Computational Psychiatry Needs Time and Context

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

1 Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, RI

2 Department of Clinical Psychology, Leiden University, Leiden, The Netherlands

3 Carney Institute for Brain Science, Brown University, Providence, RI

Corresponding author:

Peter F. Hitchcock

Department of Cognitive, Linguistic, and Psychological Sciences

Brown University

Providence, RI 02912

Email: peter_hitchcock@brown.edu

Other author e-mail addresses:

Eiko Fried: eiko.fried@gmail.com

Michael Frank: michael_frank@brown.edu

In press for Annual Review of Psychology, 73 (2022). Please cite as (DOI will become active upon publication):

Hitchcock, P.F., Fried, E.I., Frank, M.J. (in press). Computational Psychiatry Needs Time and Context. Annual Review of Psychology, 73. https://doi.org/10.1146/annurev-psych-021621-124910

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.

Key words: 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 each other? What modeling 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., 2019a), 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 (Maia & Frank, 2011; Wang & Krystal, 2014).

Despite its promise, computational psychiatry has as 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 over-optimism, rather than fundamental limitations. The field also may have had difficulty recognizing how different mental health problems are from each other. As such, computational psychiatry 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 will need to incorporate modeling strategies suited to even the most complex problems (see also, Gillan & Rutledge, 2021; Moutoussis et al., 2017).

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; American Psychiatric Association, 2013). Many phenotypes are poor—they lack reliability, validity, and are highly heterogeneous—and as such permit limited conclusions about mechanisms ("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; Cuthbert & Kozak, 2016; Gillan et al., 2017; Kotov et al., 2017; 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, three prominent alternatives to sharp diagnostic delineation have recently been developed. The Research Domain Criteria (RDoC) assumes that mental health symptoms arise from mixtures of individual differences in cognitive and emotional processes (Cuthbert & Kozak, 2016). The Hierarchical Taxonomy of Psychopathology uses factor analytic methods to investigate symptom co-occurrence patterns across a broad, transdiagnostic space of mental health problems (Kotov et al., 2017). And 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 delineate individuals with one mental health problem from those with another 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 (S. A. 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 versus without the problem (similar to infectious diseases such as measles). In this case, grouping 500 patients diagnosed with Major Depressive Disorder (MDD) into the same category and investigating biological markers compared to a healthy control group would be a sound scientific step. But many mental health problems instead appear to be best understood as complex systems—interactions between **neurocomputational processes (**see Terms and Definitions) and socioenvironmental contexts unfolding over time (Boyd, 1991; Fried & Cramer, 2017; Kendler et al., 2011). These may well differ greatly among the 500 MDD patients just described (Cai, Choi, & Fried, 2020). The utility of essentialist thinking thus depends on the nature of the problem (Brick et al., 2020; McNally, 2020).

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; S. A. 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 will suggest three heuristics for estimating the essentiality of a mental health problem. We will 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, 2020) and essentialist thinking obfuscates the importance of these interactions. **Section 3** will then 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 comprehensive review due to citation limitations.

Terms and Definitions

Neurocomputational process – An input-output transformation and the neural machinery that effects it.

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 that; 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 (such as employing computational modeling strategies and process-pure tasks that can reveal 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 only due 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 of this section are meant to build intuition about indicators and practical consequences of meaningful heterogeneity through a series of examples.

In particular, we will consider Parkinson's Disease, Schizophrenia, and MDD as running examples of high, moderate, and low essentiality disorders. To situate this discussion, we will often draw on the neurocomputational functions of corticostriatal circuitry and dopamine (DA) in decision making, motivation, and reinforcement learning and how dysfunctions or alterations in this circuitry relate to mental health (Maia & Frank, 2011).

