and affective disorder, respectively. One can see that initial oscillations between schizophrenia and schizoaffective diagnoses are subverted by the therapeutic intervention. At this point, the latent pathophysiology is drawn to its (point) attractor at zero, the most likely diagnosis becomes remission and the symptom scores regress to their normal values of zero. In short, this models a successful intervention in a pathophysiological process that shows chaotic oscillations expressed in terms of fluctuating symptoms and differential diagnosis. Later we will see that, in the absence of therapy, these chaotic oscillations would otherwise produce a relapsing-remitting progression with an ambiguous diagnosis that fluctuates between schizophrenic and schizoaffective. This intervention is formally similar to what is known anecdotally as *chaos control* (e.g., Rose 2014). This example suggests that the goal of therapy is less about countering pathological deviations and more a subtle problem of suppressing chaotic or turbulent neurohormonal processes that are equipped with many self-organizing feedback mechanisms. Heuristically, the role of a clinician becomes much more like the captain of the ship that uses prevailing winds to navigate toward calmer waters.

This particular example is not meant to be definitive or valid in any sense. It is just one of a universe of potential models (or hypotheses) about the way psychiatric morbidity is generated. (Discussion will return to procedures for comparing models in the next section.) This example does allow us, however, to make a few key points about the nature of pathology and its expression. First, in any generative model of psychopathology there is a fundamental distinction between (time-invariant) parameters and (time-sensitive) states. This distinction can be regarded as the formal homologue of the distinction between *trait* and *state* abnormalities. For example, the patient illustrated above had a particular set of parameters θ_i^p determining the family of trajectories (and their attracting sets) of pathophysiology. Simply knowing these parameters, however, does not tell us anything about the pathological state of the patient at a particular time. To determine this, one needs to infer the latent pathophysiology in terms of the current state $x(t)$ using model fitting or inversion. This presents a difficult (but solvable) problem, because we have to estimate both the parameters (traits) and states of a patient to determine their trajectory in the short term.

The second distinction this sort of model brings to the table is the distinction between parameters that are patient specific and those conserved over the population to which the model applies. In statistical terms, this corresponds to the difference between *random* and *fixed* effects, where patient-specific effects model random variations in traits that may reflect predisposing factors

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.

(e.g., genetic predisposition). Conversely, other parameters may be fixed over patients and determine the canonical form of nosology.

In the example above, this distinction is illustrated by the difference between parameters that are specific to each patient or pathophysiology θ_i^p and those that are inherent in the nosology θ_i^n. The nosological parameters define a generic mapping from pathophysiology to psychopathology that is conserved over patients. Understanding this distinction is important practically, because nosological parameters can only be estimated from group data. Examples of this are given in the next section.

Model Inversion and Selection

Here let us consider the inversion and selection of generative models based on measurable outcomes. The ultimate aim of modeling is to predict outcomes for a particular patient. The quality of these predictions rests upon a model that is both accurate and generalizes to the sorts of patients encountered. The quality of a model is scored in terms of its evidence, given some data. However, to evaluate model evidence, one needs to be able to invert or fit data. This means that we first have to ensure that models can be inverted. In other words, can we recover the unknown parameters and latent variables responsible for clinical data? In what follows, the simulated patient above is used to see whether the latent states can be recovered, given the (known) therapeutic input and clinical outcomes (symptoms and diagnosis). Thereafter I will briefly review Bayesian model comparison and discuss its crucial role in hypothesis testing and elaborating a more mechanistic nosology for psychiatry in the future.

Model Inversion and Bayesian Filtering

The problem of estimating unknown parameters and latent states from time series data is known as deconvolution or filtering in the modeling literature. Because we have to estimate both parameters and states, this presents a *dual estimation problem* that is usually accommodated by treating parameters as very slowly fluctuating states. I will illustrate Bayesian filtering using an established procedure called *dynamic expectation maximization* (DEM). DEM was originally devised to infer latent neuronal states and the connectivity parameters generating neurophysiological signals in distributed brain networks and has been applied in a number of different contexts (Friston et al. 2008). Special cases of DEM include Kalman filtering (when the states are known and the state-space model is linear).

