9

The Computation of Collapse

Can Reliability Engineering Shed Light on Mental Illness?

Angus W. MacDonald III, Jennifer L. Zick,
Theoden I. Netoff, and Matthew V. Chafee

Abstract

Computational modeling in psychiatry has generally followed from efforts to understand cognitive processes (McClelland and Rumelhart 1986) or the nervous system (Hodgkin and Huxley 1952). This stands to reason: psychiatric disorders are disorders of thought and central nervous system activity. Although there are few contributions to psychiatry from probability theorists and engineers (Shewhart 1938; Miner 1945; Lusser 1958), the tools developed for quality control of metal fatigue and failed rockets may point to a useful approach for thinking about mental illness. This chapter argues that the *computational science of collapse*, which describes the manner and likelihood of failures in complex systems, provides a framework in which to use computational modeling for relating mechanisms to behavioral outcomes. This science, known as reliability engineering, is a branch of applied probability theory that has now been used for almost a century to help understand and predict how inorganic, complex systems break down. The idea of a fault tree analysis is introduced, a tool developed in reliability engineering which may be able to incorporate and provide a broader structure for more traditional computational models. Finally, Some of the current challenges of psychiatric classification are unpacked, and discussion follows on how this framework might be adapted to provide a unifying framework for classification and etiology.

Toward a Reliability Engineering Framework for the Central Nervous System

The reliability engineering framework provides a fresh perspective on the way in which we ask questions of, and report, our data. Historically, reliability engineering developed over the twentieth century as mechanical devices became

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.

increasingly complicated. In the 1920s, for example, Bell Labs faced the problem that many of the telephone amplifiers, which they produced, failed after being buried underground. To address this, Walter A. Shewhart pioneered the field of statistical quality control, bringing together the fields of probability theory and electrical engineering. In the 1960s, at the height of the Cold War, the same company was employed to use this approach to address problems in missile launch control systems. A principle of quality control engineering is that the reliability of an entire system declines as the number of interacting components increases (Lusser 1958). A careful description of the various ways in which interacting components could lead to a system failure highlighted weak points in the device and resulted in *fault tree analysis* (FTA).

Fault Tree Analysis on the Brain

FTA is a deductive failure analysis; the fault tree identifies how faults of individual components interact with other components resulting in overall failure.[1] To generate a fault tree, the different components of the device must be identified, as well as a description of how these component failures—called faults—interact and combine into failure modes. A fault occurs when a component is unable to perform its required function, such as a mutation in an ion channel or a neurotransmitter receptor that critically impairs synaptic communication. In combination, such faults can cause a cascade resulting in a general failure mode, such as a loss of information-processing capacity in cortical networks. In psychiatry, a symptomatic expression of this cortical network failure mode might be an impairment in some aspect of cognition, emotion, or behavior. Consistent with other authors (e.g., Redish 2013), we will propose that neuropsychiatric syndromes may be thought of as failure modes of the central nervous system.

Fortunately, the failure of a single component rarely results in a general system failure due to built-in redundancy and plasticity of the brain; this enables most people with many forms of insult to function normally in the world. In this way, a FTA makes affordances for causes that are neither necessary nor sufficient for dysfunction in and of themselves. To generate a fault tree, one identifies the different components that contribute to the failure. Consider a simple circuit with a main bulb and a back-up bulb, a power generator and a back-up battery, and a controller switch. The FTA in Figure 9.1a illustrates

[1] FTA is only one of a number of tools that may find application for understanding neuroscientific questions. Strictly speaking, FTA is a deductive, top-down method for organizing ideas, recording probabilities, and evaluating likelihoods, whereas failure mode and effects analysis is a bottom-up approach that focuses on how a fault in a single component propagates through a system. Together these two analyses constitute a failure mode effects summary, which is commonly done after, for example, airplane crashes. Both approaches may have potential for neuroscience, and at times the best tool for a question may be adapted from still elsewhere in reliability engineering's armamentarium.

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.