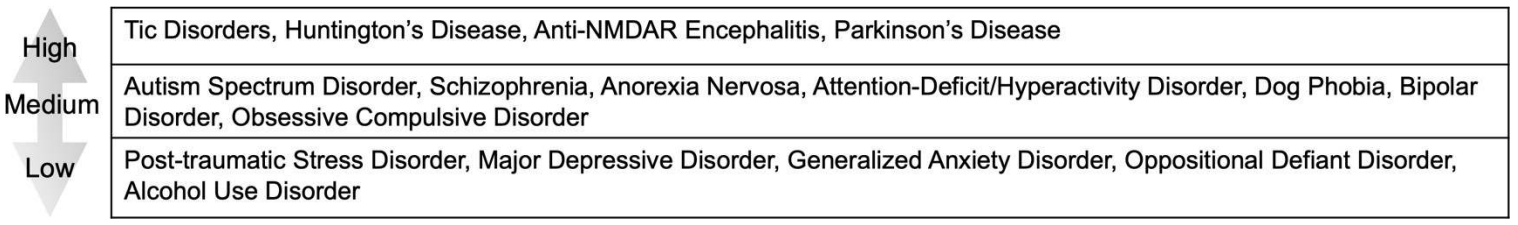


Figure 1. Estimates of whether some well-known mental health problems have high, medium, or low essentiality.

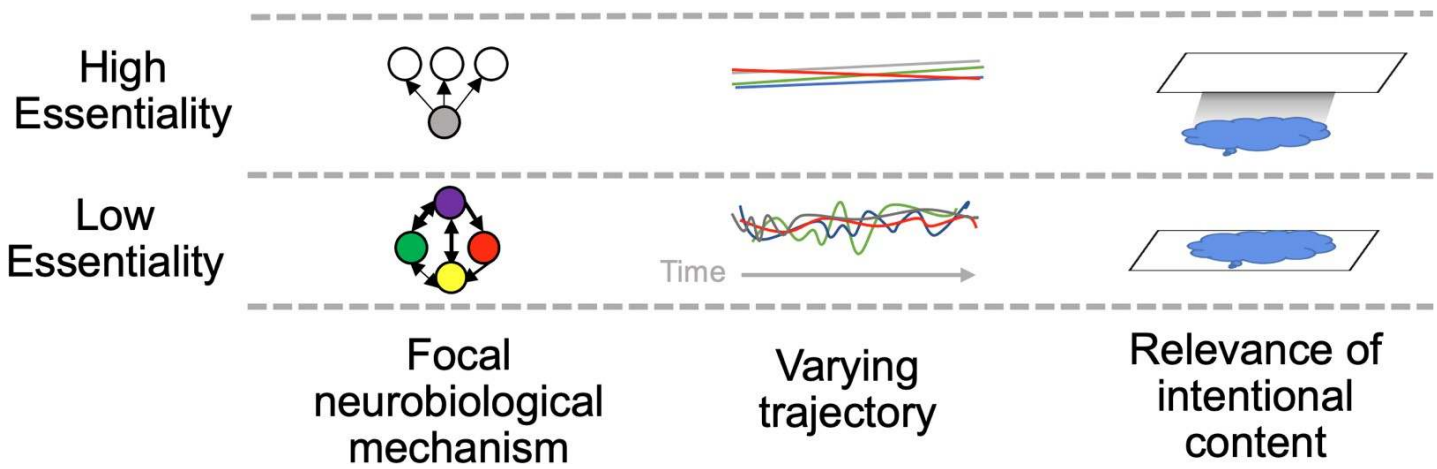


Figure 2. Visualization of three heuristics for estimating essentiality.

1. 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 each other (denoted by arrows of different widths and directions). These elements constitute low-essentiality problems.

2. High-essentiality problems follow a relatively linear and similar natural (i.e., absent intervention) course, whereas lower-essentiality problems follow variable trajectories.

3. Intentional mental content (e.g., negative schemata; represented by the blue thought bubble) is central to low-essentiality problems (e.g., Major Depression; represented by the 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.

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 **Box 1**; see *Parkinson's Disease* below for an example.) 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 20th 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 discovered to underlie many mental health problems (Kendler, 2005). More than a century later, however, this appears quite unlikely; as Kendler (2005, p. 1) 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.

Parkinson's Disease. The relatively high-essentiality disorder Parkinson's Disease 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 to learn from negative compared to positive reward prediction errors (RPEs; Wiecki & Frank, 2010). DA medications reverse these biases by restricting DA levels to an artificially high range, preventing 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 PD, dopamine depletion impacts on not just motor striatal circuits but also on those interacting with prefrontal cortex. Accordingly, in the computational models, this mechanism alters gating not only of motor actions, but also cognitive ones, such as the entrance of cortical content into working memory. Empirical work confirms parallels in how motor actions and working memory content are gated, and that these are related to striatal DA mechanisms in PD (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, separately from the effect of DA on weighting costs vs benefits (Frank et al., 2007). Indeed, deep brain stimulation of the STN reduces the decision threshold and partially remediates motor deficits, but 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. It 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., 2018).