Figure 11.2 shows the results of Bayesian filtering when applied to the symptom and diagnostic time series shown in the previous figure. The format is similar to the middle panel of Figure 11.1; however, here, the colored lines correspond not to the true values generating data but to the estimated

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.

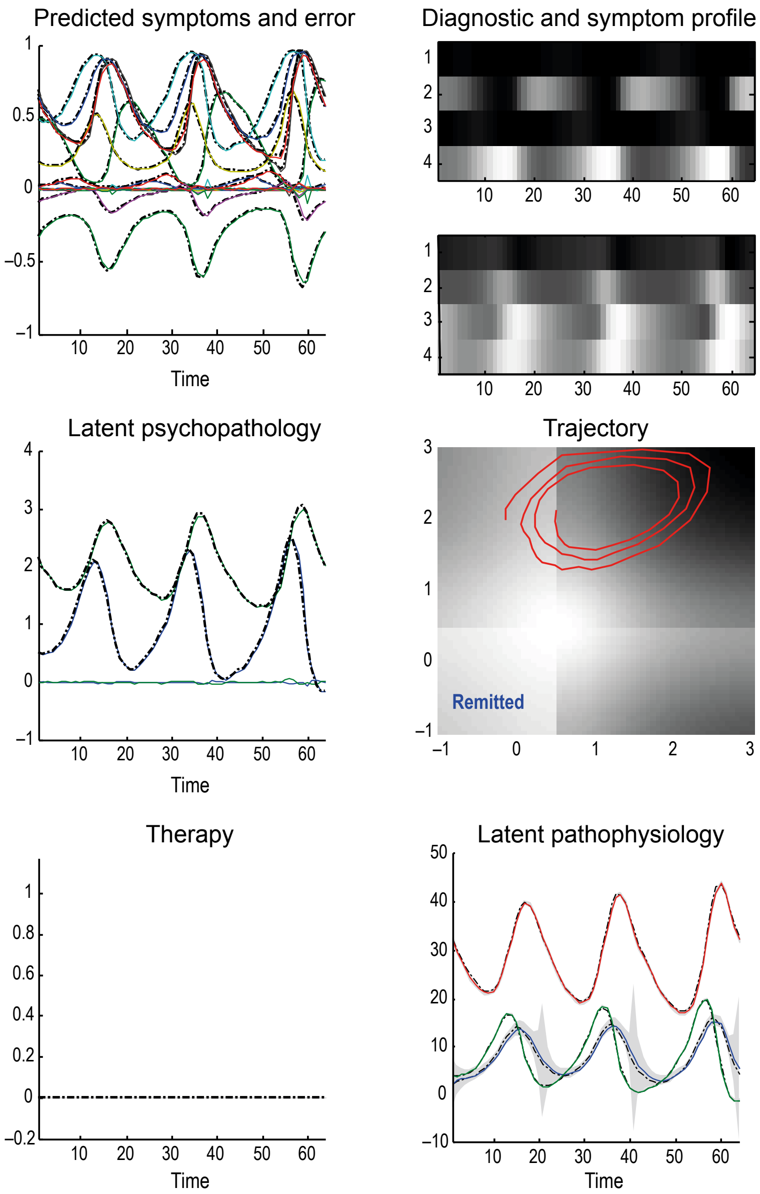


Figure 11.2 Bayesian filtering or model inversion of simulated clinical time series data: this figure reports the results of model inversion using the data generated by the model described in Figure 11.1. In this instance, the clinical data were simulated, without any therapeutic intervention, over 64 weeks and show the results using a similar format to the middle panels in Figure 11.1. Here, the solid lines represent predictions, while the broken black lines are the true values and actual outcomes being predicted. The gray areas correspond to 90% confidence intervals on hidden states. (*continued on next page*)

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 11.2 (continued) The colored lines correspond to clinical outcomes (upper left panel) and latent psychopathology (middle left panel). The symptom scores and differential diagnosis are shown as functions of time (upper left) and in image format (upper right). The top image shows the changes in differential diagnosis, with the diagnosis of remission in the first row; the lower panel shows the fluctuations in the first four symptom scores. Note that, in the absence of treatment, the chaotic fluctuations between schizophrenia and schizoaffective regimes of latent psychopathology slowly increase in amplitude.