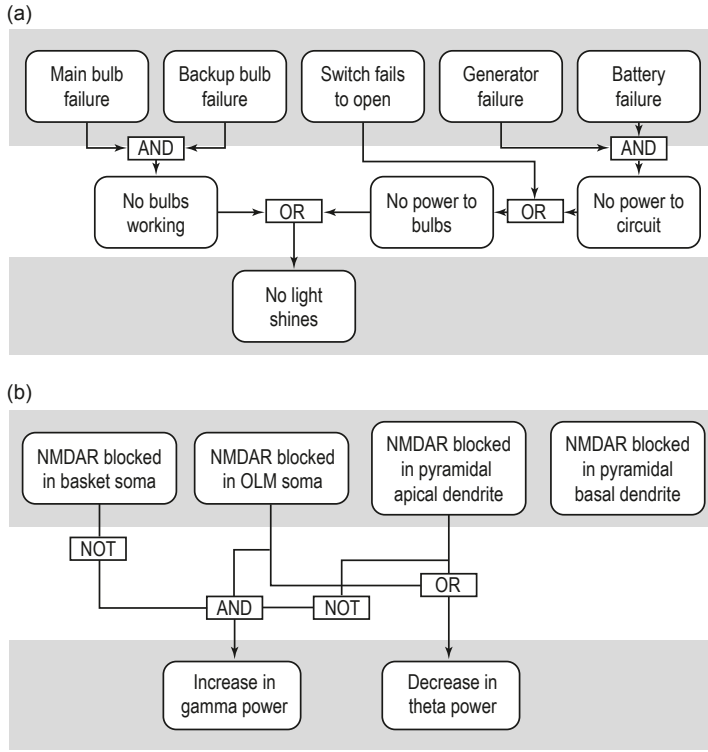

Figure 9.1 Examples of fault trees. (a) Typical fault tree of a simple mechanical circuit (after Rae and Lindsay 2004). (b) A fault tree generated from results of Neymotin et al. (2011). They used a computational model to investigate the conditions under which the power of theta frequency oscillations (3–12 Hz) decrease while the power of gamma frequency oscillations (30–100 Hz) increase, as seen in animals and human patients after ketamine administration. In their computational model, a decrease in theta power resulted when NMDA receptors (NMDARs) were blocked in either the somas of oriens-lacunosum moleculare (OLM) cells or in the apical dendrites of pyramidal cells, regardless of the function of NMDARs in other cell types. In the same model, an increase in gamma power occurred only when NMDARs were blocked in the somas of OLM cells and NMDARs were *not* blocked in basket cells or the apical dendrites of pyramidal cells. Thus the only combinations that generated both an increase in gamma power and a decrease in theta power involved blocking NMDARs in OLM soma with intact NMDARs in basket cells and pyramidal apical dendrites; the state of NMDARs in the pyramidal basal dendrites did not affect these results and can thus be said to be irrelevant in this case. Generation of a fault tree from these results allows one to visualize the roles that each factor plays in two effects, gamma and theta power.

how failures in these components might interact to cause a failure mode. It does this using Boolean logic "gates," describing how faults relate to other faults through AND/OR operators. Where the rate of such component failures is known (probability of failure over a timescale, e.g., per cycle or per day),

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

the likelihood of a failure mode can additionally be calculated (or the range of likelihoods, using additional Markov chain expansions in which a distribution of possible outcomes is sampled many times).

Figure 9.1b is an example of an FTA reconceptualized from a biophysically realistic computational model of theta and gamma oscillation generation. In this case, Neymotin et al. (2011) looked at the functioning of different hippocampal cell types to examine the impact of ketamine administration. Ketamine, an NMDAR antagonist, is known to induce a schizomimetic state that has been traced to a contemporaneous decrease in theta and an increase in gamma brain waves, particularly in the hippocampus. The researchers developed a Hodgkin–Huxley style network model of hippocampal neurons consisting of 200 each of basket and oriens-lacunosum moleculare (OLM) interneurons and 800 pyramidal cells. All these cell types had NMDA receptors on the soma, with pyramidal cells additionally having them on apical dendrites. The models showed that blocking all NMDA receptors decreased *both* theta and gamma, inconsistent with experimental findings at schizomimetic doses. They reasoned that *differences in sensitivity* to ketamine of NMDA receptors on the different cell types must be the source of these preanesthetic effects. Therefore, they independently manipulated four types of insults—blocking NMDA receptors on somas of (a) basket, (b) OLM, (c) pyramidal neurons, and (d) apical dendrites of pyramidal neurons. This resulted in 16 binary combinations: 2 (normal/off) raised to the 4^{th} (types of insults). The model found that pyramidal somatic NMDA receptors were largely irrelevant to theta and gamma power, whereas turning down pyramidal apical receptors alone was enough to decrease *both* theta and gamma power. In fact, the only condition in which they observed decreased theta and increased gamma was when OLM NMDA receptors were off, while the basket interneurons and apical pyramidal NMDA receptors remained functional. This result has been translated into the Boolean logic illustrated in Figure 9.1b.