Schizophrenia. DA has also been long implicated in the middle-essentiality problem Schizophrenia (McCutcheon, 2020). Indeed, many of the disorder's positive symptoms can be accounted for by spontaneous striatal DA fluctuations that assign meaning to irrelevant events ("aberrant salience"; Kapur, 2003) and many negative symptoms 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

Box 1: 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, 1992a). We favor a definition proposed by McNally that casts dysfunction as a disrupted process operating within a larger causal system (McNally, 2011; McNally, 2001). 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 to contrast against aberrant functioning. Wachbroit (1994) has argued that a concept of normality is indispensable within biology. Normal function according to this account is not the same as statistically normal (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, 1992b). Rather, normal function by this account refers to an idealized operation of the function against which deviations can be contrasted (Wachbroit, 1994).

evidence also implicates prefrontal cortex (PFC) dysfunction that leads to context-inappropriate behavior (J. D. Cohen & Servan-Schreiber, 1992). To formally model complementary BG and PFC contributions, an extended neural network model included PFC layers that maintain stimulus-outcome associations in working memory "attractor states"; these afford specific representations about the expected values of stimuli and actions, and 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 reward prediction errors (e.g., Collins et al., 2017; Gold et al., 2012**;** Geana et al., in press**)**. 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).

Depression. A wide range of neurocomputational differences have been noted in the relatively low-essentiality problem MDD, including alterations in reward processing and cognitive control tasks, and greater experience of 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. Depression also comprises 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 as dysfunctional (as opposed to adaptive) in light of others and social and environmental factors. For example, rumination has been consistently associated with depression (reviewed in Nolen-Hoeksema et al., 2008). And 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 (though 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 a boundary between dysfunctional thinking (Dayan & Huys, 2008) and constructive thinking that helps to resolve problems, facilitate recovery, and elicit support (Andrews & Thomson, 2009). Similar, 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 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 (Grahek et al., 2019). For example, control may be allocated to self-directed mentation if such thinking is valued (cf. Agrawal, Mattar, Cohen, & Daw, 2020; Andrews & Thomson, 2009). And decreased control could be rationally learned from action-outcome statistics (Lieder 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 lower-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 on 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*—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 but another develops MDD. One reason multinfal problems are challenging to model is that the causes of mental unhealth appear at different causal distances from acute problem onset. Heuristically, these can be classified into distal and proximal factors (things that happen to people, such as one's 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) as well as moderators that determine exactly how a problem unfolds (such as, a problematic behavior crystallizing into a strong habit; Nolen-Hoeksema & Watkins, 2011). In lower-essentiality problems, the dynamic interrelations between these elements (that are operative at different time scales) partly *constitute* the problem (McNally, 2020). 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, yet for which it is still 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 compounded by factors 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 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 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 .74 in men), implying that non-genetic (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 between life stress experienced and resulting symptoms, with humiliating events showing a stronger relationship to MDD and danger a stronger relationship to 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 (such as losing one's spouse) but also dependent stressors (events in which individuals play a role, such as a major fight 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., R. T. Cohen & Kahana, 2020; Van Vugt et al., 2012). Hence, rumination may simultaneously increase the future availability of negative thoughts while decreasing 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 themself 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 will 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 *about* something, such as a belief about oneself, 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 themself. Changing these beliefs may assist in this person's ability to cope, but it would 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 themself 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 (this is sometimes referred to as "moral injury") may report mental health symptoms (e.g., low mood, lost motivation, shame and guilt; Litz, 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 with these changes, eventually larger-scale neural circuits). Yet, there is no reason to think that the specific details of the neural instantiation of these engrams themselves 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 trying to manipulate their brain. Similar, when the causal loci of a mental health problem involve specific intentional mental content, intervening on such content (Eronen, 2020a) 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 people act in ways when they are under pressure that are fundamentally different from the values that they espouse when they have more time to reflect. And the computational expense of certain ways of thinking might also help us to 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 values violations, and other types of intentional content are core to many mental health problems. 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 comprise individual differences transacting with social and environmental contexts over time. Such contexts, rather than dysfunctions or neurocomputational propensities alone, in part determine which mental health elements arise subsequently based on the conclusions that people draw (i.e., the intentional content that emerges) within 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 only *appears* to have low essentiality due to poor phenotyping (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 is key to avoiding confounds in estimating essentiality. **Section 3.1.** will review efforts to improve phenotypic precision in computational psychiatry (**Figure 3**).