trajectories based upon Bayesian filtering (as implemented with the Matlab routine *spm_DEM.m*). In this example, I simulated clinical progression in the absence of any therapeutic intervention (as shown by the flat line in the lower left panel). In the absence of any check on pathophysiology, chaotic oscillations of slowly increasing amplitude emerge over a period of 64 weeks. These fluctuations are shown in terms of a trajectory in the state-space of psychopathology (lower right panel) and as functions of time (middle row). Here, the solid lines correspond to posterior expectations (the most likely trajectories) that are contained within 90% Bayesian confidence intervals (gray areas). The true values are shown as dotted black lines. In this example, the true and estimated values were almost identical. This is because very low levels of random fluctuations were used—with log precisions of twelve, eight, and four—to control the amplitude of random effects at the level of outcomes, psychopathology, and pathophysiology, respectively (see Figure 11.1). Precision is the inverse variance or amplitude.

The upper panels show the resulting fluctuations in symptom and diagnostic scores as a function of time in graphical format (upper left panel) and in image format (upper right panel). One can see clearly that the differential diagnosis of schizophrenia and schizoaffective disorder vacillate every few months, reflecting an unstable and ambiguous diagnostic picture. Filtering was then repeated but with a therapeutic intervention at 32 weeks. The simulated response and inferred latent states are shown in Figure 11.3. These reproduce the results of Figure 11.1 and show the success of the intervention—as indicated by the emergence of a remitted diagnosis as time progresses (solid blue line on the upper left and cyan circle on the upper right).

In these illustrations, I estimated both the unknown states generating (simulated) clinical data and the patient-specific (trait) parameters governing pathophysiological dynamics. The estimated and true parameters are shown in the upper left panel of Figure 11.4: estimated values are shown as gray bars, true values in black, and white bars show 90% confidence intervals. The accuracy of these estimates is self-evident, with a slight overconfidence that is characteristic of approximate Bayesian inference implicit in dynamic expectation maximization (MacKay 1995). Although these estimates show that, in principle, one can recover the traits and states of a particular subject at a particular time, I used the true values of the nosological parameters coupling pathophysiology 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.