The point may go without saying, but the purpose of the examples in Figure 9.1 is to illustrate the mechanisms underlying FTA and make explicit the parallels between mechanical and biological circuits. Several elements are omitted from these examples for the purpose of simplicity. First, the power of invoking probability theory is not illustrated. The probability of the co-occurrence of two independent events (pA AND pB = pA*pB) or the occurrence of either of two such events (pA OR pB = pA + pB – (pA*pB)) are familiar from introductory statistics courses. The impact of NOT, or inhibition, is straightforward. By simple extension, the inclusion of additional forms of logic (X-OR, 3 OUT OF 4, etc.) can also be readily incorporated. Furthermore, the effect of earlier event probabilities can then propagate through the FTA to examine the rate at which a general failure mode should occur. Second, many features of biological systems are not dichotomous and thus do not fall into simple categories. In such an event, classical probability theory can be augmented with a Markov chain and other methods, which involve resampling

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.

probability distributions.[2] This complicates the statistics but allows more dimensional variations of these concepts to be captured. Third, the events within the system need not be all within the organism. For example, stressors or treatments from outside the system can be modeled as externally controlled variables and can thereby alter the outcome (the rate of a specific fault and therefore the rate of general failure modes).

With these considerations in place, we come closer to imagining how FTA might provide a way of thinking about neuropsychiatric diseases and how different risk factors and treatments might interact. In this regard, the reliability-engineering framework is similar to other computational neuroscience approaches, insofar as the goal reflects a dissatisfaction with the correlative relationships that undergird much of what is known about mental illness. Like other computational approaches, the tool reflects specific causal hypotheses. Also, like other computational approaches, the tool risks becoming an armchair exercise, unless it generates hypotheses that are testable, either in patients or animal models. For example, the models can be generative—by pointing out domains and connections about which too little is known—and they can be tested and refined by assessing the extent to which known risks predict the characteristics of faults within a neural system and rates of failure modes (diseases) within a population.

Reliability Engineering on the Brain

Why have not reliability-engineering approaches been embraced more in psychiatry? One answer may be that biology in general, and the nervous system in particular, have failure rates and causes that are hard to quantify. The vast majority of the literature involving reliability engineering in the biological and medical sciences involves the traditional reliability of various medical devices. In the brain, reliability-engineering approaches are most evident in those places where engineers have had to share space with neuroscientists, such as in the study and manufacture of computer–brain interfaces (e.g., Polikov et al. 2005; Yousefi et al. 2015).

While reliability engineers may not tread into neuroscientific territory, there is a useful precedent in the work of the psychologist and neuroscientist Robert Glassman (1987). Glassman drew on Lusser's work in missiles and rocketry to speculate that component faults and failure modes were likely evolutionary constraints that led to redundancy and parallel processing in the brain (Lusser 1958). Lusser's law states that the reliability of components in series is equal to the product of the reliability of its component subsystems (an observation that

[2] Rather than a deterministic model, such simulation tools allow for probabilities to be assigned to several outcomes. For example, Markov chains examine a sequence of events using random draws to determine how the sequence proceeds. This then produces the distribution of probabilities informed by the internal structure of the sequences.