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

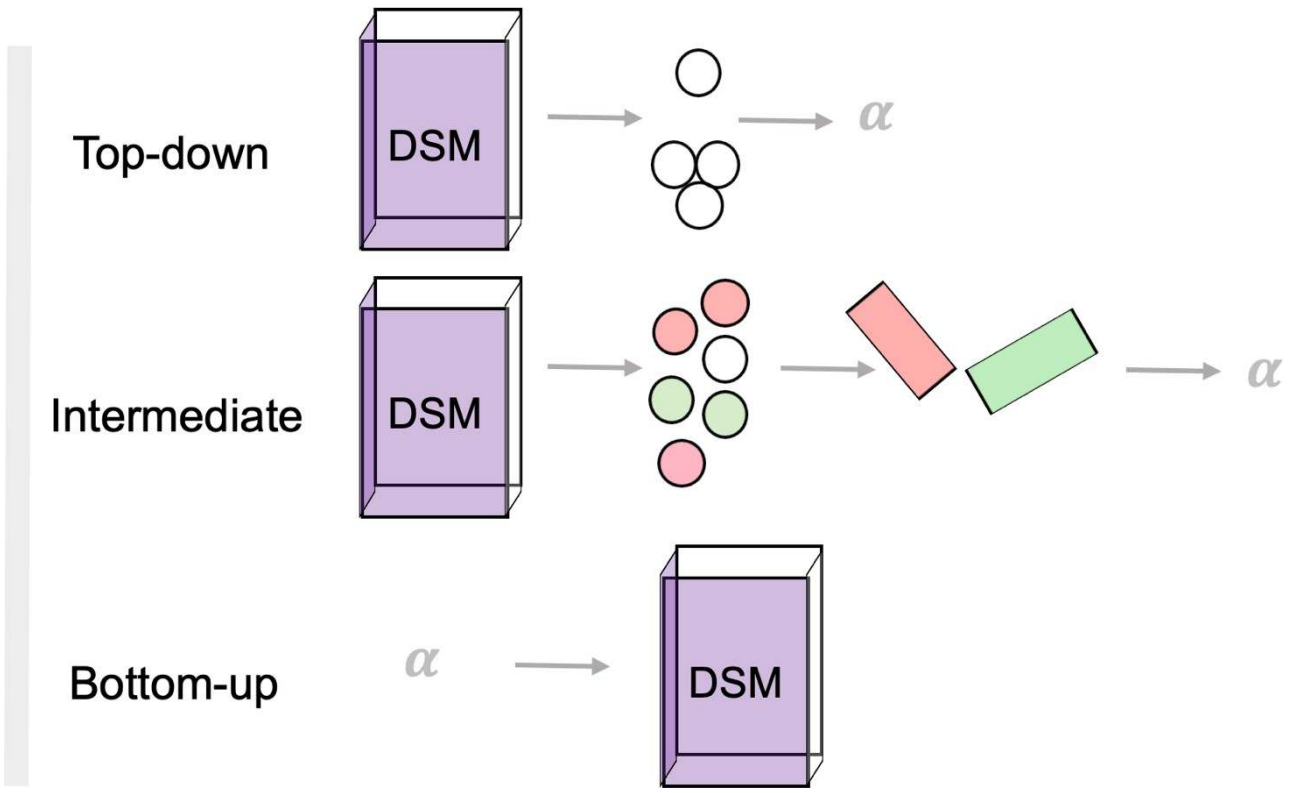


Figure 3. Approaches to improving phenotypic precision.

Top-down approaches begin with symptoms or symptom clusters (represented by white circle and white-circle cluster) and relate these to processes inferred via computational psychiatry methods (such as differences in learning rate, represented by an $\boldsymbol{\alpha}$ parameter).

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.

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 which 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.