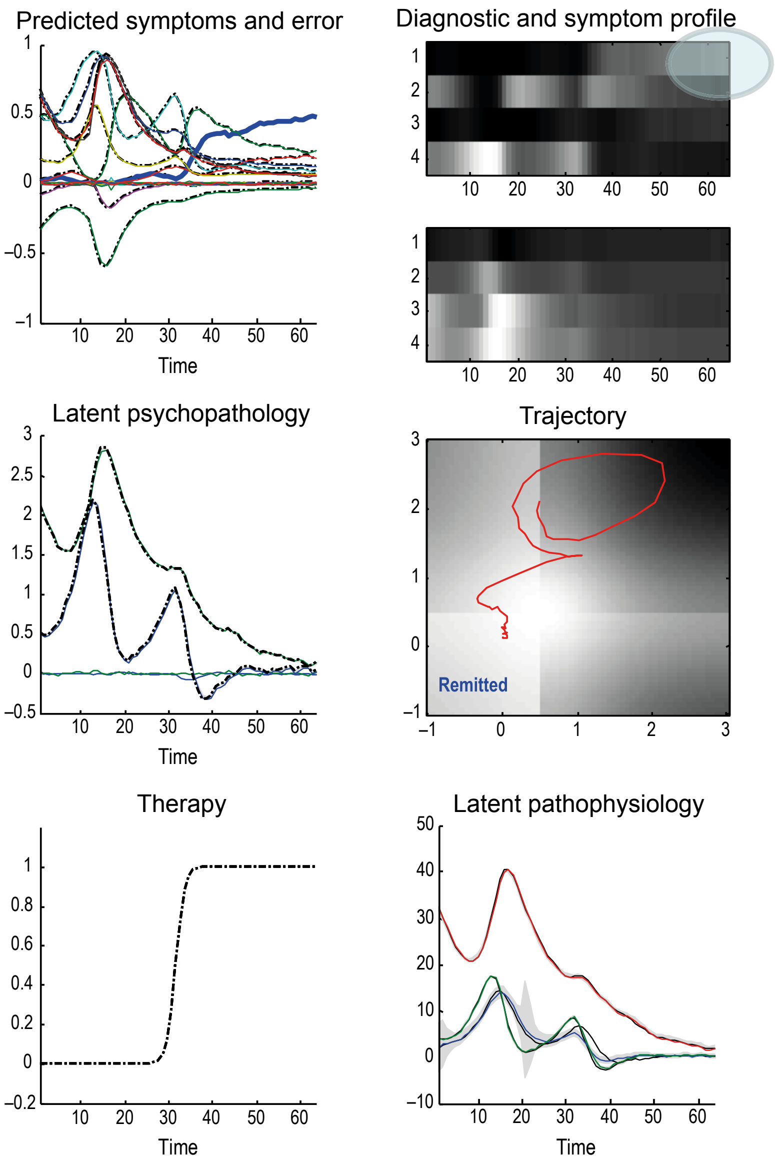
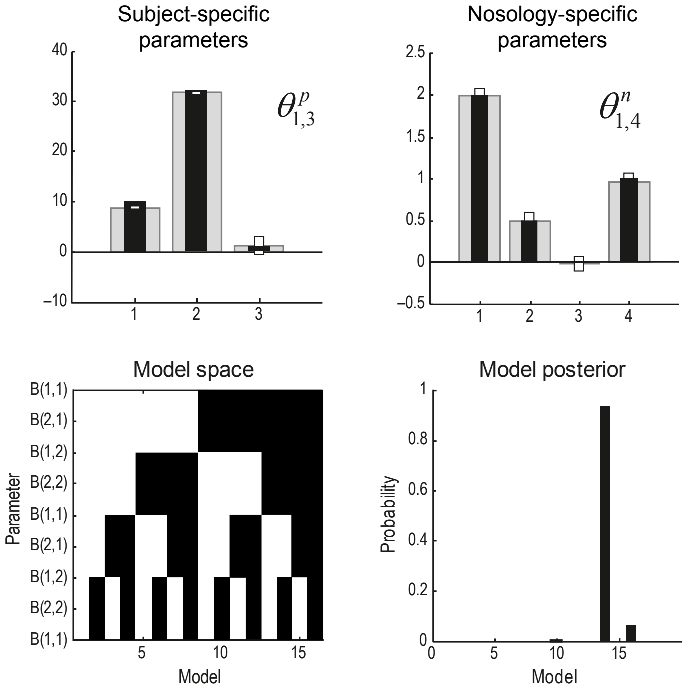


Figure 11.3 Inferred psychopathology and pathophysiology in response to treatment. This figure uses the same format as Figure 11.2. The only difference here is that a therapeutic intervention was introduced that destroyed the chaotic attractor, replacing it with a point attractor in the remitted regime of latent psychopathology. As a consequence, the pathophysiological variables flow toward zero and the symptom scores normalize. At the same time, the most probable diagnosis becomes one of remission (solid blue line in the upper left panel; see also the cyan circle in the corresponding predictions in image format).

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.


Subject-specific parameters

Nosology-specific parameters

Model space

Model posterior

Figure 11.4 Bayesian model identification and comparison: this figure summarizes the results of model identification (parameter estimation) and model selection using Bayesian model reduction. Upper panels show the posterior estimates of subject-specific (left) and nosological (right) parameters based on single subject time series and (eight subject) group data, respectively. Gray bars correspond to the posterior mean; white bars report 90% Bayesian confidence intervals. These are superimposed on black bars that correspond to the true values used to simulate the clinical data. The lower left panel shows the combinations of nosological parameters that define 16 competing models that were compared using Bayesian model reduction. This comparison entails evaluating the evidence for each model; namely the probability of the data under each model, having marginalized over unknown parameters and states. Model evidence is also known as the marginal or integrated likelihood. Under uninformative or flat priors over models, this also corresponds to the model posterior. The posterior probability over 16 models for the group data is shown on the lower right, suggesting that a model that precludes coupling between the second physiological state and first psychopathological state (mediated by the third nosological parameter) have greater evidence than all other models.

psychopathology (and generating clinical outcomes) during model inversion. One might ask: Can these parameters also be estimated?

