Published in final edited form as:

Annu Rev Psychol. 2022 January 04; 73: 243–270. doi:10.1146/annurev-psych-021621-124910.

Computational Psychiatry Needs Time and Context

Peter F. Hitchcock[1], **Eiko I. Fried**[2], **Michael J. Frank**[1,3]

[1]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island 02912, USA

[2]Department of Clinical Psychology, Leiden University, 2333 AK Leiden, The Netherlands

[3]Carney Institute for Brain Science, Brown University, Providence, Rhode Island 02192, USA

Abstract

Why has computational psychiatry yet to influence routine clinical practice? One reason may be that it has neglected context and temporal dynamics in the models of certain mental health problems. We develop three heuristics for estimating whether time and context are important to a mental health problem: Is it characterized by a core neurobiological mechanism? Does it follow a straightforward natural trajectory? And is intentional mental content peripheral to the problem? For many problems the answers are no, suggesting that modeling time and context is critical. We review computational psychiatry advances toward this end, including modeling state variation, using domain-specific stimuli, and interpreting differences in context. We discuss complementary network and complex systems approaches. Novel methods and unification with adjacent fields may inspire a new generation of computational psychiatry.

Keywords

computational psychiatry; network approach; state versus trait; domain specificity; temporal dynamics; functional analysis

1. INTRODUCTION

Computational psychiatry is a burgeoning research field that applies methods, formalisms, and theories from the computational cognitive neurosciences to mental health. The last decade has seen an explosion of research in both theory-based (formal accounts of mental health) and data-driven (predictive modeling using many variables) approaches. Attesting to the field's promise, several studies have found that predictions of diagnostic categories or symptoms could be improved by including latent parameters estimated through computational models fit to brain or behavioral data (reviewed in Huys et al. 2021, Maia & Frank 2011, Wang & Krystal 2014). Here we focus on emerging challenges as computational psychiatry matures (Browning et al. 2020, Williams 2016): How can the field help us understand how mental health problems differ from one another? What modeling

eiko.fried@gmail.com .

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at http://www.annualreviews.org/errata/psych

strategies are needed for different kinds of problems? And what methods will be helpful for modeling temporal dynamics and the social and environmental contexts in which mental health problems emerge?

The allure of computational psychiatry is that it is organized around theories such as reinforcement learning, dynamical systems, neural networks, Bayesian decision making, and sequential sampling. These theories span many fields, including mathematics, computer science, and computational cognitive neuroscience. Thus, unlike many psychological theories with shallow roots in basic science (Haslbeck et al. 2021), computational psychiatry theories build from deep terrain, ranging from mathematical theories to biological sciences. Computational psychiatry offers principled techniques to link processes across levels of analysis (see Eronen 2019). In particular, it provides distinct vantage points on neurocomputational functions, from rational analysis of the problem being solved to algorithmic details of specific solutions to plausible biological implementations (Huys et al. 2016, Maia & Frank 2011, Wang & Krystal 2014).

Despite its promise, computational psychiatry has yet had little influence on clinical practice (Rutledge et al. 2019). A running joke in the field is that the number of reviews hyping the field's promise has outpaced its empirical results. With the benefit of retrospect, however, it was perhaps unrealistic to predict dramatic and near-immediate progress on a topic as complex as mental health. Early disappointment may have come from overoptimism rather than fundamental limitations of the field. Computational psychiatry also has had difficulty recognizing how different mental health problems are from one another. As such, it may have been slow to adopt sufficiently distinct modeling strategies for problems that drastically differ. We propose that, to accelerate progress, the next generation of computational psychiatry research will need to incorporate modeling strategies suited to even the most complex problems (see also Gillan & Rutledge 2021, Moutoussis et al. 2017).

Neurocomputational process:

an input-output transformation and the neural machinery that effects it

A key challenge in early computational psychiatry has been the field's reliance on diagnostic systems that are widely acknowledged to be flawed, such as the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; Am. Psychiatr. Assoc. 2013). Many phenotypes are poor—they lack reliability and validity and are highly heterogeneous—and as such they permit limited conclusions about mechanisms (i.e., "garbage in, garbage out"). Yet, much early computational psychiatry research (including our own) recruited healthy controls and compared them to individuals with one mental health disorder (or severity cutoff) as conceived by the DSM. Diagnostic systems delineate static and categorically distinct mental health problems, yet many problems are best thought of as mixtures of dynamically interacting and dimensionally varying processes (Borsboom 2008, Gillan et al. 2017, Kotov et al. 2017, Kozak & Cuthbert 2016, Nelson et al. 2017). Dimensional and transdiagnostic approaches have thus been increasingly utilized in computational psychiatry (Gillan & Seow 2020, Gillan et al. 2017, Gueguen et al. 2021, Wiecki et al. 2015). In psychopathology research broadly intended, three prominent alternatives to sharp diagnostic delineation

have been recently developed. First, the Research Domain Criteria (RDoC) assumes that mental health symptoms arise from mixtures of individual differences in cognitive and emotional processes (Kozak & Cuthbert 2016). Second, the Hierarchical Taxonomy of Psychopathology (HiTOP) uses factor analytic methods to investigate symptom co-occurrence patterns across a broad, transdiagnostic space of mental health problems (Kotov et al. 2017). Third, the network approach to psychopathology views mental health problems as dynamic systems of elements that interact within and across diagnostic boundaries (Borsboom 2008, Fried & Cramer 2017).

Although these three approaches differ in many respects, they concur that it is unwise to attempt to cleanly distinguish individuals with one mental health problem from individuals with another mental health problem at a single point in time. This critique comes down to the perils of essentialist thinking about mental health problems. Essentialist thinking focuses attention away from the superficial features of a phenomenon and toward an internal mechanism or property assumed to give rise to it (Gelman 2004). This is unproblematic if mental health problems are indeed characterized by a "single, well-defined etiological agent" (Kendler et al. 2011, p. 1144) that is both necessary and sufficient to distinguish individuals with and without the problem (as if it were an infectious disease). If this were the case, grouping 500 patients diagnosed with major depressive disorder (MDD) into the same category and investigating their biological markers compared to those of a healthy control group would be a sound scientific method. However, many mental health problems appear to be best understood as complex systems—i.e., interactions between neurocomputational processes and socioenvironmental contexts unfolding over time (Boyd 1991, Fried & Cramer 2017, Kendler et al. 2011). These may differ greatly among the 500 MDD patients just described (Cai et al. 2020). The utility of essentialist thinking thus depends on the nature of the problem (Brick et al. 2021, McNally 2021).

For simplicity, we will hereafter refer to disorders as varying along a spectrum of essentiality, from high to low. Critically, this term is only meant as a shorthand for the utility of essentialist thinking (i.e., the psychological process; Gelman 2004) about a problem. It is not a claim that some or all mental health problems have essences, for instance. In Section 2, we suggest three heuristics for estimating the essentiality of a mental health problem. We argue that many mental health problems may have modest or fairly low essentiality. Essentialist thinking is not helpful for such problems because interdependent, temporally extended interactions partly constitute them (McNally 2021), and essentialist thinking obfuscates the importance of these interactions. In Section 3 we review developments in computational psychiatry and adjacent fields that move us toward capturing the dynamic interactions of even medium- and low-essentiality problems by modeling time and context. Note that throughout we focus on examples rather than offering a comprehensive review due to citation limitations.

2. THREE HEURISTICS FOR ESTIMATING THE ESSENTIALITY OF A MENTAL HEALTH PROBLEM

This section develops three heuristics for estimating the essentiality of mental health problems. Figure 1 shows estimates of essentiality for some well-known mental disorders. Note that an estimate is just an estimate; it is subject to change as more is learned. Moreover, each heuristic alone provides only limited information about a disorder's essentiality; the heuristics should be combined to triangulate on an estimate. Figure 2 depicts the three heuristics.

A challenge in estimating essentiality is that poor phenotyping can make a problem appear to have lower essentiality than it truly does (e.g., due to lack of understanding or misclassification). A well-established aim of computational psychiatry, closely aligned with initiatives such as the RDoC, is to improve phenotypic precision (Redish & Gordon 2016). Computational psychiatry offers powerful tools to build bridges between phenotypes defined by the current diagnostic systems and an emerging neurocomputational ontology (Poldrack & Yarkoni 2016). Ultimately, this may allow the current system of symptom-level descriptions to be partly reformulated as mixtures of neurocomputational processes (e.g., Drysdale et al. 2017) that have been refined through a combination of measurement innovation and theory (e.g., by employing computational modeling strategies and process-pure tasks that can reveal the differences underlying superficially similar symptoms and behaviors).

Yet, even if we could perfectly phenotype problems at any one point in time, we argue that there would still be a spectrum of essentiality. This is because the variability that we see among mental health problems is not due only to variability in how well we currently understand them (i.e., in our current knowledge of the underlying processes and our way of clustering these processes). The problems themselves can also have what we call meaningful heterogeneity. This is heterogeneity that arises due to the interdependence of the elements that constitute the problem, which makes it difficult to classify them at any one point in time and out of context (Lydon-Staley et al. 2021, Nelson et al. 2017). The three heuristics described in this section are meant to illustrate the indicators and practical consequences of meaningful heterogeneity through a series of examples.

In particular, we consider Parkinson's disease, schizophrenia, and MDD as running examples of high-, moderate-, and low-essentiality disorders, respectively. To situate this discussion, we draw on the neurocomputational functions of corticostriatal circuitry and dopamine (DA) in decision making, motivation, and reinforcement learning and on how dysfunctions or alterations in this circuitry relate to mental health (Maia & Frank 2011). We introduce each section with one or two questions to frame the discussion.

2.1. Neurobiological Mechanism Heuristic

Does a single, well-specified neurobiological mechanism cause the mental health problem? Would repairing it resolve the problem?

High-essentiality problems are caused by impairment of a specific, core neurobiological mechanism, beginning in a well-defined temporal window and leading to the disorder's primary signs and symptoms. Note that a single neurobiological mechanism can lead to more than one neurocomputational dysfunction (see Section 2.1.1 and the sidebar titled What Does Dysfunction Mean in a Mental Health Context?). The paradigmatic example of a clear biological etiology and resulting neurobiological impairment is general paresis of the insane, today known as late-stage syphilis. In the early twentieth century, this disorder was famously discovered to be caused by the spiral-shaped bacterium *Treponema pallidum*, which produces frontotemporal atrophy. This raised the prospect that simple etiologies would soon be found to underlie many mental health problems (Kendler 2005). More than a century later, however, this appears quite unlikely; as Kendler (2005, p. 433) has noted, "we can expect no more 'spirochete-like' discoveries." Although most mental health problems are more etiologically complex than general paresis of the insane, there still appears to be substantial variation in the extent to which they are characterized by a core neurobiological mechanism.

2.1.1. Parkinson's disease.—Parkinson's disease is a relatively high-essentiality disorder that involves the progressive denervation of DA neurons of the substantia nigra, preferentially targeting dorsal striatum of the basal ganglia (BG) early in the disease (Cools et al. 2001). In computational models, a healthy dynamic striatal DA range is required for adaptive action selection and reinforcement learning. Chronic DA depletion in Parkinson's disease leads to a bias toward learning more from negative than from positive reward prediction errors (RPEs; Wiecki & Frank 2010). DA medications reverse these biases by restricting DA levels to an artificially high range, preventing the DA "dips" that normally accompany negative RPEs, as captured by computational modeling (Frank 2005). Confirming model predictions, relative to healthy controls, unmedicated Parkinson's disease patients showed impaired learning from positive RPEs but relatively enhanced learning from negative RPEs; medications reversed this bias, impairing learning from negative RPEs (Frank et al. 2004). This pattern may explain some of the adverse effects of DA medications, such as impulsivity, and has been replicated at least 15 times (some of which are reviewed in Collins & Frank 2014).