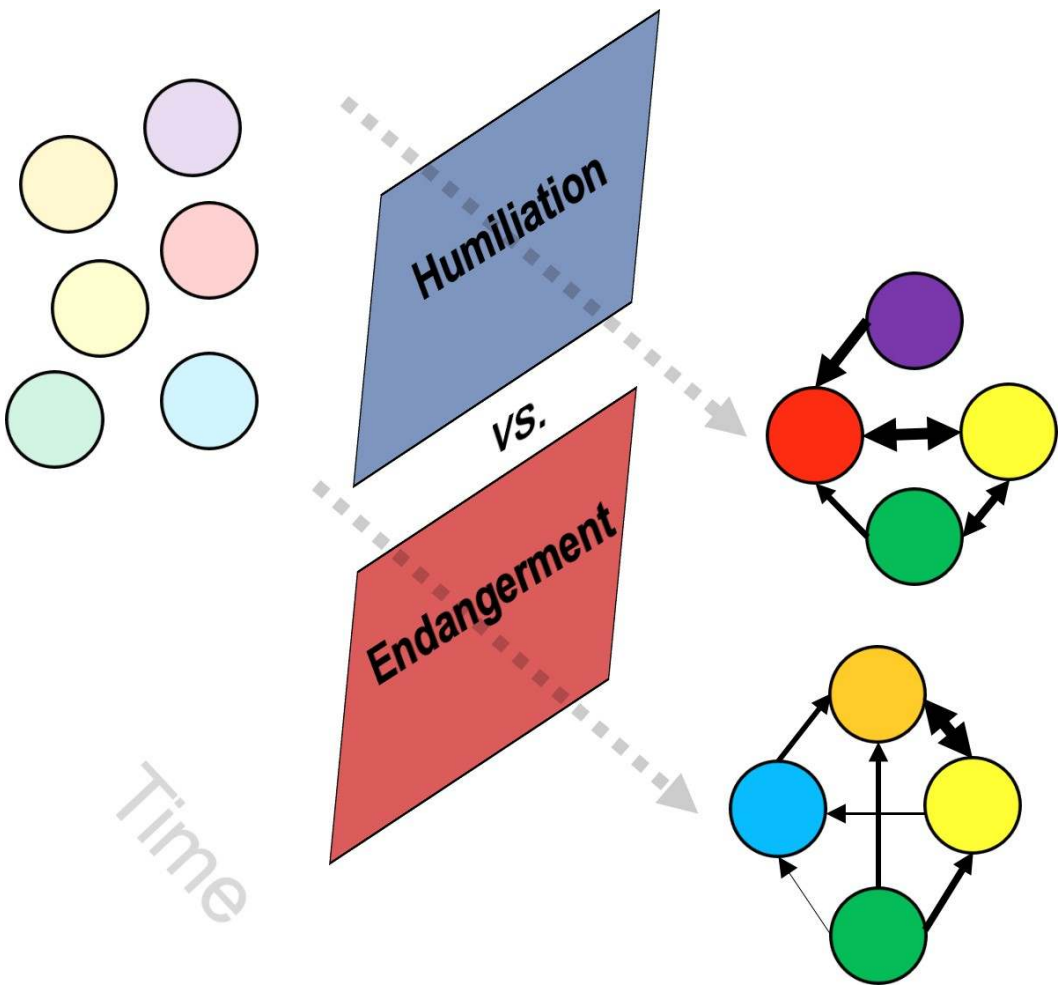


Figure 4. An example of meaningful heterogeneity.

Various mental health problem elements, such as elements of Major Depression Disorder or Generalized Anxiety Disorder, might arise in some individual (partially transparent colored dots). The specific elements that in fact arise in a given time frame (filled-in dots), and their relations to each other (arrows), are determined in part by 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).

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).

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 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 only very weakly related to symptoms such as feeling like a failure, crying, and lost appetite (Beevers et al., 2019). 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—given that these 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 reported that amygdalar activity evoked by computational-model-derived "associability" (increased attention proportional to prediction error, here specifically in a loss condition) was more related to avoidance/numbing and hyperarousal than re-experiencing symptom clusters of PTSD (Brown et al., 2018). However, note that obtaining replicable symptom clusters for common mental health problems has been challenging (e.g., Armour et al., 2015).

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, 1992b). 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 processes previously subsumed under an aggregating construct into primitive units; 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). Similar, 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 could help to stratify pharmacological approaches. Ultimately, we will likely need dynamic, quantitative, aggregative methods to iteratively refine the 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).