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.

led Lusser in the 1950s to dismiss the possibility of reaching the Moon because of the complexity—and therefore the low reliability—of the rockets required). Building on the theorems underlying Lusser's observations, Glassman derived the observation that the brain's apparent series-parallel networking operations had evolved to overcome inevitable faults in particular neurons or neural systems over the course of a lifetime using built-in redundancies, such as large neural populations firing in concert. Glassman also saw in Lusser's law the principles underlying *diaschisis*, which is the alteration of functions in brain regions far removed from a damaged area. Diaschisis might then reflect the brain's manner for overcoming such failures such that at first it is unable to produce the given behavior at all but, with recovery, it can again produce the given behavior only by accepting a lower level of precision (a higher fault rate) from some components.

We cannot speak to the extent that these ideas informed Glassman's subsequent work. It is clear that they did not lead a stampede of neuroscientists to seek out training in reliability engineering. However, in such matters, the selection of a target problem can make all the difference. In the next section we will try again, this time by applying the framework of FTA to the challenge of integrating classification and etiology in the study of mental disorders.

Comparing Frameworks for Classification and Etiology of Mental Disorders

Although testing, or even proposing, a formal FTA for a specific mental illness is beyond the scope of this chapter, we believe that it will be a useful complement to other computational approaches in the future. Even in the absence of a realized FTA model, the reliability-engineering approach provides a framework that contrasts with the two frameworks for thinking about psychiatric and personality disorders currently prominent in the field. *Frameworks* are the premises and concepts that tacitly guide our research. For example, the number of angels that can dance on the head of a pin is now a byword for a pointless debate, but it was once a subject of serious discourse; we have long since retired the framework that led to those arguments. Are there parts of our framework for studying psychiatric disorders similarly ready for retirement?

To address this issue we will unpack the two prominent frameworks in psychiatry, which we term the neo-Kraepelinian and the reverse-engineering frameworks, and then contrast them with the reliability-engineering framework. In particular, we will examine how these frameworks affect the way in which we link the causes of a disorder to its symptoms. Since classification and cause are so central to psychiatric research, these are domains where an incorrect framing of the questions could lead us hopelessly astray. We will argue that both the neo-Kraepelinian and the reverse-engineering

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.

frameworks are misaligned with the nature of psychiatric disorders. Our aim will then be to use the tools of reliability engineering to make this discrepancy more explicit.

The Neo-Kraepelinian Framework

The founder of modern psychiatry, Emil Kraepelin (Kraepelin and Diefendorf 1907; Kraepelin 1919), was the original proponent of the quasi-medical framework used most widely in psychiatry today. This framework came to prominence in the 1970s, supplanting the psychoanalytic framework used in the second edition of the Diagnostic and Statistical Manual (DSM-II). This neo-Kraepelinian framework posited distinct categories of illness that could be assigned to someone who had a sufficient number of observable symptoms. The framework was codified in the Feighner criteria (Feighner et al. 1972), the Research Diagnostic Criteria (RDC; Spitzer et al. 1975), and eventually the third edition of DSM (DSM-III). These codes suggested that symptoms were useful for determining whether someone fulfilled the necessary and sufficient conditions for diagnosis. However, it was also acceptable for patients to share a diagnosis without sharing any symptoms. It was hoped that a formal diagnostic framework would decrease idiosyncratic noise, increase the reliability of diagnoses, and harmonize practice across laboratories and clinics (for a critique of these aspirations, see Markon et al. 2011). This served to reify the search for natural categories with distinct etiological and pathophysiological characteristics (Hyman 2010). Within this framework, theories about how the neural functions of, for example, schizophrenia patients may be distinct from the neural functions of depressed, alcoholic, or obsessive-compulsive patients were immediately salient and substantive.

Forty years on, the premises and concepts of the neo-Kraepelinian framework are hampering progress toward the grand challenges of psychiatric research (Persons 1986; Van Os et al. 1999; Krueger and MacDonald 2005; Markon et al. 2005; Hyman 2010). There is increasing evidence that upstream genetic, cellular, and neural system impairments are shared across distinct disorders, even between categorically distinct disorders. To follow up on our example, schizophrenia has at times been thought of as a categorically distinct psychiatric disorder. It is somewhat surprising, then, that 50% of people with schizophrenia also fulfill criteria for comorbid substance abuse at some point, and 50% fulfill criteria for depression (Buckley et al. 2009). People with schizophrenia are also at a 12-fold greater risk for obsessive-compulsive disorder (Pokos and Castle 2006), whereas those with obsessive-compulsive disorder are at a fourfold greater risk for schizophrenia (Tien and Eaton 1992). The levels of comorbidity between other mental disorders can be equally as high. In any case, psychopharmacology and psychotherapy frequently use the same medications and techniques in practice across different diagnoses (for additional critique, see MacDonald 2013). At some point, the neo-Kraepelinian