Other Parkinson's disease sequelae arise as a consequence of this core pathology. This pattern is common to many high-essentiality problems: A core neurobiological mechanism can lead to multiple neurocomputational dysfunctions. In Parkinson's disease, dopamine depletion affects not only the motor striatal circuits but also those interacting with the prefrontal cortex (PFC). Accordingly, in the computational models, this mechanism alters gating not only of motor actions but also of cognitive ones, such as the entrance of cortical content into working memory. Empirical work confirms that there are parallels in how motor actions and working memory content are gated, and that these are related to striatal DA mechanisms in Parkinson's disease (Salmi et al. 2020, Wiecki & Frank 2010). Within a given corticostriatal circuit, DA depletion also induces hyperactivity of the subthalamic nucleus (STN). According to the computational model, this hyperactivity leads to elevated decision thresholds for initiating actions, which is separate from the effect of DA on weighting costs versus benefits (Frank et al. 2007). Indeed, deep brain stimulation

of the STN reduces the decision threshold and partially remediates motor deficits, but it can accordingly lead to a distinct sort of impulsivity, preventing patients from adaptively elevating the decision threshold when needed for cognitive control (Cavanagh et al. 2011, Frank et al. 2015, Herz et al. 2016). Thus, the same computational model ties together several cognitive, motivational, and motor sequelae of Parkinson's disease resulting from a core neurobiological mechanism: DA denervation in the BG. The model therefore suggests how varying rates of dysfunction in these pathways can help to explain Parkinson's disease subtypes, such as those where gait freezing predominates (Matar et al. 2019).

2.1.2. Schizophrenia.—Schizophrenia is a middle-essentiality problem in which DA has long been implicated (McCutcheon et al. 2020). Indeed, many of the disorder's positive symptoms can be accounted for by spontaneous striatal DA fluctuations that assign meaning to irrelevant events (defined as aberrant salience; Kapur 2003), and many negative symptoms can be explained by weaker adaptive DA responding to motivationally significant events (Gold et al. 2015, Maia & Frank 2017). Yet, it is clear that dysregulated striatal signals alone are an insufficient account of schizophrenia; much evidence also implicates PFC dysfunction that leads to context-inappropriate behavior (Cohen & Servan-Schreiber 1992). In a formal model of the complementary contributions of BG and PFC, an extended neural network includes PFC layers that maintain stimulus-outcome associations in working memory "attractor states"; these afford specific representations about the expected values of stimuli and actions as well as rapid adjustment to recent outcomes (Frank & Claus 2006). Experiments disentangling these contributions with quantitative modeling revealed that schizophrenia patients mostly struggled with PFC-like computations (e.g., reduced contributions of working memory and expected value, reduced top-down biasing of learning), with relatively spared incremental reinforcement learning from RPEs (e.g., Collins et al. 2017, Geana et al. 2021, Gold et al. 2012). This conclusion is also supported by neuroimaging (Dowd et al. 2016) and is consistent with other dynamical systems models of deficient attractor states in schizophrenia (e.g., Durstewitz & Seamans 2008).

2.1.3. Depression.—MDD is a relatively low-essentiality problem in which a wide range of neurocomputational differences have been noted, including alterations in reward processing and cognitive control tasks, experience of more negative emotions, and proneness to self-referential, ruminative thinking (Goldstein & Klein 2014, Kaiser et al. 2015, Keren et al. 2018, Snyder 2013). Yet, in contrast to Parkinson's disease, where there is a focal pathological aberration of midbrain DA neurons, the processes implicated in MDD develop over a long time and in close interaction with one another. Depression also constitutes a heterogeneous phenotype (Fried & Nesse 2015): Differences documented at the group level are not reliably present among individual patients (e.g., Webb et al. 2016).

Critically, it is unclear which observed alterations in MDD should be thought of as dysfunctional (as opposed to adaptive) in light of other alterations and of social and environmental factors. For example, rumination has been consistently associated with depression (reviewed in Nolen-Hoeksema et al. 2008). Neuroimaging studies confirm altered activity patterns in depression in many areas implicated in self-referential processing and attentional control (Kaiser et al. 2015). These patterns are sometimes interpreted as aberrant,

yet it is unclear what distinguishes maladaptive from adaptive repetitive thinking about oneself (but see Watkins 2008 for one delineation). Intuitively, intense and protracted thinking can be important after a serious setback to one's life plans. Stressful life events tend to precipitate MDD (Kendler et al. 2000); hence, it is unclear where to mark the boundary between dysfunctional thinking (Dayan & Huys 2008) and constructive thinking that helps to resolve problems, facilitate recovery, and elicit support (Andrews & Thomson 2009). Similarly, depressed individuals on average show performance decrements in cognitive control–demanding tasks (Snyder 2013). Yet, operating from a computational perspective on cognitive control allocation, Grahek and colleagues (2019) have emphasized that merely observing a difference in a control-demanding task is uninformative about whether the difference emanates from dysfunction per se or from learned control-allocation decisions. For example, control may be allocated to self-directed mentation if such thinking is valued (see also Agrawal et al. 2020, Andrews & Thomson 2009), and decreased control could be rationally learned from action-outcome statistics (Lieder et al. 2013, Shenhav et al. 2013). To the experimenter's eye, these learned differences—products of a properly functioning control system—would (typically) lead to a performance pattern indistinguishable from cognitive control dysfunction (Grahek et al. 2019).

In sum, research points to a relatively specific core dysfunction in Parkinson's disease, whereas schizophrenia arises from a more complicated interaction between striatal and PFC dysfunction and other interrelated neurocomputational processes (reviewed in Valton et al. 2017). MDD involves an even more complicated set of alterations, many of which are difficult to interpret out of context (e.g., whether the alteration helps or harms in coping with recent life stress).

2.2. Variable Trajectory Heuristic

Would the problem manifest in the same way irrespective of neurocomputational and social and environmental context?

High-essentiality problems follow a stereotyped natural course (absent intervention), whereas low-essentiality problems involve the contingent interactions of neurocomputational and social and environmental processes over time. This makes it difficult to predict the specific trajectory of such problems (Henry et al. 2020). This heuristic thus concerns a continuum along which problems fall: from following an ordered and linear progression to comprising interacting elements that lead to ramifying trajectories over time.

At the heart of this heuristic is the degree of multifinality—that is, the extent to which the same predisposing constellation of factors leads to divergent outcomes (Cicchetti & Rogosch 1996). For instance, a bias to attend to negative information has been implicated as a risk factor for various internalizing disorders, yet it is unclear why one individual develops obsessive-compulsive disorder whereas another develops MDD. One reason multifinal problems are challenging to model is that the causes of mental unhealth appear at different causal distances from the acute onset of the problem. Heuristically, these can be classified into distal versus proximal factors (i.e., things that happen to people, such as having certain genes or having experienced child abuse, versus things that vary over time within individuals, such as one's current propensity to ruminate or tolerance for ambiguity) and

moderators that determine exactly how a problem unfolds (e.g., a problematic behavior crystallizing into a strong habit; Nolen-Hoeksema & Watkins 2011). In lower-essentiality problems, the dynamic interrelations between these elements, which are operative at different time scales, partly constitute the problem itself (McNally 2021). For instance, in MDD, processes such as negative schemas, rumination, cognitive control differences, interpersonal stress, and a conflict-laden work environment can mutually reinforce each other (Fried & Cramer 2017, Kendler et al. 2011).

In contrast, for higher-essentiality problems, there is a more direct path from distal risk factors to core neurobiological mechanism, concomitant dysfunction(s), and resulting symptoms. For instance, in contrast to many mental health problems, single-gene mutations confer strong risk for Parkinson's disease (though note that various genes leading to somewhat different etiologies are implicated, hence Parkinson's disease may be further subtyped eventually; Weiner 2008). The hallmark of Parkinson's disease is denervation of DA neurons, leading to well-characterized problems that follow a fairly ordered progression over time. It is important to note that even this relatively high-essentiality disorder is dependent on the social milieu and environment. This follows from the aforementioned findings that DA denervation in Parkinson's disease leads to exaggerated learning from negative outcomes (in the unmedicated state; Frank 2005). In addition to having direct detrimental effects on motor performance, this denervation can induce progressive aberrant learning that amplifies symptom progression in a context-dependent fashion (Beeler et al. 2012). It is noteworthy that some degree of social and environmental dependence is present even toward the farthest end of the essentiality spectrum, such as in Huntington's disease, which has a single genetic cause but for which it is nonetheless unclear when symptoms will manifest (Wiecki et al. 2016).

In schizophrenia, there appears to be a more temporally extended and interactive pathway to disorder development. Schizophrenia involves distal risk factors, including a complex suite of genetic risk factors that are thought to be at least partly responsible for cognitive impairments that become evident over childhood and adolescence (McCutcheon et al. 2020). Stress caused by difficulties in functioning due to these impairments, and compounding factors such as childhood abuse, familial stress, and social marginalization (Egerton et al. 2016), are thought to alter the function of the striatal DA system by adulthood (McCutcheon et al. 2020). As noted, altered striatal DA signaling may serve to imbue irrelevant events with salience (via spontaneous firing) and to prevent appropriate responding to relevant events (via lower adaptive firing; Maia & Frank 2017). Disorganized and inappropriate responding resulting from these dysfunctions may in turn promote social ostracism and fuel the development of idiosyncratic beliefs, such as negative views about oneself and one's abilities, leading to emotional symptoms and further functional impairment (Perivoliotis et al. 2009).

MDD (and other internalizing disorders with which it is highly comorbid) appear to show an intricate interdependency with the social and environmental context and to be highly dependent on the formation of specific beliefs. Strikingly, the genetic correlation between MDD and generalized anxiety disorder (GAD) has been estimated at 1 in women (and 0.74 in men), implying that nongenetic (e.g., socioenvironmental) factors play a crucial

role in determining the unique symptoms of these problems (Kendler et al. 2007). Indeed, there appears to be some specificity in the relationship between life stress experienced and resulting symptoms, with humiliating events showing a stronger relationship with MDD and danger showing a stronger relationship with GAD (although loss is comparably associated with both and with mixed presentations; Kendler et al. 2003).

Hammen (2005) has emphasized that stressful life events include not only independent stressors (e.g., losing one's spouse) but also dependent stressors (events in which individuals play a role, e.g., fighting with one's spouse). This suggests a transaction between depression risk factors and stress-generating behavior in challenging situations. For instance, rumination and worry among individuals prone to MDD and GAD may disrupt reinforcement learning about external contingencies (Hitchcock et al. 2021, Whitmer et al. 2012). Because rumination involves accessing negative memories within a negative affective context, it may also make negative memories more accessible in the future (e.g., Cohen & Kahana 2020, Van Vugt et al. 2012). Hence, rumination may simultaneously increase the future availability of negative thoughts and decrease the chance of adaptively behaving in similar (external) situations in the future (see Hitchcock et al. 2021 for discussion). Depending on what outcomes this leads to, different symptoms could result. For instance, an individual who experiences substantial humiliation may develop depression symptoms, whereas someone who finds themselves in ensnaring or dangerous situations could develop general anxiety symptoms (Kendler et al. 2003). This latter possibility may be especially likely if the individual becomes pessimistic about their ability to act safely in general (Zorowitz et al. 2020). Longitudinal investigation confirms that there is a complex interplay between the tendency to ruminate, impaired performance in control-demanding activities, dependent stress generation, and subsequent depression and anxiety symptoms (Snyder & Hankin 2016). As we discuss in Section 3, we think these complex interactions imply that time and context must be more fully incorporated into computational psychiatry models if we are to predict and model precisely problems such as MDD and GAD.

2.3. Relevance of Intentional Content Heuristic

Is mental content about something (such as beliefs and values) critical to the problem? Is intervening on such content an important lever to intervene on in the problem?

Mental health problems vary in the importance of intentional content: content that is about something, such as a belief about oneself, the significance attributed to a personally meaningful event, or a value about how one ought to live. This heuristic thus concerns the extent to which such content is central or peripheral to a mental health problem. For example, consider Parkinson's disease and MDD. A Parkinson's disease patient will experience substantial functional and occupational impairment as the disorder progresses, which may lead to negative views about themselves. Changing these beliefs may assist in this person's ability to cope, but it will not fix the root problem: midbrain DA denervation. In contrast, negative views about oneself are arguably core to MDD; they partly constitute the problem (Kendler et al. 2011). Evidence-based psychotherapeutic interventions specifically target such negative schemata and can lead to considerable improvement.