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 et al., 2017; Gillan & Daw, 2017; Gillan & Seow, 2020). Studies using this approach have reported specificity in neurocomputational processes associated with distinct regions of phenotypic space (e.g., Gillan et al., 2016; Roualt et al., 2018). For instance, Roualt and colleagues (2018) found, using computational modeling applied to a perceptual decision-making task, that individuals who endorsed more compulsive behavior and intrusive thoughts (as per 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 (via another factor which included apathy symptoms) showed the opposite pattern: less confidence but relatively higher discernment of which choices were correct (Roualt et al., 2018). Parallel to these developments in computational psychiatry, there are efforts underway in clinical science more broadly to delineate relations among symptoms and disorders transdiagnostically, such as the Hierarchical Taxonomy Of Psychopathology (Kotov et al., 2017).

This intermediate approach is not without challenges. For one, dimensional summaries depend (of course) on the questionnaires that they are summarizing. To establish factor structure replicability, computational psychiatrists have tended to use similar questionnaires as an original set of studies by Gillan and colleagues, yet these may not encompass all processes of interest (see Watts et al., 2020, for an interesting perspective on this issue). Gillan and Seow (2020) noted therefore that dimensions from prior studies (and questionnaires from which they are constructed) must be iteratively refined to enable continued progress. Other challenges relate to interpretational and measurement challenges which arise whenever symptom questionnaires are used. Symptoms can co-vary for a number of reasons, and methods finding 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 also complementary strengths and weaknesses. It is also worth noting here 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 (**Box 2**). 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 toward taming this heterogeneity, and thereby modeling mental health in the dimensionality within which even low-essentiality problems reside.

435

Box 2: Algorithmic Modeling Issues and Innovations in Computational Psychiatry

[Supplemental Material in publication]

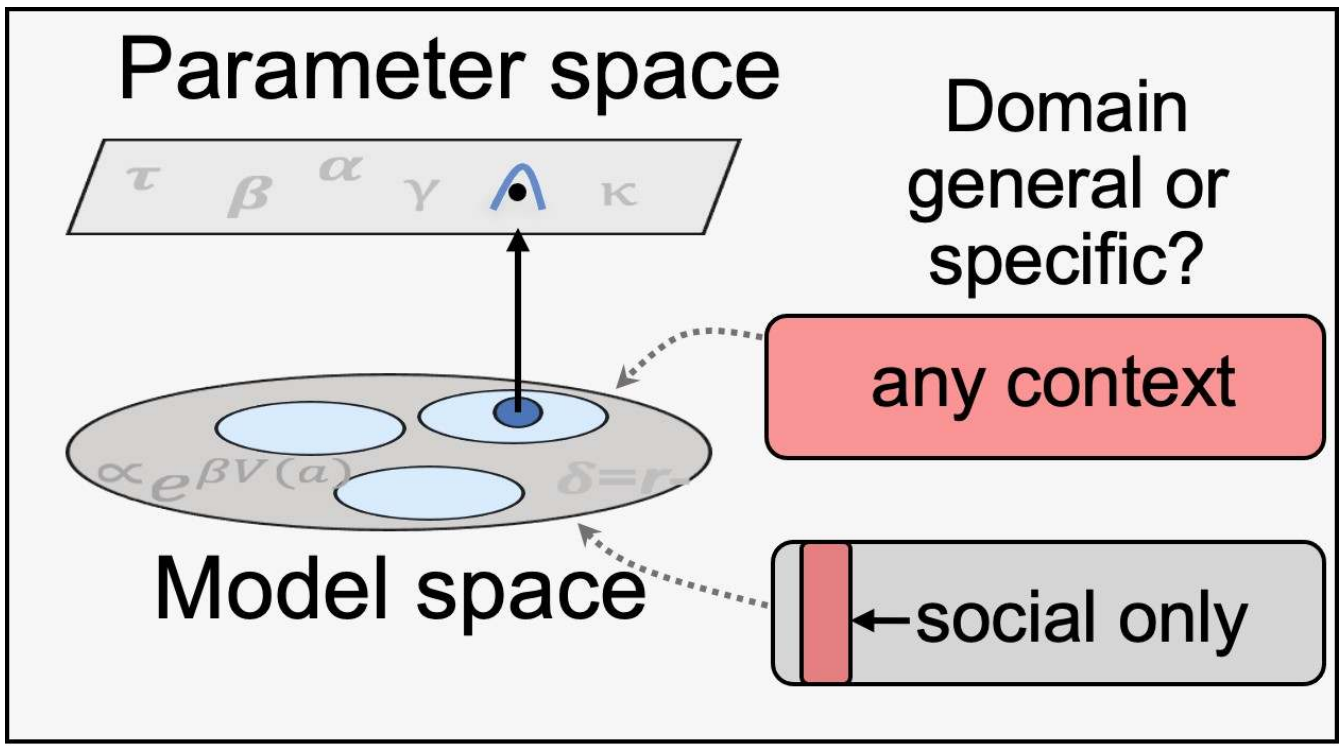
Algorithmic computational models can concisely capture key dynamics, including those that emerge from biologically detailed models, and describe individual differences and within-subject changes relevant to mental health. Such models can also lead to dangerous misinterpretations if used incorrectly, however (see Wilson & Collins, 2019 for a tutorial). Here, we focus on a set of related challenges that are especially relevant to computational psychiatry.