To illustrate this estimation, the upper right panel of Figure 11.4 reports the estimates (and 90% confidence intervals) based on any *empirical Bayesian* analysis (Kass and Steffey 1989) of eight simulated patients (using *spm_dcm_peb.m*). Again, the estimates are remarkably accurate, suggesting that, in principle, it is possible to recover parameters that are conserved over subjects.

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.

Empirical Bayesian analysis of dynamic causal modeling estimators refers to the hierarchical modeling of within- and between-subject effects which may or may not be treated as random effects. Note that the third nosological parameter has an estimated value of zero. This is important because it leads us into the realm of Bayesian model comparison and evidence-based hypothesis testing.

Bayesian Model Comparison and Hypothesis Testing

Above, we saw that this sort of model can, in principle, be inverted such that underlying (latent) psychopathological and pathophysiological states can be inferred, in the context of (unknown) subject-specific parameters or traits. However, this does not mean that the model itself has any validity or will generalize to real clinical data. In other words, how do we know whether we have a good model?

This is a question of model comparison. In short, the best model provides an accurate explanation for the data with the minimum complexity. The model evidence reflects this, because model evidence is equal to accuracy minus complexity. The complexity term is important and ensures that models do not overfit data, and will thus generalize to new data. The model evidence is evaluated by marginalizing (averaging over) unknown parameters and states to provide the probability of some data, under a particular model. The model here is defined in terms of the number of states (and parameters) and how they depend upon each other. A simple example of model comparison is provided in Figure 11.4 (lower panels).

In our model, there are a number of ways in which the physiological states could influence psychopathology. There are two physiological states that can influence two pathophysiological states, creating four possible dependencies that may or may not exist. This leads to $16 = 2^4$ models which cover all combinations of nosological parameters (see the lower left panel of Figure 11.4). We can evaluate the evidence for each of these 16 models by inverting all 16 and evaluating the evidence or, as illustrated here, inverting the model with all four parameters in place and computing the evidence of all reduced models, with one or more parameters missing. This is known as *Bayesian model reduction*, which is an efficient way of performing Bayesian model comparison (Friston and Penny 2011). The results of this model comparison are shown in the lower right panel of Figure 11.4 and suggest that the posterior probability of model 15 is much greater than any of the others. In this model, the influence of the second pathophysiological state on the first psychopathological state has been removed. Removing this parameter reduces model complexity without any substantial loss in accuracy and therefore increases model evidence. We might have guessed that this was the case by inspecting the posterior density of the third nosological parameter mediating this model component (see the upper right panel).

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 is a rather trivial example of model comparison but illustrates an important aspect of dynamic causal modeling; namely, the ability to test and compare different models or hypotheses. Although not illustrated here, one can imagine comparing models with different numbers of pathophysiological states and different forms of dynamics. One could even imagine comparing models with a different graphical structure. One interesting example here would be the modeling of psychotherapeutic interventions that might influence pathophysiology through experience-dependent plasticity. This would necessitate comparing models in which therapeutic intervention influenced psychopathology, which couples back to pathophysiology, through the parameters of its dynamics. This is illustrated by the dotted arrows in Figure 11.1.

Many other examples lend themselves to speculation: crucial examples involve an increasingly mechanistic interpretation of pathophysiology, in which pathophysiological states could be mapped onto neurotransmitter systems through careful (generative) modeling of electrophysiological and psychophysical measurements (Stephan and Mathys 2014). One could also contemplate comparing models with different sorts of inputs or causes, ranging from social or environmental perturbations (e.g., traumatic events) to genetic factors (or their proxies like family history). Questions about whether and where genetic polymorphisms affect pathophysiology are formalized by simply comparing different generative models that accommodate effects on different states or parameters. For example, do models that include genetic biases on physiological parameters have greater evidence than models that do not?