As another example, consider a soldier who unintentionally killed a civilian in combat (see Litz et al. 2009). Trauma-informed guilt reduction (TrIGR) psychotherapy guides clients who have incurred guilt from these kinds of experiences to reinstate the event's complete context: distinguishing the knowledge they had at the time from that which they accrued later; recalling which actions were actually available then (rather than which actions they wish had been available); and identifying their specific responsibility (which typically reveals that their actions were embedded in a complex causal chain). Elaborating the context of such an experience with a psychotherapist may not bring full relief, but it can help to move a client from seeing themselves as deserving of unrelenting and lifelong shame toward living consistently with their values now (Norman et al. 2014).

An individual who has experienced an event or set of events that challenged their values and moral sense (sometimes referred to as moral injury) may report mental health symptoms (e.g., low mood, lost motivation, shame and guilt; Litz et al. 2009). Finding the best lever (Redish & Gordon 2016, p. 19) for intervening on these symptoms would probably require understanding the injurious memory and the beliefs that have developed around it; this would seem especially plausible if dialogue (via TrIGR, for example) improved the person's symptoms. Of note, such an intervention undoubtedly would change memory and judgment engrams distributed through the person's brain (and, eventually, larger-scale neural circuits). Yet, there is no reason to think that the specific details of the neural instantiation of these engrams would be especially interesting. A more useful level of analysis for understanding this person's difficulties is at the level of their specific memories, judgments, and beliefs (Eronen 2019, Kendler 2005). By analogy, if I want to convince someone that I have a blue bandanna in my closet, I will almost assuredly have more success if I tell them as much directly rather than if I try to manipulate their brain. Similarly, when the causal loci of a mental health problem involve specific intentional mental content, intervening on such content (Eronen 2020b) may be the most direct route to effecting change.

A perhaps underappreciated point in computational psychiatry is that computational theories can inform clinical principles relevant to intervening on intentional content. For instance, inverse-planning models formalize theory-of-mind inferences about an agent's goals and objectives from their actions in situations (Baker et al. 2017); potentially, such models could elucidate how one draws inferences about one's own actions (see Gillan et al. 2017 for a similar proposal). Understanding the computational costs of different action-selection strategies can help to explain how factors such as time pressure and proximity to threat mandate the use of fundamentally different ways of responding (Mobbs et al. 2020). This could help to explain why, when they are under pressure, people act in ways that are fundamentally different from the values they espouse when they have more time to reflect. The computational expense of certain ways of thinking might also help us understand why we tend to save (amortize) costly computations for later reuse (Dasgupta & Gershman 2021), possibly including inferences about our own character made under or in the wake of duress. In fact, this may even help to explain why we tend not to recompute past inferences unless we have a strong motivation to do so—indeed, why we may not do so even if we have since acquired relevant new information (an observation that has puzzled many a psychotherapist who has observed their client express flatly contradictory beliefs that were formed in different contexts).

Of note, moral injury provides a particularly clear example of the relevance of intentional content in mental health, yet beliefs, self-judgments, perceived violations of values, and other types of intentional content are core to many mental health problems (see also Gu et al. 2019). That intentional content is especially important in lower-essentiality problems follows from the two previous heuristics. Lower-essentiality problems do not involve a core mechanism that leads to generic neurocomputational deficits, but rather they comprise individual differences transacting with social and environmental contexts over time. Such contexts, rather than dysfunctions or neurocomputational propensities alone, partly determine which mental health elements will arise based on the conclusions that people draw (i.e., the intentional content that emerges) in such situations.

2.4. Concluding Thoughts on Our Three Heuristics for Estimating Essentiality

We offered three complementary heuristics for estimating the essentiality of a mental health problem: whether a single and specific neurobiological mechanism is core to the problem; whether the problem follows a straightforward natural course or is characterized by divergent trajectories (multifinality); and whether intentional mental content (beliefs, values, etc.) are core or peripheral to the problem. Note that although we used diagnostic categories in our running examples for familiarity, essentiality could be estimated for more granular representations (e.g., endophenotypes), subsuming representations (e.g., higher-order factors; Kotov et al. 2017), or multidimensional profiles (Wiecki et al. 2015) or "biotypes" (Drysdale et al. 2017) if these are consistently replicated and refined in a way that enables categorization. For this reason, we refer throughout to "mental health problems" for simplicity and generalizability.

3. NEW METHODS TO MODEL LOWER-ESSENTIALITY PROBLEMS IN COMPUTATIONAL PSYCHIATRY

An important challenge to estimating essentiality is the possibility that a disorder may only appear to have low essentiality due to poor phenotyping (i.e., improper clustering and superficial understanding), and that perhaps it would be possible to derive a higher-essentiality disorder (or disorders) through improved phenotyping. Enhancing phenotypic precision is critical to continued progress in computational psychiatry, and in the current context it is key to avoiding confounds in estimating essentiality. Section 3.1 reviews efforts to improve phenotypic precision in computational psychiatry (Figure 3).

However, even if we reached perfect phenotyping, there would still likely be a spectrum of essentiality, because many mental health problems are characterized by meaningful heterogeneity: that is, heterogeneity that arises from the interdependency of the elements constituting the problem, which confounds attempts to categorize the problem at any single point in time and without an understanding of the context in which it arose. Sections 3.2 and 3.3 focus on modeling dynamics unfolding in context over time to tame meaningful heterogeneity (Figure 4).

3.1. Refining Phenotypes

A key step toward more precise phenotyping is discovering (possibly high-dimensional) clusters of neurocomputational alterations. There are a few strategies for discovering such clusters (see also Maia & Frank 2011): top-down (from the diagnostic systems to neurocomputational processes), bottom-up (working from well-defined neurocomputational processes to mental health phenomena), and intermediate (e.g., using data-driven approaches to summarize questionnaire-based data from the diagnostic systems and then relating these summaries to neurocomputational processes).

3.1.1. Top-down approaches.—A number of computational psychiatry studies have taken steps to move beyond diagnostic categories. One strategy is to report differential relationships between neurocomputational processes and specific symptoms. Beevers and colleagues (2019) reported that estimated drift rate (a rate parameter in computational models that assume information is sequentially sampled over time) for negative words in the self-referential encoding task strongly related to depression symptoms such as sadness and self-dislike, yet it only weakly related to symptoms such as feeling like a failure, crying, and lost appetite. A symptom-centric approach may be particularly valuable for poorly phenotyped problems such as MDD (i.e., those with very different risk factors, neurobiological correlates, relationships to functional impairment, etc.; Fried & Nesse 2015). Diagnostically minded theorists have also emphasized that there is special value in understanding the processes that underlie hallmark (disorder-specific) symptoms, because they carve phenotypic space at its joints (Spitzer et al. 2007). For instance, from a nosological perspective, there may be special value in understanding flashbacks in post-traumatic stress disorder (PTSD) due to their specificity to this disorder, whereas symptoms such as negative beliefs about oneself and the world are much less specific to PTSD.

Another approach that begins with the diagnostic categories is to use common clusters of symptoms. For instance, Brown and colleagues (2018) reported that amygdalar activity evoked by computational-model-derived associability (i.e., increased attention proportional to prediction error, here specifically in a loss condition) was more related to avoidance/numbing and hyperarousal than reexperiencing symptom clusters of PTSD. Note, however, that obtaining replicable symptom clusters for common mental health problems has been challenging (e.g., Armour et al. 2015).

3.1.2. Bottom-up approaches.—A fundamental challenge to top-down research that begins with the DSM diagnostic system is that the signs and symptoms collected in this manual were deliberately described at a superficial level rather than in terms of underlying processes. The aspiration was to enable reliable diagnosis by clinicians of different theoretical orientations who disagreed about the underlying processes (Wakefield 1992a). However, a critical aim for psychopathology science, including computational psychiatry, is to move beyond such superficial descriptions. Computational cognitive neuroscience offers powerful tools for fractionating into primitive units processes that were previously subsumed under an aggregating construct. Computational psychiatry seeks to fractionate the processes specifically relevant to mental health (Maia & Frank 2011); that is, it takes a bottom-up approach that begins with well-defined processes and relates these to mental

health phenomena. Underscoring the importance of this endeavor, many symptoms within the current diagnostic manuals (and constructs in the wider psychopathology vernacular) are turning out to be “suitcase terms”—terms that obscure precise distinctions (Minsky 2007). For example, anhedonia, a cardinal symptom of MDD that is also present (or similar to symptoms described) in numerous other mental health problems (McCabe 2018), involves distinct components, only some of which are altered in MDD (Huys et al. 2013, Keren et al. 2018, Treadway & Zald 2011). Similarly, impulsivity can arise from a variety of mechanisms, including valuation asymmetries related to striatal DA (Frank 2005), alterations in decision-threshold activity during conflict via PFC-STN interactions (Frank et al. 2007), and differences in how future rewards are discounted (McClure et al. 2004). Once such decompositions are confirmed, they should influence our strategies with top-down phenotypes; for instance, the discovery that individuals with attention-deficit/hyperactivity disorder could be distinguished by type of impulsivity can help to stratify pharmacological approaches. Ultimately, we will likely need dynamic, quantitative, and aggregative methods to iteratively refine our diagnostic systems, especially if the pace of discovery of strongly supported mental health–relevant decompositions quickens. Emerging data-driven neurocomputational ontologies offer inspiration (Poldrack & Yarkoni 2016).

3.1.3. Intermediate approaches.—An intermediate strategy is to begin with questionnaires related to diagnostic categories (i.e., problems or symptoms commonly seen in patients with a specific disorder) but then use dimension reduction techniques such as factor analysis to derive data summaries that cut across diagnostic symptoms, which can then be related to neurocomputational processes (e.g., Gillan & Daw 2017, Gillan & Seow 2020, Gillan et al. 2017). Studies using this approach have reported specificity in neurocomputational processes associated with distinct regions of phenotypic space (e.g., Gillan et al. 2016, Rouault et al. 2018). For instance, Rouault and colleagues (2018) found, using computational modeling applied to a perceptual decision-making task, that individuals who endorsed more compulsive behavior and intrusive thoughts (based on a data-driven summary factor with transdiagnostic symptoms including schizotypal symptomatology) were more confident in their choices, yet poorer in their ability to discern which choices were actually correct; by contrast, individuals endorsing more depression and anxiety symptoms (based on another factor including apathy symptoms) showed the opposite pattern: less confidence but relatively higher discernment of which choices were correct (Rouault et al. 2018). Parallel to these developments in computational psychiatry, efforts are underway in clinical science more broadly to delineate relations among symptoms and disorders transdiagnostically, such as the HiTOP (Kotov et al. 2017).

This intermediate approach is not without challenges. For one, dimensional summaries depend (of course) on the questionnaires they are summarizing. To establish factor structure replicability, computational psychiatrists have tended to use questionnaires similar to the ones employed in an original set of studies by Gillan and colleagues (reviewed in Gillan & Seow 2020), yet these may not encompass all processes of interest (see Watts et al. 2020 for an interesting perspective on this issue). Gillan & Seow (2020) noted therefore that dimensions from prior studies (and the questionnaires from which they are constructed) must be iteratively refined to enable continued progress. Other challenges relate to interpretational

and measurement challenges that arise whenever symptom questionnaires are used. Symptoms can covary for a number of reasons, and the methods that find dimensions based on symptom covariation often provide little insight into the data-generating mechanisms behind the covariation (Bringmann & Eronen 2018). For instance, symptoms can correlate due to a common cause (e.g., sweats and aches arising from a fever) or because one symptom causes another (e.g., worry causing insomnia; Borsboom 2008, Kendler et al. 2011). They can also covary for more artificial reasons, such as semantic overlap among items (e.g., feeling sad, feeling blue, and feeling depressed in a prominent depression scale; Fried & Cramer 2017), response styles that have nothing to do with questionnaire content (e.g., tending to answer "strongly agree"), and implicit theories (e.g., guessing that one is answering a questionnaire about depression; Podsakoff et al. 2012). Identifying and extracting components, factors, or dimensions from such instruments thus does not by itself establish reliable or valid intermediary phenotypes between symptoms and disorders (see Leising et al. 2020 for an accessible overview of some of these issues).