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.

medical framework became more useful to insurance adjusters and lawyers than to patients, clinicians, or even researchers. In the words of Thomas Insel, director of the National Institute of Mental Health (NIMH) when DSM-5 was released: "Patients with mental disorders deserve better."[3] For these reasons, some scientists are moving away from the neo-Kraepelinian framework toward something new, which we refer to as informal reverse engineering.

The Informal Reverse-Engineering Framework

Reverse engineering involves analyzing a complex system related to a function to determine the mechanisms underlying that function. This framework is already implicit in much neuroscience research, while NIMH's Research Domain Criteria (RDoC) perspective (explicitly named in recognition of the RDC framework it replaces) is the most codified version at this time. "The mandate for RDoC is to consider psychopathology in terms of maladaptive extremes along a continuum of normal functioning, to promote a translational emphasis" (Ford et al. 2014:S296). At the core of RDoC is a matrix with rows consisting of functional dimensions organized into five broad categories (positive valence systems, negative valence systems, cognition, social processes, and arousal). The columns of the matrix are levels, or units, of analysis ranging downward to genes and upward to behavior and symptoms.[4] Thus, the framework strives to organize extant knowledge about a multitude of cognitive and affective processes with research findings about brain networks, neurons, neurotransmitters, proteins, and genes (Insel and Cuthbert 2009; Stanislow et al. 2010; Cuthbert and Kozak 2013; Ford et al. 2014). The principle motivating RDoC is that patients who are sorted according to some shared functional deficits (e.g., in working memory, attention, executive control) will have more in common in terms of brain functioning than do patients placed in the same diagnostic groups according to neo-Kraepelinian schemes. The hope is that this new framework should facilitate the discovery of the underlying neural mechanisms that cause neuropsychiatric disease. RDoC is based on several reasonable, but untested, assumptions:

- A focus on functional deficits will direct research toward causal biological mechanisms more rapidly than a focus on clinical symptoms.
- Patients grouped based on functional deficits will be more homogenous with respect to underlying biological mechanisms than grouping based on clinical symptoms.

[3] April 2013 blog post, available at http://www.nimh.nih.gov/about/director/index.shtml. For partial retraction, see http://www.nimh.nih.gov/news/science-news/2013/dsm-5-and-rdoc-shared-interests.shtml (accessed July 7, 2016).

[4] http://www.nimh.nih.gov/research-priorities/rdoc/constructs/rdoc-matrix.shtml (accessed July 8, 2016).

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.

- Clinical symptoms and functional deficits derive from a common set of biological mechanisms, so that studying one will provide insight into the other (which is necessary if treatments that improve functional deficits are also to improve clinical symptoms).

If patients with similar functional deficits do not end up sharing more in terms of a common set of underlying neurobiological deficits, it would suggest that either there are no meaningful categories of neuropsychiatric disease with unique neural signatures, or that new functional axes closer to neural functioning need to be identified. In spite of the uncertainty of these propositions at this stage, it seems that shifting toward functional deficits and away from clinical symptoms will enable a tighter link between biology and behavior in neuropsychiatric research. For this reason, RDoC holds enormous potential for accelerating discovery.

We refer to the RDoC approach as an instance of "informal reverse engineering" in the present context because, although it seeks to identify the biological causes underlying behavioral deficits in neuropsychiatric disease, the RDoC framework does not attempt to provide a quantitatively rigorous or unifying framework for achieving this. A potential limitation of the informal reverse engineering in general, and RDoC as a manifestation of it, is that by isolating psychological constructs from each other (rows in the RDoC grid), attention is drawn away from the sources of *structure* in psychopathology. It is this structure of covariation in cognitive and affective dysfunctions within and across patients that initially led to the delineation of diagnostic entities early in the twentieth century. Even when statistical relationships are discerned, it is not easy to append these into the cumulative science of mental illness. In particular, RDoC does not make affordances for how multiple causal factors interact, nor does it provide a basis for predicting how a complex set of interacting neural systems that collectively malfunction as a result of the disease will respond to interventions intended to normalize brain function. In short, it is not clear that RDoC in its current form will identify treatments. Alternatively, it is possible that with perfect knowledge, functional deficits in working memory, attention, or executive control may turn out to result from diverse biological causes. Under such circumstances any new grouping (or dimension) would not reduce heterogeneity within groups at all, which would limit its usefulness as well.