The potential importance of model comparison should not be underestimated. Here I have tried to give a flavor of its potential. It is also worth noting that this field is an area of active research; fast and improved schemes for scoring large model spaces are continually being developed (e.g., Viceconti et al. 2015). One can construe an exploration of model space as a greedy search over competing hypotheses and a formal statement of the scientific process. This may be especially relevant for psychiatry, which addresses the specific problem of integrating both physiological and psychological therapies, and points to the need for generative models that map between these two levels of description. In the final section, let us turn to the more pragmatic issue of predicting response to treatment for an individual patient.

Prediction and Personalized Psychiatry

Let us assume that we used Bayesian model comparison to optimize our generative model of psychosis and prior probability distributions over its parameters. Can we now use the model to predict the outcome of a particular intervention in a given patient? In the previous section, we saw how clinical data from a single subject could be used to estimate subject-specific parameters (traits) and states at a particular time. In fact, the parameter estimates in Figure

11.4 were based on the first 32 weeks of data *before any treatment began*. This means that we now have estimates of how this particular subject would respond from any physiological state and the physiological state at the end of the period of assessment. Given these (posterior probability) estimates, there are several ways in which we can predict clinical prognosis and responses to different treatments. The simplest way would be to sample from the posterior distribution and integrate the generative model with random fluctuations to build a probability distribution over future states. We will take a related but simpler approach and apply Bayesian filtering to null data with zero precision; in other words, data that has yet to be acquired. This finishes a predictive distribution over future trajectories based on the posterior estimates of the subject's current parameters and states.

Figure 11.5 shows the results of this predictive filtering using the same format as Figure 11.3. However, there are two crucial differences between Figures 11.5 and 11.3. First, we are starting from latent states that are posterior estimates of the subject's current state and, more importantly, the trajectories are pure predictions based upon pathophysiological dynamics. One can see that the predicted response to treatment (at 16 weeks) has a similar outcome to the actual treatment (although the trajectories are not exactly the same, when comparing the predicted and actual outcomes in Figures 11.3 and 11.5, respectively). Figure 11.6 shows the same predictions in the absence of treatment, again showing the same pattern of fluctuation between schizophrenia and schizoaffective diagnosis encountered in Figure 11.2.

The right panel of Figure 11.6 also includes trajectories with increasing levels of therapeutic intervention (ranging from 0 to 2). The final outcomes of these interventions are summarized on the lower left in terms of the probability of a diagnosis of remission at 48 weeks. This illustrates the potential for predictive modeling of this sort to provide dose-response relationships and explore different therapeutic interventions (and combinations of interventions). In this simple example, there is a small probability that the patient would remit without treatment, which dips and then recovers to levels of around 50% with increasing levels of therapy. The apparent spontaneous recovery would not, however, be long lasting, as can be imputed from the chaotic oscillations in Figure 11.1.

Conclusion

In this chapter, I have illustrated what a computational nosology could look like using simulations of clinical trajectories, under a canonical generative model. The potential of this approach to nosological constructs can be motivated from a number of perspectives. First, it resolves the dialectic between categorical diagnostic constructs (e.g., DSM) and those based on latent dimensions of psychopathology or pathophysiology (e.g., RDoC). Both constructs play an

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.