In sum, bottom-up, top-down, and intermediate strategies have a natural synergy; each approach has limitations, but they also have complementary strengths and weaknesses. It is also worth noting that algorithmic computational models in computational psychiatry play a special bridging role in that they can connect clinical phenomena and observations to biologically realistic models. Yet, algorithmic models too have limitations and require substantial caution (see Supplemental Text). A more fundamental challenge than any of these particular limitations is that only so much progress can be made by refining static and decontextualized phenotypes, due to the challenge of meaningful heterogeneity (Figure 4). The next sections review emerging developments for incorporating time and context in order to tame this heterogeneity, and thereby expand the dimensionality of our models to a space within which even low-essentiality problems reside.

3.2. Capturing Domain-Specific and Time-Varying Phenomena in the Real World

We have argued that rather than arising from a core neurobiological mechanism, lower-essentiality problems comprise dynamically changing neurocomputational processes interacting with situations and social milieus encountered over time. This calls for an expansion of the focus of computational psychiatry away from looking exclusively for trait-like dysfunctions and toward understanding time-varying alterations in context (see also Radulescu & Niv 2019, Scholl & Klein-Flugge 2018).

3.2.1. Modeling state variation.—Many mental health problems are far from static; they follow stages or exhibit oscillations and change and transact in important ways with social and environmental contexts. Addiction, for example, has been described as following distinct stages, and neurocomputational processes may vary dynamically by stage, while possibly retaining an invariant multidimensional structure (Gueguen et al. 2021). A neurocomputational account of bipolar disorder produces oscillations whereby mood and reward appraisal interact in a positive feedback loop (Eldar & Niv 2015, Mason et al. 2017). MDD (and possibly many other internalizing disorders) is both precipitated by life stress and associated with stress-generating behavior (Hammen 2005), possibly due to a complex

interplay between dynamically changing propensities and stressful experiences (Hitchcock et al. 2021, Snyder & Hankin 2016).

Time-varying phenomena present a challenge to task assays performed at one cross-section in time, as these are predicated on the assumption that the processes under study are stable (i.e., trait-like; Rodebaugh et al. 2016). However, if time-varying phenomena can be harnessed, they present opportunities, in that phenomena that signal transition points in mental health could be detected for prediction and intervened upon for prevention. Exemplifying this possibility, Konova and colleagues (2020) administered a task, which distinguished comfort with known risk (via monetary gambles where the probabilities were known) from unknown risks (via monetary gambles where probabilities were partially occluded), up to 15 times over a period of 7 months to individuals receiving community treatment for opioid use. Using computational modeling, the researchers estimated individual propensities to take known and unknown risks and submitted these as one-time-back predictors in logistic regression models predicting opioid use. They found that tolerance for unknown (i.e., ambiguous) risks alone significantly predicted subsequent use. This result was especially compelling because data were collected from a parallel cohort of healthy controls, among whom the model-derived predictors were relatively stable over time; by contrast, the predictors' stability was lower among the individuals struggling with opioid use, likely due in part to meaningful variation that facilitated prediction (Konova et al. 2020).

3.2.2. Incorporating domain-specific stimuli or contexts.—Another method for understanding neurocomputational differences in context is to use domain-specific stimuli or contexts rather than generic (e.g., fractal) stimuli. Frey and colleagues found that individuals with elevated depression symptoms showed slower incremental learning in two social tasks: one that involved picking items for a party and then seeing how each item was judged by other (putative) participants (Frey et al. 2021), and another that involved gradually learning how happy or fearful different people tended to be by repeatedly guessing each person's emotion and then seeing them make a neutral or happy/fearful face (Frey & McCabe 2020b). Those who were slower to learn in the first study also reported spending more time quarreling or engaging in other unpleasant social activities in their everyday lives (Frey et al. 2021). Another interesting finding by this research group was that, in the face-learning task, nondepressed participants who underwent serotonin depletion showed similar patterns of sluggish learning and altered neural activity as the depressed participants (Frey & McCabe 2020a).

One limitation of these studies is that they did not directly compare social and nonsocial contexts, making it difficult to determine whether participants were characterized by a generic decision-making alteration or one specific to social settings (see Pulcu & Browning 2017). Addressing this issue, Lamba and colleagues (2020) investigated behavior in a game where participants received an initial monetary endowment and invested portions of it on a trial-wise basis with (they were told) a human partner or slot machine, which would subsequently return varying amounts; they were told the human participant would receive quadruple the invested amount before apportioning the return. In reality, the amount that the human partner/machine returned was rigged and drifted slowly over time, mimicking

real-world situations in which fortunes or attitudes change gradually (such as a job interview that takes a slow but steady turn for the worse). Participants across a spectrum of generalized anxiety symptoms struggled to stop investing in slot machines that began shorting them on returns; however, lower-anxiety participants rapidly adjusted when their human partners did the same, possibly reflecting a swift ability to detect exploitation in this social context. By contrast, higher-anxiety participants were similarly slow to adjust investments to human partners who became more miserly as they were to adjust to slot machines. The use of matched social and nonsocial contexts allowed the researchers to conclude that the difficulty in responding to gradual uncertainty among anxious participants was (mostly) specific to the social domain (Lamba et al. 2020).

3.2.3. Connecting lab-based observations to real-life behavior.—

Complementary to research that brings idiosyncratic and ecologically valid stimuli into the lab is work that relates lab-observed differences to behavioral variation in everyday life. Eldar and colleagues (2018) reported a tour-deforce example of how to connect modeling, real-world behavior, and multimodal measurement. In their study, ten individuals completed a reinforcement-learning task twice per day on their smartphones while portable systems recorded electroencephalography and heart-rate data. Computational modeling revealed individual differences related to dissociable fast and slow learning processes: Participants with stronger neural decodability of the fast-learning process (according to machine-learning methods) showed an improvement in their mood a few hours later, whereas those with stronger decodability of the slow-learning process showed higher mood the following day (Eldar et al. 2018).

In general, smartphones offer an unprecedented opportunity for so-called digital phenotyping, including high-frequency or even ubiquitous collection of certain types of mental health–relevant data with minimal participant burden (see Gillan & Rutledge 2021 for an authoritative review).

3.2.4. Understanding alterations in context.—

A theme of this section has been the importance of understanding empirically observed neurocomputational alterations in context, rather than merely documenting that an alteration exists. One area of computational psychiatry in which a shift has been evident in how to interpret observed differences is the investigation of model-free versus model-based strategies in reinforcement learning. Briefly, model-free reinforcement-learning algorithms are those that solve trial-and-error learning tasks without an explicit representation of the world, whereas model-based strategies represent aspects of the world such as reward distributions and transition probabilities. An impactful set of studies used the so-called two-step task (Daw et al. 2011) to infer participants' model-free and model-based propensities. Early studies suggested that a tendency to employ model-based control emerges over development (Decker et al. 2016) and implicated decreased model-based control in obsessive-compulsive disorder (Gillan et al. 2015) and compulsive decision making broadly (Gillan et al. 2016). This seemed to imply that a trait-like and domain-general propensity toward model-free over model-based control contributes to faulty decision making and psychiatric disorders. This may be correct to an extent, but recent work has also shifted the focus toward understanding how different

contexts and goals influence the type of strategy used.[1] This includes theoretical accounts that implicate incorrect model-based reasoning in depression (Huys et al. 2015) and suggest a spectrum of model-free to model-based reasoning depending on the speed under which a decision must be made (e.g., Keramati & Smittenaar 2016). A study involving the two-step task showed that people increased model-based control when incentivized to do so, cutting against the notion of a fixed capacity; surprisingly, the researchers also found that individuals high on sensation seeking and on an anxious-depressed dimension were especially responsive to incentives to use model-based control (Patzelt et al. 2019). In a reinforcement-learning task with a social framing, Hunter and colleagues (2019) found that individuals with elevated social anxiety symptoms showed increased model-based control specifically in response to "upward-counterfactual" feedback (Hunter et al. 2019). Finally, building on behavioral neuroscience research, Mobbs and colleagues (2020) argued that the same animal will tend to employ a spectrum of strategies depending on its proximity to threat: from hardwired responses when threat is extremely close to multi-step, model-based reasoning when threat is very far. Overall, this recent work reflects a shift in emphasis toward the differential use of model-free versus model-based strategies based on demand and context.

3.3. Measuring Dynamics and Person-Specific Processes and Developing Formal Mental Health Systems

This section reviews methods for modeling temporal and within-person dynamics, which we have argued are especially important in medium- and lower-essentiality problems (see also Gillan & Rutledge 2021, Huys et al. 2021, Scholl & Klein-Flugge 2018).

3.3.1. Modeling dynamics.—Recent frameworks that conceptualize mental disorders as complex systems of interacting processes have developed novel network methods to model dynamic changes to mental health over time (Beltz & Gates 2017, Borsboom 2008, Bringmann et al. 2013, Fried & Cramer 2017, McNally 2021, van de Leemput et al. 2014). These network models are statistical representations of node-and-edge relationships between mental health elements (most commonly symptoms, although other variables are increasingly incorporated; Fried & Cramer 2017). These elements are often assessed by self-report; hence, they are subject to similar limitations as those mentioned above in the context of intermediate approaches. This includes that the methods typically provide only weak information about the structure of mental health problems (Bringmann & Eronen 2018).

Notwithstanding these modeling limitations, the network approach has drawn important attention to the ontology of mental health (McNally 2021). Additionally, recent network modeling developments may provide more information about the structure of mental health problems and potentially point to novel intervention targets. These include recent methods that leverage control theory to attempt to infer the most controllable node within a network, which could be a fruitful target for psychotherapy (Henry et al. 2020).

[1]Note that in their earliest work Daw and colleagues (2005) already emphasized that context should normatively influence the strategy used.

Predictability methods estimate how well each node in a network can be predicted by all other nodes in terms of variance explained, potentially revealing how important a node (e.g., sleep difficulties) is within a broader system (e.g., depression). Moreover, the average predictability of all nodes in a network can (under some critical assumptions) provide insight into how well (or poorly) the included elements reflect the full system. For instance, a review of 18 network studies found that depression, PTSD, and anxiety had higher average predictability than psychosis, suggesting that some elements (possibly including a neurocomputational common cause) were not represented in the psychosis network (Haslbeck & Fried 2017). Methods from complex-system analysis could also aid our understanding of the structure and dynamics of various problems. These methods build on the properties of complex systems, such as their leaving signatures like autocorrelation and increasing variance near transition points, regardless of their specific constitutive elements. An influential paper argued that rising autocorrelation and variance among emotions signals a “critical slowing down” that augurs a depressed state, similar to critical transitions observed in fields such as ecology (van de Leemput et al. 2014).

In computational psychiatry, there is a rich tradition of modeling neural dynamics (recently reviewed in Durstewitz et al. 2020), yet there has been much less focus on the externally observable dynamic elements of mental health systems. A notable exception are the models developed by Eldar and colleagues that produce oscillatory dynamics (Eldar & Niv 2015, Mason et al. 2017). These frameworks model individual differences relevant to bipolar disorder via an interdependence between mood and evaluation. In this approach, a mood-biasing parameter (assumed to be trait-like) can produce dynamics such that perceived rewards sometimes far exceed expectations, leading to large positive surprises that send mood rocketing upward, and sometimes fall far short of expectations, leading in turn to crushing disappointments after reward omission that drive mood downward. Remarkably, the administration of a selective serotonin reuptake inhibitor (SSRI) appeared to modulate this parameter, leading rewards to be more impactful when in a good mood, and in turn further increasing mood. This might lead to a slow but steady increase in the proportion of felicitous experiences, eventually leading to greater well-being over time. Thus, this finding may help to explain the gradual effects of SSRIs as well as the increased susceptibility to mood instability that these drugs appear to induce among a subset of individuals (Michely et al. 2020). Computational psychiatry theories that predict these kinds of temporally extended dynamics offer a glimpse into how risky predictions concerning how elements of mental health systems interrelate can be derived and then tested on data collected in the real world —leading to an iterative refinement of model and theory (Figure 5). For instance, this model predicts trait-like individual differences as well as drug effects on mental health elements— expectations, subsequent gloomy and glorifying appraisals of surprising experiences, and domino effects on mood. These could be tested by applying network models (such as moderated network models; Haslbeck et al. 2019) to data reported by participants over time, in order to capture varying drug effects or between-subject trajectories related to the mood-biasing parameter.