Sharing information across levels and representing uncertainty. Much computational psychiatry research to date has estimated model parameters as fully "random effects" at a local level (e.g., subject and time point). However, more refined hierarchical Bayesian approaches allow parameter estimates for each individual to be informed by the larger group—specifically, by dynamically "shrinking" estimates optimally based on how similar individuals are to each other. This procedure improves parameter estimation even for individuals (A. Gelman & Hill, 2007) and provides estimates of the uncertainty around parameters' values (via posterior distributions). Several toolboxes exist to facilitate hierarchical inference for common reinforcement-learning and decision-making models; for bespoke applications, computational psychiatrists are turning to probabilistic programming languages such as Stan and PyMC3. **Box 2 Figure** depicts the distinction between estimating a point estimate of a parameter's value versus a distribution over its credible values by (respectively) the dot versus distribution in parameter space (that uncertainty estimates in hierarchical Bayesian models are constrained by multiple data levels is not shown).

Moving beyond selecting a single best model. Although hierarchical modeling is increasingly applied to capture clustering within data, one level that is typically still left out is model identity. In particular, researchers typically select a single best model and either shoehorn it onto all subjects or discard subjects (Hitchcock et al., 2017) or they describe participant behavior through a mix of separately estimated models (e.g., Harlé et al., 2017). However, when participant behavior tends to be describable through some mix of models, model identity can also be viewed as a random effect. A recently developed toolbox does just that. It allows researchers to estimate behavior at the group level as arising from a mix of models (e.g., mostly a Reinforcement Learning model with a fixed learning rate, but also some Kalman Filter with a decreasing learning rate) and to weight subject-level parameter estimates from each model proportional to its estimated responsibility for the subject's behavior (Piray et al., 2019). **Box 2 Figure** depicts the distinction between a single "best" model versus the wider space of models under consideration in a given study (respectively) by the blue circle and wider light blue circle.

Moving beyond a limited model space. Another bottleneck in translating from theory to testable models is that the whole set of models that a given study tends to consider has typically itself been limited due to the need to select among models with closed-form likelihood expressions. (A likelihood function maps from an empirically observed data point to a chance of observing it based on a model and set of parameter values.) However, the set of models with such expressions is much smaller than those of theoretical interest. For instance, a common model for fitting choice and reaction time data is the standard drift diffusion model (DDM). In many cases other models may be more appropriate for capturing the underlying decision dynamics (e.g., time-varying decision bounds), yet the DDM is used anyway for convenience. Approximate Bayesian Computation (ABC) methods are gaining traction in computational cognitive neuroscience to overcome this challenge, but traditionally are extremely computationally intensive. Recent developments overcome this challenge by using neural networks to "amortize" expensive computations, thereby allowing a dramatic expansion in the number of hierarchical Bayesian models available to the computational psychiatry community (Fengler et al., 2020). **Box 2 Figure** depicts the set of modeling approaches that are of theoretical interest as the gray model space, within which only a subset of models, represented by light blue circles, are typically fit.