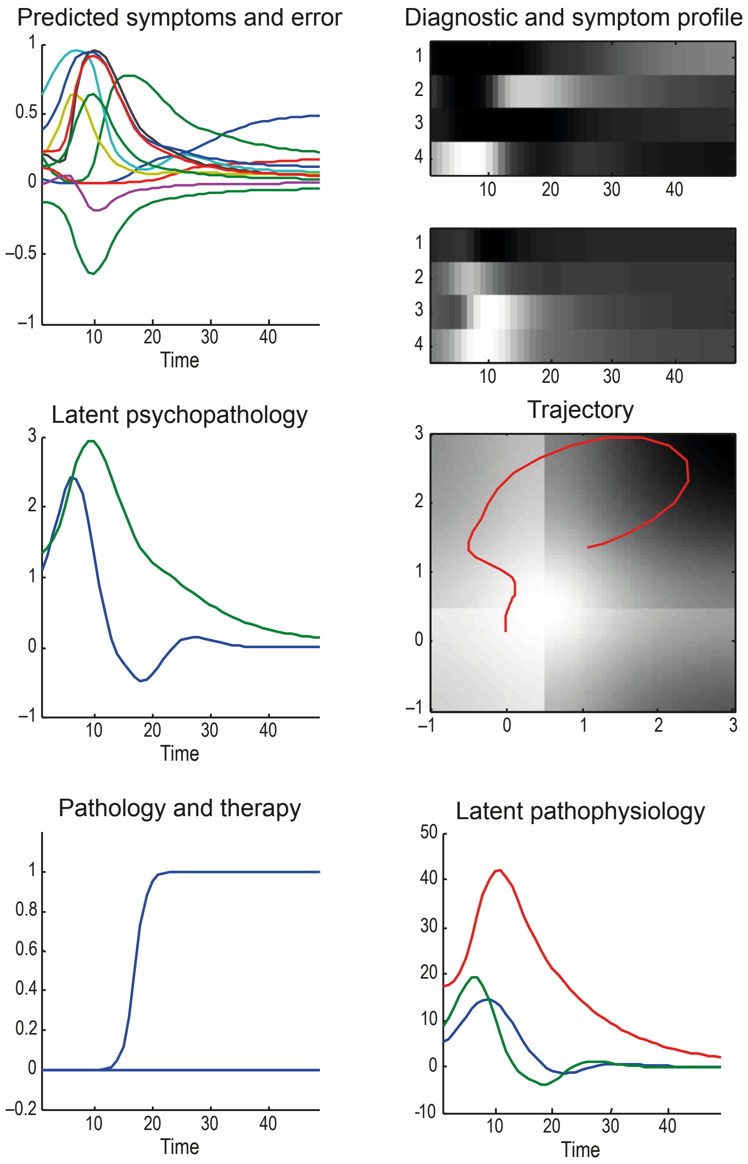


Figure 11.5 Predicted responses to treatment. This figure uses the same format as Figure 11.3; however, here, the results are purely predictive in nature. In other words, the predictions are driven entirely by pathophysiological dynamics based on subject-specific estimates of the model parameters and starting from the state last estimated on the basis of an assessment prior to therapy.

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.