3.3.2. Capturing person-specific processes.—Due to the divergent trajectories of lower-essentiality problems (i.e., multifinality), measuring, modeling, and understanding

person-specific patterns are especially important. One striking example of how person-specific patterns can dissociate from group-level patterns is Simpson's paradox—the fact that, for example, coffee consumption may perfectly positively correlate with neuroticism between subjects, even if the relationship is negative within subjects (i.e., these individuals become less neurotic when they consume coffee; Kievit et al. 2013). Such a possibility should trouble computational psychiatrists, because a tacit assumption in much task-based research is that finding an altered pattern between mentally unhealthy and healthy individuals (or groups) is the first step toward developing a remedial within-subject intervention. Notably, the fact that extrapolating from between-person to within-person patterns—or more generally from groups to subgroups, groups to individuals, or averages across time to temporal patterns (Kievit et al. 2013)—can lead to misleading conclusions appears to be of more than theoretical concern, with a recent computational psychiatry study providing an interesting example. As mentioned above, a longitudinal investigation by Konova and colleagues (2020) found that opioid use could be predicted by a one-time-back measure of tolerance for ambiguous risk. On average between groups, however, a quite different pattern emerged: Tolerance of known risk, which was not a significant predictor of subsequent opioid use, was the only different marker among the recovering and healthy control groups (see also Gueguen et al. 2021 for discussion of this result).

Hierarchical modeling (including frequentist mixed-effects models and hierarchical Bayesian models; see Supplemental Figure 1) offers a statistically principled approach to modeling between- and within-subject effects, and it enjoys widespread use in computational neuroscience and psychiatry. Multilevel vector auto-regressive (VAR) models enable the estimation of some specific temporal effects, permitting examination, for example, of how various emotions predict themselves and other emotions over time (Lydon-Staley et al. 2021). This allowed researchers to corroborate clinical insights such as the idea that, among neurotic individuals, worry strengthens the duration and transition between negative emotions (Bringmann et al. 2013). To date, such models have largely relied on self-reports, but an exciting future avenue is to use multimodal methods, including neurocomputational markers derived from computational psychiatry methods, to estimate the elements in such networks with higher precision. This is especially important to overcome the problems inherent to the investigation of suitcase constructs, such as worry, that may encompass so many primitive processes that their relationships to other items are confounded (Eronen 2020a).

Despite their advantages, hierarchical methods alone are of course unable to resolve the limitations inherent in attempting to extrapolate from between-subjects data to within-subject patterns. Moreover, from the perspective of informing person-specific interventions, hierarchical methods can distort individual patterns that may be important (due to their imposition of distributions that can alter patterns from the raw data, especially outlying points). In particular, hierarchical methods may sometimes mask patterns operative within individuals over time that could be important—to psychotherapy conceptualizations, for example. Drawing on a rich tradition of single-case designs (Barlow & Hersen 1973), psychotherapy-minded research is seeing an efflorescence of methods aimed at capturing and capitalizing on within-subject patterns (Wright & Woods 2020). Potentially offering the best of both worlds, methods such as the GIMME algorithm seek to capture time-series

patterns reliably present within a group and at the same time extract idiographic patterns (Beltz & Gates 2017).

An exciting avenue for future research is to connect these person-specific approaches that offer rigorous methods for functional conceptualizations of mental health with computational psychiatry accounts. What the latter have to offer are new clinical principles for the next generation of psychotherapies built upon basic (e.g., computer and decision) sciences (Moutoussis et al. 2018, Niv et al. 2021). It is worth noting that there are natural complementarities among the functional-analytic tradition in behavior therapy, which seeks to understand why behavior occurs in a context with an eye toward modifying it (Burger et al. 2020, Hofmann & Hayes 2019); the network approach, which views mental health problems as causally related elements interacting over time (McNally 2021); and the bounded (computational) rationality perspective in the decision and computer sciences, which seeks to model decision making under limited resources, and which can explain how what might appear to be dysfunctional responding is actually rational in light of context and constraints (Gershman et al. 2015, Russek et al. 2020, Simon 1990).

3.3.3. Formalizing mental health systems.—A landmark development toward modeling time and context is the recent development by Robinaugh and colleagues (2019) of a large-scale mental health system (in this case, panic disorder). This system implements the network approach vision of interacting mental health elements within a detailed computational model that can simulate mental health dynamics. Notably, this system was recently extended to model the effect of functional-analytic interventions for panic disorder (Burger et al. 2020), thereby demonstrating a parallel functionality to the ability of biologically detailed computational neuroscience models to simulate the dynamics of specific interventions, such as an increase in tonic dopamine. Robinaugh et al.'s (2019) model has not yet incorporated rich biological detail, nor has it been paired with algorithmic approaches to concisely summarize key model behaviors that can be applied to describe individual differences between people; these are exciting avenues for future research. Integrating this type of approach with powerful techniques from the mainstream of computational psychiatry may eventually enable time and context to be rigorously incorporated into computational psychiatry, providing insights and targeted intervention opportunities even for low-essentiality problems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Thank you to Zachary Cohen, Romy Frömer, Ivan Grahek, Louis Gularte, Alexander Kline, and Amrita Lamba for feedback on parts of this manuscript, and to Jim Gold, Yael Niv, Jeff Poland, Dan Scott, and Isabel Berwian and the rest of the Computational Psychotherapy Group for exchanges that contributed to its ideas. P.F.H. was supported by National Institute of Mental Health (NIMH) grant F32 MH123055. M.J.F. was supported by NIMH grants P50 MH119467 and R01 MH084840-08A.

DISCLOSURE STATEMENT

M.J.F. receives consultant fees from Hoffman-La Roche pharmaceuticals on topics related to computational psychiatry and is the recipient of several grants from the National Institutes of Health for research on computational psychiatry. The authors are not aware of any other affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Agrawal M, Mattar MG, Cohen JD, Daw ND. 2020. The temporal dynamics of opportunity costs: A normative account of cognitive fatigue and boredom. bioRxiv 2877276. 10.1101/2020.09.08.287276

Am. Psychiatr. Assoc. 2013. Diagnostic and Statistical Manual of Mental Disorders. Arlington, VA: Am. Psychiatr. Publ. 5th ed. [DSM-5]

Andrews PW, Thomson JA Jr. 2009. The bright side of being blue: depression as an adaptation for analyzing complex problems. Psychol. Rev. 116(3):620–54 [PubMed: 19618990] Makes the case that depressed mood and rumination facilitate recovery and elicit needed social support.

Armour C, Műllerová J, Elhai JD. 2015. A systematic literature review of PTSD's latent structure in the Diagnostic and Statistical Manual of Mental Disorders: DSM-IV to DSM-5. Clin. Psychol. Rev 44:60–74 [PubMed: 26761151]

Baker CL, Jara-Ettinger J, Saxe R, Tenenbaum JB. 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nat. Hum. Behav. 1:0064

Barlow DH, Hersen M. 1973. Single-case experimental designs: uses in applied clinical research. Arch. Gen. Psychiatry 29(3):319–25 [PubMed: 4724141]

Beeler JA, Frank MJ, McDaid J, Alexander E, Turkson S, et al. 2012. A role for dopamine-mediated learning in the pathophysiology and treatment of Parkinson's disease. Cell Rep. 2(6):1747–61 [PubMed: 23246005]

Beevers CG, Mullarkey MC, Dainer-Best J, Stewart RA, Labrada J, et al. 2019. Association between negative cognitive bias and depression: a symptom-level approach. J. Abnorm. Psychol. 128(3):212–27 [PubMed: 30652884]

Beltz AM, Gates KM. 2017. Network mapping with GIMME. Multivar. Behav. Res. 52(6):789–804

Borsboom D 2008. Psychometric perspectives on diagnostic systems. J. Clin. Psychol. 64(9):1089–108 [PubMed: 18683856]

Boyd R 1991. Realism, anti-foundationalism and the enthusiasm for natural kinds. Philos. Stud. 61(1–2):127–48

Brick C, Hood B, Ekroll V, de-Wit L. 2021. Illusory essences: a bias holding back theorizing in psychological science. Perspect. Psychol. Sci In press

Bringmann LF, Eronen MI. 2018. Don't blame the model: reconsidering the network approach to psychopathology. Psychol. Rev. 125(4):606–15 [PubMed: 29952625]

Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, et al. 2013. A network approach to psychopathology: new insights into clinical longitudinal data. PLOS ONE 8(4):e60188 [PubMed: 23593171]

Brown VM, Zhu L, Wang JM, Frueh BC, King-Casas B, Chiu PH. 2018. Associability-modulated loss learning is increased in posttraumatic stress disorder. eLife 7:e30150 [PubMed: 29313489]

Browning M, Carter CS, Chatham C, Den Ouden H, Gillan CM, et al. 2020. Realizing the clinical potential of computational psychiatry: report from the Banbury Center Meeting, February 2019. Biol. Psychiatry 88(2):e5–10 [PubMed: 32113656]

Burger J, van der Veen DC, Robinaugh DJ, Quax R, Riese H, et al. 2020. Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: a computational model of functional analysis. BMC Med. 18(1):99 [PubMed: 32264914]

Cai N, Choi KW, Fried EI. 2020. Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. Hum. Mol. Genet. 29(R1):R10–18 [PubMed: 32568380]

Cavanagh JF, Wiecki TV, Cohen MX, Figueroa CM, Samanta J, et al. 2011. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nat. Neurosci. 14(11):1462–67 [PubMed: 21946325]

Cicchetti D, Rogosch FA. 1996. Equifinality and multifinality in developmental psychopathology. Dev. Psychopathol. 8(4):597–600

Cohen JD, Servan-Schreiber D. 1992. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol. Rev. 99(1):45–77 [PubMed: 1546118]

Cohen RT, Kahana MJ. 2020. A memory-based theory of emotional disorders. bioRxiv 817486. 10.1101/817486

Collins AGE, Albrecht MA, Waltz JA, Gold JM, Frank MJ. 2017. Interactions among working memory, reinforcement learning, and effort in value-based choice: a new paradigm and selective deficits in schizophrenia. Biol. Psychiatry 82(6):431–39 [PubMed: 28651789]

Collins AGE, Frank MJ. 2014. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol. Rev. 121(3):337–66 [PubMed: 25090423]

Cools R, Barker RA, Sahakian BJ, Robbins TW. 2001. Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. Cereb. Cortex 11(12):1136–43 [PubMed: 11709484]

Dasgupta I, Gershman SJ. 2021. Memory as a computational resource. Trends Cogn. Sci. 25(3):240–51 [PubMed: 33454217]

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron 69(6):1204–15 [PubMed: 21435563]

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. 8(12):1704–11 [PubMed: 16286932]

Dayan P, Huys QJM. 2008. Serotonin, inhibition, and negative mood. PLOS Comput. Biol. 4(2):e4 [PubMed: 18248087]

Decker JH, Otto AR, Daw ND, Hartley CA. 2016. From creatures of habit to goal-directed learners: tracking the developmental emergence of model-based reinforcement learning. Psychol. Sci. 27(6):848–58 [PubMed: 27084852]

Dowd EC, Frank MJ, Collins A, Gold JM, Barch DM. 2016. Probabilistic reinforcement learning in patients with schizophrenia: relationships to anhedonia and avolition. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 1(5):460–73 [PubMed: 27833939]

Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, et al. 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 23(1):28–38 [PubMed: 27918562]

Durstewitz D, Huys QJM, Koppe G. 2020. Psychiatric illnesses as disorders of network dynamics. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 6(9):865–76 [PubMed: 32249208]

Durstewitz D, Seamans JK. 2008. The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-O-methyltransferase genotypes and schizophrenia. Biol. Psychiatry 64(9):739–49 [PubMed: 18620336]

Egerton A, Valmaggia LR, Howes OD, Day F, Chaddock CA, et al. 2016. Adversity in childhood linked to elevated striatal dopamine function in adulthood. Schizophr. Res. 176(2–3):171–76 [PubMed: 27344984]

Eldar E, Niv Y. 2015. Interaction between emotional state and learning underlies mood instability. Nat. Commun. 6:6149 [PubMed: 25608088] Shows a coupling between affect and appraisal and links individual differences therein to hypomanic symptoms.