We argue for a more quantitatively rigorous framework that simultaneously affords an account of how causes that are neither necessary nor sufficient in and of themselves can result in a disorder, and how within-diagnosis heterogeneity and between-diagnosis comorbidity arise. With that said, there are many components of reverse engineering, even informal reverse engineering, that are a salubrious and necessary part of any FTA of a biological system. First among these are computational models.

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.

Fault Tree Analysis Framework for Syndromes and Computational Models in Psychiatry

If you don't know where you're going, you don't have very much control over where you arrive. The framework offered by reliability engineering allows us to envision what a description of psychopathology might look like and provides a set of tools to move forward. This possibility goes beyond the classification of people into disease categories or the measurement of people along cognitive dimensions, and begins to reintegrate causes into our conceptualization of mental disorder. Further, the framework provides a means for describing both how a single therapy might affect several different disorders and how different therapies can all reduce the same symptom.

As is well established elsewhere in this volume, computational models are mathematical formalizations of hypotheses. As noted above from the work of Neymotin et al. (2011; see also Figure 9.1b), integrating computational modeling with FTA in the context of psychiatric disease allows us to relate different risk factors to disease pathophysiology forming the outcome of a model. Such computational models can enable a mechanistic understanding of the linkages between faults that may occur in the tree; they can also relate mechanisms understood at one scale of the fault tree to outcomes at another scale.

The schematized illustration of FTA in Figure 9.2 demonstrates a number of properties of an expanded FTA framework that make it desirable for understanding the etiological and classification problems in psychiatry. Levels of analysis are illustrated in a series of gray bands, and the relationships between those gray bands is explicitly illustrated with simple logic gates. Two symptoms (A and B) represent two failure modes of the system. The co-occurrence (or comorbidity) between them is shown to be the result of the relationships among four cognitive/affective processes. Because of the dual role played by cognitive/affective process C in the two symptoms, the likelihood of co-occurrence is greater than chance (and could be explicitly calculated and compared to empirical rates), depending on the presence of other processes. Cognitive/affective process C is not the only source of comorbidity, as cell process B plays a role in three of the cognitive processes, rendering those, in turn, nonindependent.

Suppose symptoms A and B are symptoms of a given disorder. In a number of psychiatric disorders, two patients can share a diagnosis without sharing symptoms. Similarly, each patient can in turn closely resemble someone who does not fulfill the criteria. How might this be explained by FTA? In our schematic, a patient with an impairment in cognitive/affective processes A, B, C, and D will express symptom A but not B; whereas another with impairments in processes C and D will only show symptom B. This is troubling from a neo-Kraepelinian perspective, where a disease is recognized by fulfilling a series of necessary and sufficient conditions because they have a specific etiology. On the other hand, a FTA framework provides a tool for thinking about

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.