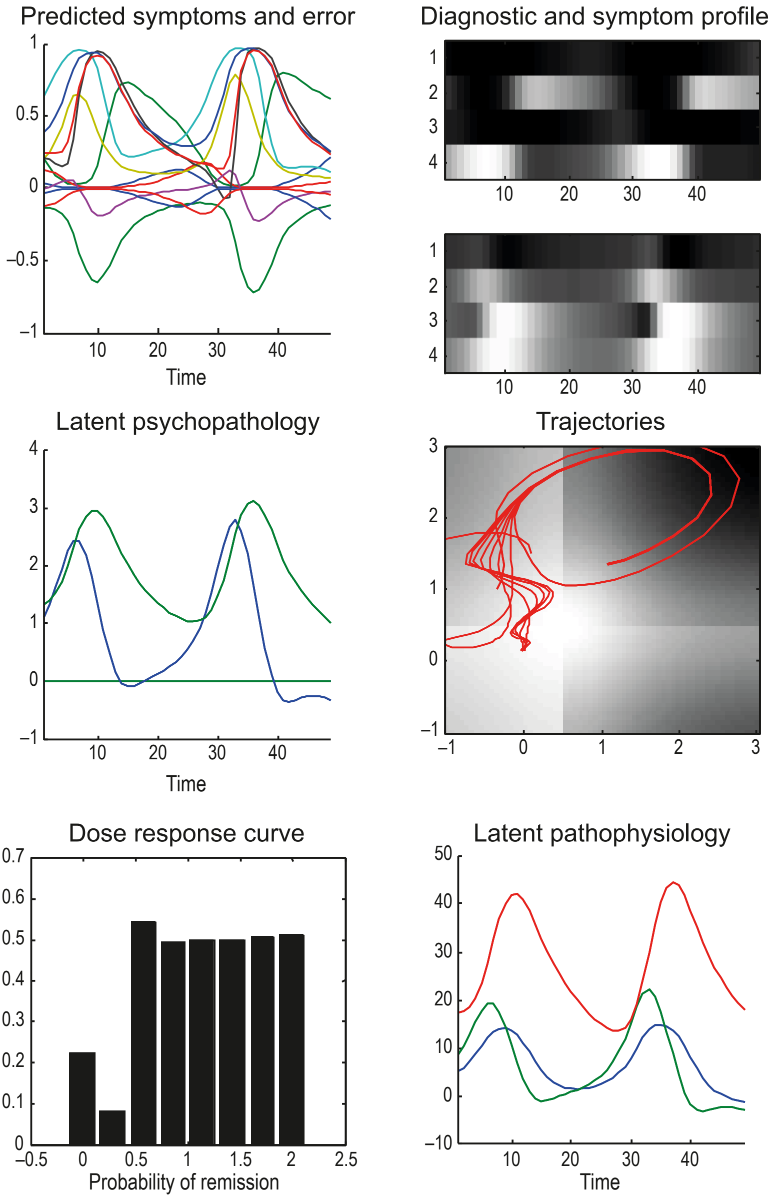


Figure 11.6 Predicted responses to different treatments. This figure uses the same format as Figure 11.5; however, here, it shows the results in the absence of any treatment. The trajectories in the middle right panel report eight simulations over increasing levels of therapeutic intervention. The endpoints of these trajectories are summarized on the lower left in terms of the probability of receiving a diagnosis of remission. This can be regarded as a predicted dose-response curve, illustrating the potential of the model to optimize treatment strategies.

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.

essential role in a generative modeling framework, as diagnostic outcomes and latent causes, respectively. To harness their complementary strengths, it is only necessary to determine how one follows from the other, which is an inherent objective of model inversion and selection. Second, I have tried to emphasize the potential for an evidence-based approach to nosology that can operationalize mechanistic hypotheses in terms of Bayesian model comparison. This provides an integration of basic research and clinical studies which could, in principle, contribute synergistically to an evidenced-based nosology. Finally, I have illustrated the practical utility of using the predictions of (optimized) generative models for individualized or precision psychiatry (Chekroud and Krystal 2015), in terms of providing probabilistic predictions of responses to therapy.

Although I have emphasized the provisional nature of this approach, it should be acknowledged that one could analyze existing clinical data using the model described in this chapter with existing algorithms. Indeed, there are hundreds of publications using dynamic causal modeling to infer the functional coupling among hidden neuronal states in the neuroimaging literature. In other words, it would be relatively simple to apply the techniques described above to existing data at the present time. The real challenge, however, lies in searching the vast model space to find models that are sufficiently comprehensive yet parsimonious to account for the diverse range of clinical measures—in a way that generalizes from patient to patient. This challenge is not necessarily insurmountable: one might argue that if we invested the same informatics resources in psychiatry as has been invested in weather forecasting and geophysical modeling, then considerable progress could be made. Ultimately, one could imagine model-based psychiatric prognosis being received with the same confidence with which we currently accept daily weather forecasts. There are, of course, differences between psychiatric and meteorological forecasting. For example, the latter must handle the "big data problem" with a relatively small model space. Conversely, psychiatry may have to contend with a "big theory problem," with a relatively large model space but more manageable data sets.

Perhaps the more important contribution of a formal nosology is not in the pragmatic application to precision medicine (i.e., through the introduction of prognostic apps for clinicians), but in the use of Bayesian model selection to test increasingly mechanistic hypotheses and pursue a deeper understanding of pathogenesis in psychiatry. This is the way in which dynamic causal modeling has been applied in computational neuroscience and, as such, is just a formal operationalization of the scientific process.

Acknowledgments

KJF is funded by the Wellcome Trust. The author declares no conflicts of interest.