Eldar E, Roth C, Dayan P, Dolan RJ. 2018. Decodability of reward learning signals predicts mood fluctuations. Curr. Biol. 28(9):1433–39.e7 [PubMed: 29706512]

Eronen MI. 2019. The levels problem in psychopathology. Psychol. Med. 51(6):927–33 [PubMed: 31549600]

Eronen MI. 2020a. Causal discovery and the problem of psychological interventions. New Ideas Psychol. 59:100785

Eronen MI. 2020b. Interventionism for the intentional stance: true believers and their brains. Topoi 39(1):45–55

Frank MJ. 2005. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J. Cogn. Neurosci. 17(1):51–72 [PubMed: 15701239]

Frank MJ, Claus ED. 2006. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychol. Rev. 113(2):300–26 [PubMed: 16637763]

Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, et al. 2015. fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. J. Neurosci. 35(2):485–94 [PubMed: 25589744]

Frank MJ, Samanta J, Moustafa AA, Sherman SJ. 2007. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science 318(5854):1309–12 [PubMed: 17962524]

Frank MJ, Seeberger LC, O'Reilly RC. 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306(5703):1940–43 [PubMed: 15528409] Shows a mechanistic learning difference in the high-essentiality problem Parkinson's disease.

Frey A-L, Frank MJ, McCabe C. 2021. Social reinforcement learning as a predictor of real-life experiences in individuals with high and low depressive symptomatology. Psychol. Med. 51(3):408–15 [PubMed: 31831095]

Frey A-L, McCabe C 2020a. Effects of serotonin and dopamine depletion on neural prediction computations during social learning. Neuropsychopharmacology 45(9):1431–37 [PubMed: 32330925]

Frey A-L, McCabe C 2020b. Impaired social learning predicts reduced real-life motivation in individuals with depression: a computational fMRI study. J. Affect. Disord. 263:698–706 [PubMed: 31784119]

Fried EI, Cramer AOJ. 2017. Moving forward: challenges and directions for psychopathological network theory and methodology. Perspect. Psychol. Sci. 12(6):999–1020 [PubMed: 28873325]

Fried EI, Nesse RM. 2015. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Med. 13:72 [PubMed: 25879936]

Geana A, Barch DM, Gold JM, Carter CS, MacDonald AW III, et al. 2021. Using computational modelling to capture schizophrenia-specific reinforcement learning differences and their implications on patient classification. Biol. Psychiatry. In press. 10.1016/j.bpsc.2021.03.017

Gelman SA. 2004. Psychological essentialism in children. Trends Cogn. Sci. 8(9):404–9 [PubMed: 15350241]

Gershman SJ, Horvitz EJ, Tenenbaum JB. 2015. Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. Science 349(6245):273–78 [PubMed: 26185246]

Gillan CM, Fineberg NA, Robbins TW. 2017. A trans-diagnostic perspective on obsessive-compulsive disorder. Psychol. Med. 47(9):1528–48 [PubMed: 28343453]

Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND. 2016. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife 5:e11305 [PubMed: 26928075]

Gillan CM, Otto AR, Phelps EA, Daw ND. 2015. Model-based learning protects against forming habits. Cogn. Affect. Behav. Neurosci. 15(3):523–36 [PubMed: 25801925]

Gillan CM, Rutledge RB. 2021. Smartphones and the neuroscience of mental health. Annu. Rev. Neurosci. 44:129–51 [PubMed: 33556250] Two computational psychiatrists make the case for the importance of smartphones in advancing mental-health research.

Gillan CM, Seow TXF. 2020. Carving out new transdiagnostic dimensions for research in mental health. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 5(10):932–34 [PubMed: 32532686]

Gold JM, Waltz JA, Frank MJ. 2015. Effort cost computation in schizophrenia: a commentary on the recent literature. Biol. Psychiatry 78(11):747–53 [PubMed: 26049208]

Gold JM, Waltz JA, Matveeva TM, Kasanova Z, Strauss GP, et al. 2012. Negative symptoms and the failure to represent the expected reward value of actions: behavioral and computational modeling evidence. Arch. Gen. Psychiatry 69(2):129–38 [PubMed: 22310503]

Goldstein BL, Klein DN. 2014. A review of selected candidate endophenotypes for depression. Clin. Psychol. Rev. 34(5):417–27 [PubMed: 25006008]

Grahek I, Shenhav A, Musslick S, Krebs RM, Koster EH. 2019. Motivation and cognitive control in depression. Neurosci. Biobehav. Rev. 102:371–81 [PubMed: 31047891] Analyzes the possible sources of apparent cognitive control impairments in depression.

Gu X, FitzGerald THB, Friston KJ. 2019. Modeling subjective belief states in computational psychiatry: interoceptive inference as a candidate framework. Psychopharm. 236(8):2405–12

Gueguen MCM, Schweitzer EM, Konova AB. 2021. Computational theory-driven studies of reinforcement learning and decision-making in addiction: What have we learned? Curr. Opin. Behav. Sci. 38:40–48 [PubMed: 34423103]

Hammen C 2005. Stress and depression. Annu. Rev. Clin. Psychol. 1:293–319 [PubMed: 17716090]

Haslbeck JMB, Borsboom D, Waldorp LJ. 2021. Moderated network models. Multivar. Behav. Res. 56(2):256–87

Haslbeck JMB, Fried EI. 2017. How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. Psychol. Med. 47(16):2767–76 [PubMed: 28625186]

Haslbeck JMB, Ryan O, Robinaugh D, Waldorp L, Borsboom D. 2019. Modeling psychopathology: from data models to formal theories. PsyArXiv, Dec. 10. 10.31234/osf.io/jgm7f

Henry TR, Robinaugh D, Fried EI. 2020. On the control of psychological networks. PsyArXiv, Apr. 1. 10.31234/osf.io/7vpz2

Herz DM, Zavala BA, Bogacz R, Brown P. 2016. Neural correlates of decision thresholds in the human subthalamic nucleus. Curr. Biol. 26(7):916–20 [PubMed: 26996501]

Hitchcock P, Forman E, Rothstein NJ, Zhang F, Kounios J, et al. 2021. Rumination derails reinforcement learning with possible implications for ineffective behavior. Clin. Psychol. Sci In press Argues that rumination may impede effective behavior in specific situations by impairing trial-and-error learning.

Hofmann SG, Hayes SC. 2019. The future of intervention science: process-based therapy. Clin. Psychol. Sci. 7(1):37–50 [PubMed: 30713811]

Hunter LE, Meer EA, Gillan CM, Hsu M, Daw ND. 2019. Excessive deliberation in social anxiety. bioRxiv 522433. 10.1101/522433

Huys QJM, Browning M, Paulus MP, Frank MJ. 2021. Advances in the computational understanding of mental illness. Neuropsychopharmacology 46(1):3–19 [PubMed: 32620005]

Huys QJM, Daw ND, Dayan P. 2015. Depression: a decision-theoretic analysis. Annu. Rev. Neurosci. 38:1–23 [PubMed: 25705929]

Huys QJM, Maia TV, Frank MJ. 2016. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat. Neurosci. 19(3):404–13 [PubMed: 26906507]

Huys QJM, Pizzagalli DA, Bogdan R, Dayan P. 2013. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. Biol. Mood Anxiety Disord. 3:12 [PubMed: 23782813]

Kaiser RH, Andrews-Hanna JR, Wager TD, Pizzagalli DA. 2015. Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. JAMA Psychiatry 72(6):603–11 [PubMed: 25785575]

Kapur S 2003. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am. J. Psychiatry 160(1):13–23 [PubMed: 12505794]

Kendler KS. 2005. Toward a philosophical structure for psychiatry. Am. J. Psychiatry 162(3):433–40 [PubMed: 15741457]

Kendler KS, Gardner CO, Gatz M, Pedersen NL. 2007. The sources of co-morbidity between major depression and generalized anxiety disorder in a Swedish national twin sample. Psychol. Med. 37(3):453–62 [PubMed: 17121688]

Kendler KS, Hettema JM, Butera F, Gardner CO, Prescott CA. 2003. Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. Arch. Gen. Psychiatry 60(8):789–96 [PubMed: 12912762] Shows some specificity in the life events precipitating major depression and generalized anxiety (humiliation and endangerment, respectively).

Kendler KS, Thornton LM, Gardner CO. 2000. Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the “kindling” hypothesis. Am. J. Psychiatry 157(8):1243–51 [PubMed: 10910786]

Kendler KS, Zachar P, Craver C. 2011. What kinds of things are psychiatric disorders? Psychol. Med. 41(6):1143–50 [PubMed: 20860872]

Keramati M, Smittenaar P. 2016. Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. PNAS 113(45):12868–73 [PubMed: 27791110]

Keren H, O'Callaghan G, Vidal-Ribas P, Buzzell GA, Brotman MA, et al. 2018. Reward processing in depression: a conceptual and meta-analytic review across fMRI and EEG studies. Am. J. Psychiatry 175(11):1111–20 [PubMed: 29921146]

Kievit RA, Frankenhuis WE, Waldorp LJ, Borsboom D. 2013. Simpson's paradox in psychological science: a practical guide. Front. Psychol. 4:513 [PubMed: 23964259]

Konova AB, Lopez-Guzman S, Urmanche A, Ross S, Louie K, et al. 2020. Computational markers of risky decision-making for identification of temporal windows of vulnerability to opioid use in a real-world clinical setting. JAMA Psychiatry 77(4):368–77 [PubMed: 31812982] Remarkable study showing that variation in the propensity to take ambiguous risk predicts prospective drug use.

Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, et al. 2017. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. J. Abnorm. Psychol. 126(4):454–77 [PubMed: 28333488]

Kozak MJ, Cuthbert BN. 2016. The NIMH research domain criteria initiative: background, issues, and pragmatics. Psychophysiology 53(3):286–97 [PubMed: 26877115]

Lamba A, Frank MJ, FeldmanHall O. 2020. Anxiety impedes adaptive social learning under uncertainty. Psychol. Sci. 31(5):592–603 [PubMed: 32343637]

Leising D, Burger J, Zimmermann J, Bäckström M, Oltmanns JR, Connelly BS. 2020. Why do items correlate with one another? A conceptual analysis with relevance for general factors and network models. PsyArXiv, Aug. 8. 10.31234/osf.io/7c895

Lieder F, Goodman ND, Huys QJM. 2013. Learned helplessness and generalization. In Proceedings of the Annual Meeting of the Cognitive Science Society, Vol. 35, pp. 900–5. Austin, TX: Cogn. Sci. Soc.

Litz BT, Stein N, Delaney E, Lebowitz L, Nash WP, et al. 2009. Moral injury and moral repair in war veterans: a preliminary model and intervention strategy. Clin. Psychol. Rev. 29(8):695–706 [PubMed: 19683376]

Lydon-Staley DM, Cornblath EJ, Blevins AS, Bassett DS. 2021. Modeling brain, symptom, and behavior in the winds of change. Neuropsychopharmacology 46(1):20–32 [PubMed: 32859996]

Maia TV, Frank MJ. 2011. From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. 14(2):154–62 [PubMed: 21270784]

Maia TV, Frank MJ. 2017. An integrative perspective on the role of dopamine in schizophrenia. Biol. Psychiatry 81(1):52–66 [PubMed: 27452791]

Mason L, Eldar E, Rutledge RB. 2017. Mood instability and reward dysregulation—a neurocomputational model of bipolar disorder. JAMA Psychiatry 74(12):1275–76 [PubMed: 29049438]

Matar E, Shine JM, Gilat M, Ehgoetz Martens KA, Ward PB, et al. 2019. Identifying the neural correlates of doorway freezing in Parkinson's disease. Hum. Brain Mapp. 40(7):2055–64 [PubMed: 30637883]

McCabe C 2018. Linking anhedonia symptoms with behavioural and neural reward responses in adolescent depression. Curr. Opin. Behav. Sci. 22:143–51

McClure SM, Laibson DI, Loewenstein G, Cohen JD. 2004. Separate neural systems value immediate and delayed monetary rewards. Science 306(5695):503–7 [PubMed: 15486304]

McCutcheon RA, Marques TR, Howes OD. 2020. Schizophrenia—an overview. JAMA Psychiatry 77(2):201–10 [PubMed: 31664453]

McNally RJ. 2001. On Wakefield's harmful dysfunction analysis of mental disorder. Behav. Res. Ther. 39(3):309–14 [PubMed: 11227812]