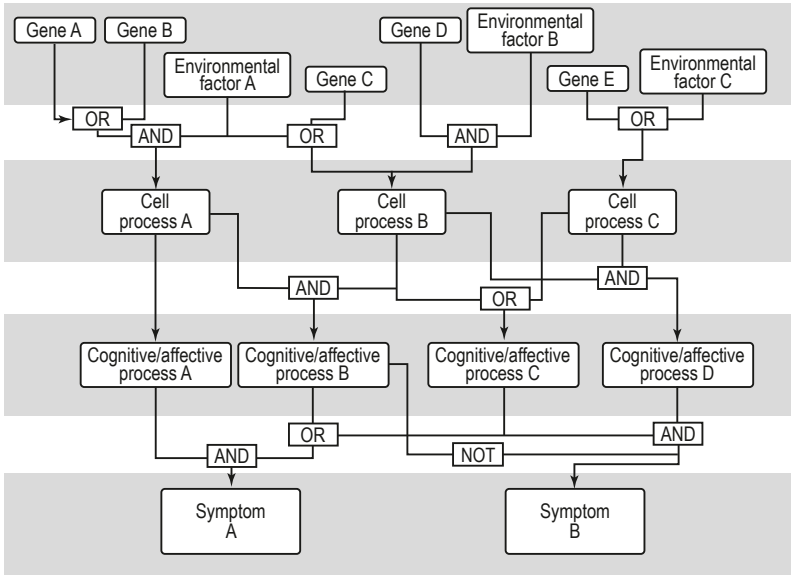

Figure 9.2 Schematic of FTA analysis illustrating the role of logic gates to integrate causes across levels of analysis and to generate nonindependence across symptoms.

syndromes. Syndromes are constructs that organize and label symptoms and other measurable signs that are often seen together. Syndromal descriptions are helpful in differential diagnosis because they prompt investigation of other features of the syndrome when the first signs are observed. Syndromes provide guidelines, but do not have rigorous necessary and sufficient conditions.

The classical psychiatric disorders are syndromes identified by the relative co-occurrence of these various failure modes, or symptoms. Our FTA example in Figure 9.2 illustrates how disorders all arise from the contributing causes of upstream faults: in some cases, several such faults will be required for a downstream failure (e.g., gene D and environmental factor B can lead to cell process B). In other cases, two alternative faults can lead to the same failure (e.g., gene E or environmental factor C causes cell process C). The result will be a statistical co-occurrence of failures in cognitive and affective processes linked to symptoms in the absence of clearly defined diagnostic boundaries. Uncovering the fuzzy logic of syndromes will be difficult to do if the focus remains on diagnostic group differences. Instead, research will add value by linking the likelihood of symptoms to the likelihood of cognitive and affective failures, and in turn by linking those failures to the likelihood of upstream faults in cellular systems, genetic polymorphisms, and environmental stressors. While this also resembles a research agenda based in informal reverse engineering, it is distinct in (a) prioritizing the importance of observing the rates of various faults, (b) making explicit their multiple causes and their various outcomes, and (c) providing potential for application of a rigorous, quantitative prediction.

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.

Honing the relationships present in the fault tree can proceed in parallel with, and in turn complement, the efforts of computational modeling to unpack the mechanisms across different levels of analysis. For example, genomic studies have identified many mutations related to schizophrenia and depressive disorder, all of which have only weak correlations to the disorder. However, identification of changes in the genetic code does not directly link the etiology to the behavioral outcome, primarily because it is unclear how genetic mutations that change the properties of an ion channel or neurotransmitter system change the function properties of neural circuits to alter how they process information. Computational models can be used to relate what we may know at the molecular and cellular scale to the observable changes in neural function, circuit dynamics, and ultimately behavior. For example, Hodgkin–Huxley neuronal models simulate ion channel and synaptic conductances to predict cellular dynamics. Channel mutations can be modeled by changing parameters and measuring the resulting changes in excitability and/or spiking patterns in artificial neural networks. These networks can be trained to "perform" behavioral tasks (e.g., process stimuli, select between available behavioral responses) that measure specific cognitive impairment in patients. Therefore, these models are useful for linking changes at the protein scale to changes in neural function and behavior. At another scale, mean-field models simulate the average firing rates of populations of neurons in brain regions. These models can be used to relate changes in excitability or connection strengths to the emergence of synchrony and population oscillations that may be measured in system-level biomarkers such as changes in fMRI, EEG, and even to cognitive deficits.