McNally RJ. 2011. What Is Mental Illness? Cambridge, MA: Harvard Univ. Press

McNally RJ. 2021. Network analysis of psychopathology: controversies and challenges. Annu. Rev. Clin. Psychol. 17:31–53 [PubMed: 33228401]

Michely J, Eldar E, Martin IM, Dolan RJ. 2020. A mechanistic account of serotonin's impact on mood. Nat. Commun. 11(1):2335 [PubMed: 32393738]

Minsky M 2007. The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind. New York: Simon & Schuster

Mobbs D, Headley DB, Ding W, Dayan P. 2020. Space, time, and fear: survival computations along defensive circuits. Trends Cogn. Sci. 24(3):228–41 [PubMed: 32029360]

Moutoussis M, Eldar E, Dolan RJ. 2017. Building a new field of computational psychiatry. Biol. Psychiatry 82(6):388–90 [PubMed: 27876357]

Moutoussis M, Shahar N, Hauser TU, Dolan RJ. 2018. Computation in psychotherapy, or how computational psychiatry can aid learning-based psychological therapies. Comput. Psychiatry 2:50–73

Nelson B, McGorry PD, Wichers M, Wigman JTW, Hartmann JA. 2017. Moving from static to dynamic models of the onset of mental disorder: a review. JAMA Psychiatry 74(5):528–34 [PubMed: 28355471]

Niv Y, Hitchcock P, Berwian IM, Schoen G. 2021. Toward precision cognitive behavioral therapy via reinforcement learning theory. In Precision Psychiatry, ed. Williams LM, Hack LM. Washington, DC: Am. Psychiatr. Assoc. In press

Nolen-Hoeksema S, Watkins ER. 2011. A heuristic for developing transdiagnostic models of psychopathology: explaining multifinality and divergent trajectories. Perspect. Psychol. Sci. 6(6):589–609 [PubMed: 26168379] How to think about multifinality by distinguishing the factors that contribute to mental health problems.

Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. 2008. Rethinking rumination. Perspect. Psychol. Sci. 3(5):400–24 [PubMed: 26158958]

Norman SB, Wilkins KC, Myers US, Allard CB. 2014. Trauma informed guilt reduction therapy with combat veterans. Cogn. Behav. Pract. 21(1):78–88 [PubMed: 25404850]

Patzelt EH, Kool W, Millner AJ, Gershman SJ. 2019. Incentives boost model-based control across a range of severity on several psychiatric constructs. Biol. Psychiatry 85(5):425–33 [PubMed: 30077331]

Perivoliotis D, Morrison AP, Grant PM, French P, Beck AT. 2009. Negative performance beliefs and negative symptoms in individuals at ultra-high risk of psychosis: a preliminary study. Psychopathology 42(6):375–79 [PubMed: 19752591]

Podsakoff PM, MacKenzie SB, Podsakoff NP. 2012. Sources of method bias in social science research and recommendations on how to control it. Annu. Rev. Psychol 63:539–69 [PubMed: 21838546]

Poldrack RA, Yarkoni T. 2016. From brain maps to cognitive ontologies: informatics and the search formental structure. Annu. Rev. Psychol. 67:587–612 [PubMed: 26393866]

Pulcu E, Browning M. 2017. Using computational psychiatry to rule out the hidden causes of depression. JAMA Psychiatry 74(8):777–78 [PubMed: 28678978]

Radulescu A, Niv Y. 2019. State representation in mental illness. Curr. Opin. Neurobiol. 55:160–66 [PubMed: 31051434]

Redish AD, Gordon JA. 2016. Computational Psychiatry: New Perspectives on Mental Illness. Cambridge, MA: MIT Press

Robinaugh D, Haslbeck J, Waldorp L, Kossakowski J, Fried EI, et al. 2019. Advancing the network theory of mental disorders: a computational model of panic disorder. PsyArXiv, May 29. 10.31234/osf.io/km37w Empirically informed, quantitative representation of panic disorder as a network of interacting elements.

Rodebaugh TL, Scullin RB, Langer JK, Dixon DJ, Huppert JD, et al. 2016. Unreliability as a threat to understanding psychopathology: the cautionary tale of attentional bias. J. Abnorm. Psychol. 125(6):840–51 [PubMed: 27322741]

Rouault M, Seow T, Gillan CM, Fleming SM. 2018. Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biol. Psychiatry 84(6):443–51 [PubMed: 29458997]

Russek EM, Moran R, McNamee D, Reiter A, Liu Y, et al. 2020. Opportunities for emotion and mental health research in the resource-rationality framework. Behav. Brain Sci. 43:e21 [PubMed: 32159474]

Rutledge RB, Chekroud AM, Huys QJ. 2019. Machine learning and big data in psychiatry: toward clinical applications. Curr. Opin. Neurobiol. 55:152–59 [PubMed: 30999271]

Salmi J, Ritakallio L, Fellman D, Ellfolk U, Rinne JO, Laine M. 2020. Disentangling the role of working memory in Parkinson's disease. Front. Aging Neurosci. 12:572037 [PubMed: 33088273]

Scholl J, Klein-Flügge M. 2018. Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. Behav. Brain Res. 355:56–75 [PubMed: 28966147]

Shenhav A, Botvinick MM, Cohen JD. 2013. The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79(2):217–40 [PubMed: 23889930]

Simon HA. 1990. Invariants of human behavior. Annu. Rev. Psychol. 41:1–19 [PubMed: 18331187]

Snyder HR. 2013. Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. Psychol. Bull. 139(1):81–132 [PubMed: 22642228]

Snyder HR, Hankin BL. 2016. Spiraling out of control: stress generation and subsequent rumination mediate the link between poorer cognitive control and internalizing psychopathology. Clin. Psychol. Sci. 4(6):1047–64 [PubMed: 27840778]

Spitzer RL, First MB, Wakefield JC. 2007. Saving PTSD from itself in DSM-V. J. Anxiety Disord. 21(2):233–41 [PubMed: 17141468]

Treadway MT, Zald DH. 2011. Reconsidering anhedonia in depression: lessons from translational neuroscience. Neurosci. Biobehav. Rev. 35(3):537–55 [PubMed: 20603146]

Valton V, Romaniuk L, Douglas Steele J, Lawrie S, Seriès P. 2017. Comprehensive review: computational modelling of schizophrenia. Neurosci. Biobehav. Rev. 83:631–46 [PubMed: 28867653]

van de Leemput IA, Wichers M, Cramer AOJ, Borsboom D, Tuerlinckx F, et al. 2014. Critical slowing down as early warning for the onset and termination of depression. PNAS 111(1):87–92 [PubMed: 24324144]

Van Vugt MK, Hitchcock P, Shahar B. 2012. The effects of mindfulness-based cognitive therapy on affective memory recall dynamics in depression: a mechanistic model of rumination. Front. Hum. Neurosci. 6:257 [PubMed: 23049507]

Wachbroit R 1994. Normality as a biological concept. Philos. Sci. 61(4):579–91

Wakefield JC. 1992a. Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder. Psychol. Rev. 99(2):232–47 [PubMed: 1594724]

Wakefield JC. 1992b. The concept of mental disorder: on the boundary between biological facts and social values. Am. Psychol. 47(3):373–88 [PubMed: 1562108]

Wang X-J, Krystal JH. 2014. Computational psychiatry. Neuron 84(3):638–54 [PubMed: 25442941]

Watkins ER. 2008. Constructive and unconstructive repetitive thought. Psychol. Bull. 134(2):163–206 [PubMed: 18298268]

Watts AL, Boness CL, Loeffelman JE, Steinley D. 2020. Does crude measurement contribute to observed unidimensionality of psychological constructs? An example with DSM-5 alcohol use disorder. PsyArXiv, June 28. 10.31234/osf.io/paxd4

Webb C, Trivedi M, Bruder G, Pizzagalli DA. 2016. Neural correlates of three promising endophenotypes of depression: evidence from the EMBARC study. Neuropsychopharmacology 41:454–63 [PubMed: 26068725]

Weiner WJ. 2008. There is no Parkinson disease. Arch. Neurol. 65(6):705–8 [PubMed: 18541790]

Whitmer AJ, Frank MJ, Gotlib IH. 2012. Sensitivity to reward and punishment in major depressive disorder: effects of rumination and of single versus multiple experiences. Cogn. Emot. 26(8):1475–85 [PubMed: 22716241]

Wiecki TV, Antoniades CA, Stevenson A, Kennard C, Borowsky B, et al. 2016. A computational cognitive biomarker for early-stage Huntington's disease. PLOS ONE 11(2):e0148409 [PubMed: 26872129]

Wiecki TV, Frank MJ. 2010. Neurocomputational models of motor and cognitive deficits in Parkinson's disease. Prog. Brain Res. 183:275–97 [PubMed: 20696325]

Wiecki TV, Poland J, Frank MJ. 2015. Model-based cognitive neuroscience approaches to computational psychiatry: clustering and classification. Clin. Psychol. Sci. 3(3):378–99

Williams LM. 2016. Precision psychiatry: a neural circuit taxonomy for depression and anxiety. Lancet Psychiatry 3(5):472–80 [PubMed: 27150382]

Wright AGC, Woods WC. 2020. Personalized models of psychopathology. Annu. Rev. Clin. Psychol. 16:49–74 [PubMed: 32070120]

Zorowitz S, Momennejad I, Daw ND. 2020. Anxiety, avoidance, and sequential evaluation. Comput. Psychiatry 4:1–17

WHAT DOES DYSFUNCTION MEAN IN A MENTAL HEALTH CONTEXT?

How to define dysfunction within a mental health context has been the subject of intense debate (e.g., McNally 2001, Wakefield 1992b). We favor a definition proposed by McNally (2011) that casts dysfunction as a disrupted process operating within a larger causal system. For instance, the heart malfunctions within the context of the circulatory system if it fails to pump blood; the amygdala malfunctions within the threat-detection system if it fails to respond to proximal threat or responds excessively to neutral stimuli (McNally 2011). This definition rests on a notion of normal function versus aberrant functioning. Wachbroit (1994) argued that a concept of normality is indispensable within biology. Normal function, according to this account, is not the same as statistically normal (i.e., average or prototypical function). For instance, a radioactive accident could render the hearts of everyone on earth dysfunctional; in this case, statistical deviation would not help to reveal dysfunction (Wakefield 1992a). Rather, normal function by this account refers to an idealized operation of the function against which deviations can be gauged (Wachbroit 1994).

SUMMARY POINTS

1. We predict that progress in the next generation of computational psychiatry will come from modeling time and context in order to tame the complexity of mental health disorders of lower essentiality.
2. Three heuristics can help to estimate essentiality: Is there a single, core neurobiological mechanism at the problem's root? Does the problem follow a straightforward natural course? Is intentional mental content (such as beliefs) distinct from the problem itself?
3. If the answer to all of these questions is yes, the problem has high essentiality. By contrast, lower-essentiality problems comprise multiple interrelated elements (not all necessarily dysfunctional) and vary greatly over time. Intentional content is important in these problems.
4. Clinical principles concerning beliefs, values, personal significance, humiliation, and other types of intentional content could be grounded in computational theories. In addition, the type of intentional content endemic to a problem can help us contextualize observed differences. For instance, do individuals with this problem invariably show differences in trial-and-error learning, or are the differences limited to specific social contexts? What does this tell us about the problem itself?
5. Mental health problems may spuriously appear to have low essentiality because of imprecise phenotyping. Computational psychiatry has much to contribute to the important project of refining phenotypes. Yet, standard approaches to deriving more precise phenotypes at a single point in time may be insufficient for lower-essentiality problems because of their temporal and contextual dependence (i.e., their meaningful heterogeneity). Modeling variation over time and in context is critical. Even when this is done, the complexity of these problems implies that it might take more time to make progress on them compared to simpler problems.
6. Algorithmic modeling has a special role in bridging levels and dimensions of analysis in computational psychiatry, although there are many technical and inferential challenges. Caution is required. Recent innovations may dramatically advance the scope and power of these models (see Supplemental Figure 1).
7. Computational psychiatry theories are beginning to make risky predictions about dynamics in the real world. Modeling and measurement techniques from adjacent areas—including network and complex-systems approaches and digital phenotyping—will be important to the next generation of computational psychiatry, especially for capturing and modeling the real-world dynamics of lower-essentiality problems and thereby enabling iterative refinement of increasingly sharp predictions.

8. The importance of context in lower-essentiality problems resonates with the perspectives of three traditions that developed largely independently: the functional-analytic tradition in behavior therapy, the bounded (computational) rationality tradition in the decision sciences, and the network approach to mental health. These shared perspectives raise the prospect of uniting computational and psychotherapy principles.

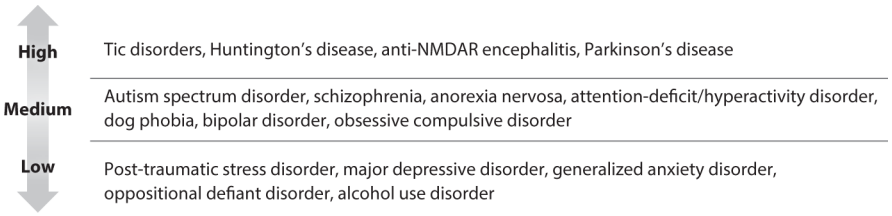

Figure 1.

Estimates of whether several well-known mental health problems have high, medium, or low essentiality. Abbreviation: NMDAR, *N*-methyl-D-aspartate receptor.

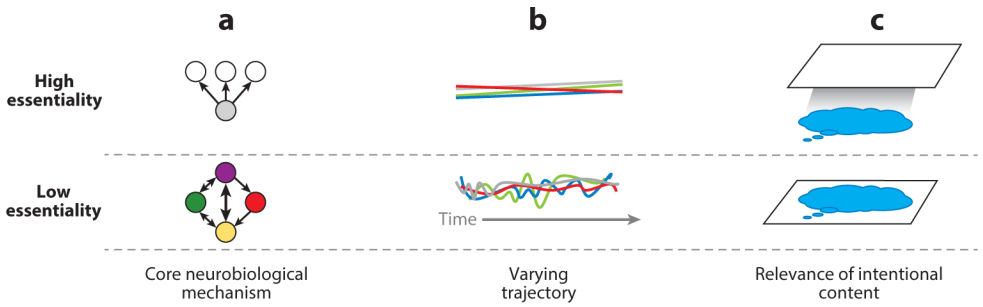

Figure 2.

Visualization of three heuristics for estimating essentiality. (*a*) High-essentiality problems comprise a set of signs and symptoms that arise from a core neurobiological mechanism, whereas low-essentiality problems are best thought of as a set of elements in varied relational patterns with one another (denoted by arrows of different widths and directions). These elements constitute low-essentiality problems. (*b*) High-essentiality problems follow a relatively linear naturalistic (i.e., absent intervention) course, whereas lower-essentiality problems follow variable trajectories. (*c*) Intentional mental content (e.g., negative schemata; *blue bubble*) is central to low-essentiality problems (e.g., major depression; *white plane*). Such content may be present in high-essentiality problems (e.g., Parkinson's disease), but it is not key to understanding such problems.

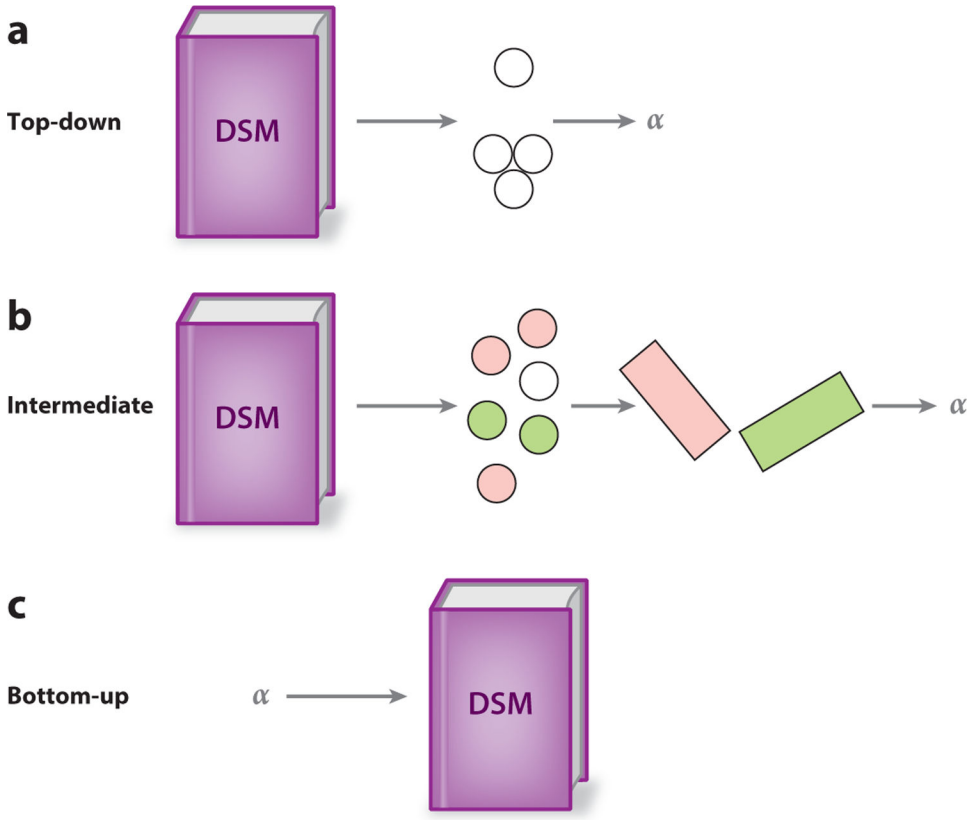

Figure 3.

Approaches to improving phenotypic precision. (*a*) Top-down approaches begin with symptoms or symptom clusters (*white circles*) and relate these to processes inferred via computational psychiatry methods (such as differences in learning rate, represented by an α parameter). (*b*) Intermediate approaches also typically use symptoms encoded in the diagnostic systems, but they use dimension-reduction techniques to derive summaries of which symptoms share variance (represented by the orthogonal planes) and then relate these summaries to inferred processes. (*c*) The bottom-up approach begins with a process well characterized by computational psychiatry methods, such as a mechanism represented by a parameter that can be distinguished from others and that often has a clear function and link to neurobiology. It then attempts to relate differences in this process to clinical phenomena, such as symptoms or diagnostic categories. Abbreviation: DSM, *Diagnostic and Statistical Manual of Mental Disorders*.

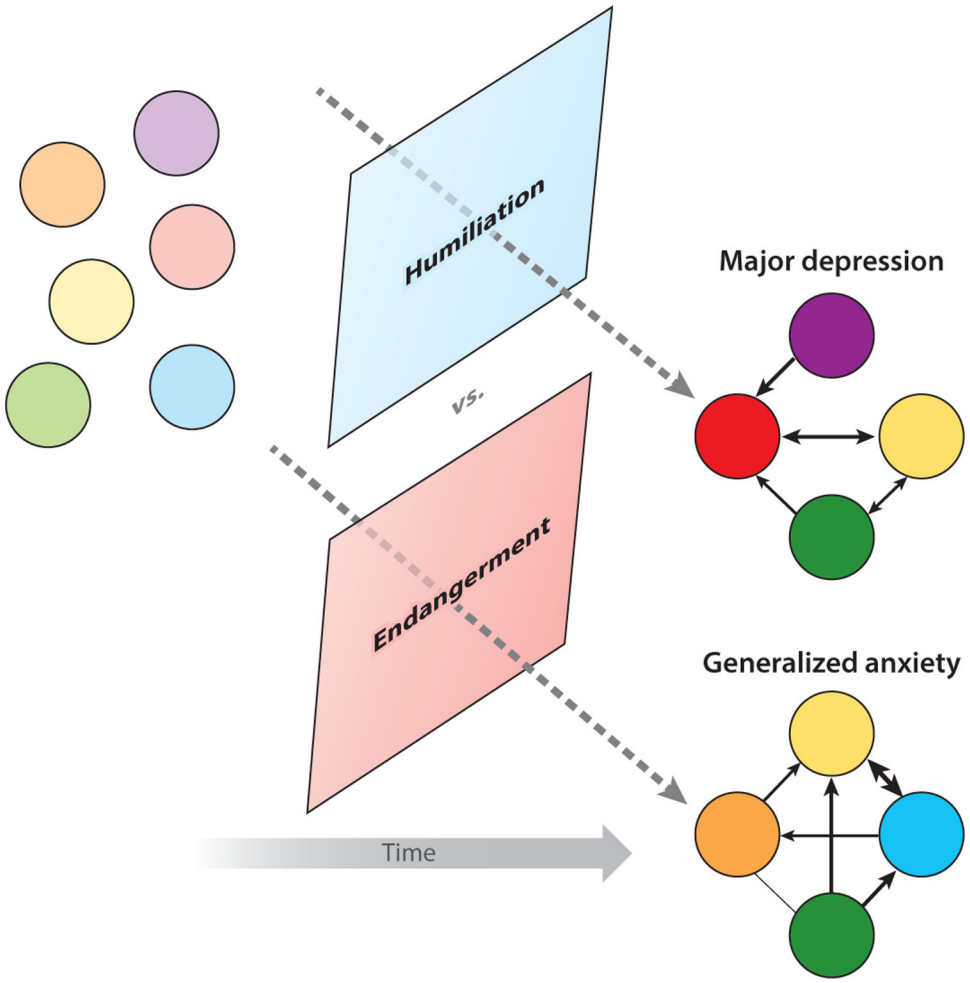

Figure 4.

An example of meaningful heterogeneity. Various mental health problem elements, such as elements of major depression disorder or generalized anxiety disorder, might arise in some individual (*pastel-colored dots*). The specific elements that arise in a given time frame (*bright-colored dots*), and their relations to each other (*arrows*), are determined in part by the socioenviromental context, such as a stressful life event involving humiliation (more likely to lead to depression) or endangerment (more likely to lead to general anxiety) (Kendler et al. 2003).

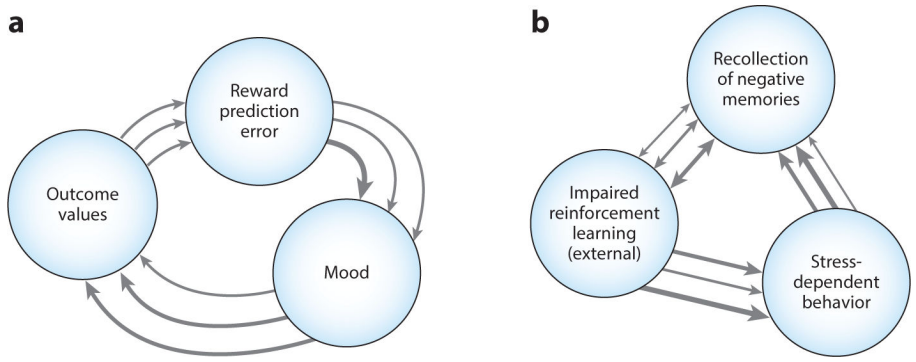

Figure 5.

Theories from recent computational accounts that predict temporal and contextual dynamics in the real world. The figure illustrates theorized interrelations between mental health elements in two recent computational psychiatry accounts. These predict real-world dynamics; hence, data could be collected over time and analyzed (e.g., via network-model representations) in order to test and iteratively refine the theories. The arrows show the theorized direction, and the arrow width the hypothetical strength, of relations for different individuals. This reflects that specific elements of the relationships between the elements may vary among people; e.g., one person may show an especially strong or weak effect of reward prediction error on mood. (*a*) Based on empirical literature on mood and reinforcement learning and computational modeling, Eldar and colleagues recently proposed a positive feedback loop between mood, appraisal of outcomes, and reward prediction error (Eldar & Niv 2015, Mason et al. 2017). (*b*) Based on empirical literature on rumination and stress-dependent behavior, Hitchcock et al. (2021) recently suggested that rumination comprises the recollection and reconsolidation of negative self-referential memories (and other cognitive processes, not depicted). And when rumination takes place at the same time as a potentially important external learning experience, it impairs reinforcement learning about the contingencies. This concurrent process may at once increase the future likelihood of recalling negative memory and engaging in stress-dependent behavior (given that avoiding the latter requires learning adaptive responses to contingencies).