An advantage of computational models is that they can assess how pharmacological, electrical, or optogenetic therapies could potentially modulate neural dynamics in networks to normalize information processing and behavior. By testing these predictions from the models, we are inherently testing our underlying hypothesis of the physiological mechanisms resulting in the disease state. Still, to be tractable, computational models necessarily focus on a small set of empirical observations and struggle to capture various syndromal and epidemiological aspects of psychiatry. For example, Voon et al. (2015) found that a bias toward model-free learning was more prevalent in people with binge eating disorder, methamphetamine addiction, and obsessive-compulsive disorder. Thus, diverse disorders of compulsivity are accompanied by an excessive tendency toward model-free learning. While a compelling mechanistic account of compulsive symptoms, it does not yet account for an important complication: the low level of comorbidity between these compulsivity disorders if they were indeed caused by a single fault. Thus, while the computation model is of itself mechanistic (and possibly correct), a full understanding of binge eating disorder, methamphetamine addiction, and obsessive-compulsive disorder will involve understanding why the same fault results in addiction, in one case, and obsessive-compulsive disorder, in another. Circumstances like this, where there is not a one-to-one mapping between a modeled fault and a particular

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.

symptom, are likely to be the norm. In this regard, we see FTA and computational modeling as complementary. The extent to which the predictions made by any given model correspond to observed probabilities of various outcomes provides useful validation of both the FTA and the computational model.

A Reliability Engineering-Aligned Research Agenda

While FTA has many attractive features for helping to systematize the etiology and classification of mental disorders, there are clear limitations to the application. First, generating a fault tree that spans genes to diseases is an immense task. Is it practical to complete an FTA for any given disorder, or is this really just a framework in which to think about disease diagnosis? We propose that the FTA may be practical for relating some specific etiologies to outcomes. In this sense, it provides a new way to summarize information and structure reviews of a given domain. While it may not yet be practical to generate a globally encompassing fault tree that relates all diseases into a single framework (i.e., an FTA of the brain), there are statistical tools that make the task less daunting. For example, probabilistic graphical models are an increasingly popular method for detecting nonlinear (including Boolean) relationships between observed variables (Praveen and Fröhlich 2013). Second, the FTA framework implies unidirectional causality: genes are responsible for cellular physiology, and physiology is responsible for behavior. Feedback loops are not, to our knowledge, accommodated in a simple way. Clearly, in biological systems there is feedback between every level, and this feedback can be incorporated into computational models. While feedback is complicated for the proposed approach, steady-state effects of such circuits and fluctuations in those steady states may be incorporated into expanded versions of FTA. Third, coupling between nodes within the FTA may be fit to data to best describe general outcomes, but may not represent any particular patient's connections. Therefore, an FTA alone may not be sufficient to provide useful guidance for selecting a given patient's therapy (cf. footnote 1 regarding failure mode and effects analysis, and failure mode effects summaries, which may, in time, provide precisely this type of guidance). Fourth, strictly speaking, FTA builds toward a single general failure mode. However, one of the features that may prove particularly useful for biological systems, such as the brain, is the way in which two FTAs can overlap. Thus it will be necessary to expand these models to exploit more fully the ways in which shared causes of two or more failure modes can be illustrated and calculated.

At present, there are only fragmentary parts of fault trees for complex mental disorders. FTAs are assembled from data collected from very disparate sources to relate causal relationships between different elements at adjacent scales and correlational relationships between more distant elements. The adoption of the FTA framework, perhaps in concert with probabilistic graphical modeling as needed, could therefore play a useful role in directing future research, for

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.

example, quantifying statistical relationships between key disease variables in existing datasets, and identifying high-value data to test computational models. There are some data sources that relate specific connections within the tree (e.g., genes to cell physiology, or cell physiology to network oscillations), but others relate scales that are not directly coupled (e.g., correlations between genes and disease prevalence). Mechanistic connections can be tested in computational models and further validated in animal models. These models can be used to identify the mechanisms or symptoms that relate the different components of the genomics and physiology into a biometric. The goal is to generate a single model that both explains the direct mechanistic connections and is also consistent with the indirect correlations. This will involve going beyond the statistical probabilities and effect sizes to which we are accustomed.

We have brought some attention to a field that has been working with quirks of nonbiological systems for almost a century. Many of the challenges to which this field of reliability engineering has struggled have parallels with biological systems. One tool from reliability engineering, FTA, provides an overarching framework for thinking about how complex systems, such as the brain, can break down. The reliability engineers' path that linked underground telephone amplifiers to Moon-bound rockets was filled with happy and some not-so-happy accidents. The path ahead for the computation of collapse may be much closer to home. We suggest that going on this journey begins with a mental shift, from the traditional medical neo-Kraepelinian framework to that of reliability engineering.

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.