Domain-general or domain-specific differences? Besides these technical issues, there is the computational psychiatry issue of whether alterations are generic or domain-specific (see **Section 3.2**)? This issue, which also determines the choice of algorithmic model, is depicted on the right side of the **Box 2 Figure**.



Box 2. Bottlenecks, sticking points, and a place to draw on theory in algorithmic computational modeling.

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 focus in 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).

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 context. 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). And MDD (and possibly many other internalizing disorders) are 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 which signal transition points in mental health could be detected for prediction and intervened upon for prevention. Exemplifying this possibility, Konova and colleagues 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 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 1-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 in part due to meaningful variation that facilitated prediction (Konova et al., 2020).

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., 2019), 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, 2020a). Those who were slower to learn in the first study reported that they also spent more time quarreling or involved in similar other unpleasant social activities in their everyday lives (Frey et al., 2019). Another interesting finding from this research group was that, in the face-learning task, non-depressed participants who underwent serotonin depletion showed similar patterns of sluggish learning and altered neural activity as the depressed participants (Frey & McCabe, 2020b).

One limitation of these studies is that they did not directly compare social and non-social 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 investigated behavior in a game where participants received an initial monetary endowment then invested portions of it on a trial-wise basis with (they were told) a human partner or slot machine, who 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 "partner"/machine returned was rigged and drifted slowly over time, mimicking real-world situations where 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 slot machines. The use of matched social and non-social 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).

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 reported a tour-de-force example of connecting modeling, real-world behavior, and multimodal measurement. Ten individuals completed a reinforcement-learning task twice per day on their smartphones while electroencephalography and heart-rate data were recorded via portable systems. Computational modeling revealed individual differences related to dissociable fast and slow learning processes; participants with stronger neural decodability (according to machine-learning methods) of the fast-learning process showed an improvement in their mood a few hours later, whereas those with more-decodable slow processes 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 and Rutledge, 2021, for an authoritative review).

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 where 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 vs. 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 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; perhaps surprisingly, the researchers also found that individuals high on sensation seeking and an anxious-depression 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 to “upward-counterfactual” feedback (Hunter, Meer, Gillan, Hsu, & Daw, 2019). Finally, building on behavioral neuroscience research, Mobbs and colleagues have argued that the same animal will tend to employ (within subject) a spectrum of strategies depending on their proximity to threat: from hardwired responses when threat is extremely close to multi-step model-based reasoning when threat is very far

[1] Note that from their earliest work, Daw and colleagues emphasized context should normatively influence the strategy used (Daw et al., 2005).

(Mobbs et al., 2020). 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; Scholl & Klein-Flugge, 2018).

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, 2020; 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 earlier 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, 2020). Additionally, recent network modeling developments may provide more information about the structure of mental health problems and potentially point to novel intervention targets. This includes 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). 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 elements (possibly including a neurocomputational common cause) were not represented in the psychosis network (Haslbeck & Fried, 2017). Methods from complex-systems analysis could also aid our understanding of the structure and dynamics of various problems. These methods build on 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 externally observable dynamic elements of mental health systems. A notable exception are models developed by Eldar and colleagues that produce oscillatory dynamics. These 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 that perceived rewards sometimes far exceed expectations, leading to large positive surprises which send mood rocketing upward, and sometimes fall far short of expectations, leading to crushing disappointments which drive mood downward (Eldar & Niv, 2015; Mason et al., 2017). 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, in turn further increasing mood. This might lead to a slow-but-steady increase in the proportion of felicitous experiences that are experienced, eventually leading to greater well-being over time. Thus, this finding may help to explain the gradual effects of SSRI--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, then tested on data collected in the real world—leading to 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 gloomful and glorifying appraisals of surprising experiences, and

580 domino effects on mood. These could be tested by applying network models, such as moderated (Haslbeck et al.,
581 2019b) network models, to data reported by participants over time, in order to capture varying drug effects or between-
582 subject trajectories related to the mood-biasing parameter.

583 Capturing Person-Specific Processes. Due to the divergent trajectories of lower-essentiality problems (i.e.,
584 multifinality), measuring, modeling, and understanding person-specific patterns are especially important. One striking
585 example of how person-specific patterns can dissociate from group-level patterns is Simpson's Paradox: that, for
586 example, coffee consumption may perfectly positively correlate with neuroticism between subjects, even if the
587 relationship is negative within subjects (i.e., these individuals
588

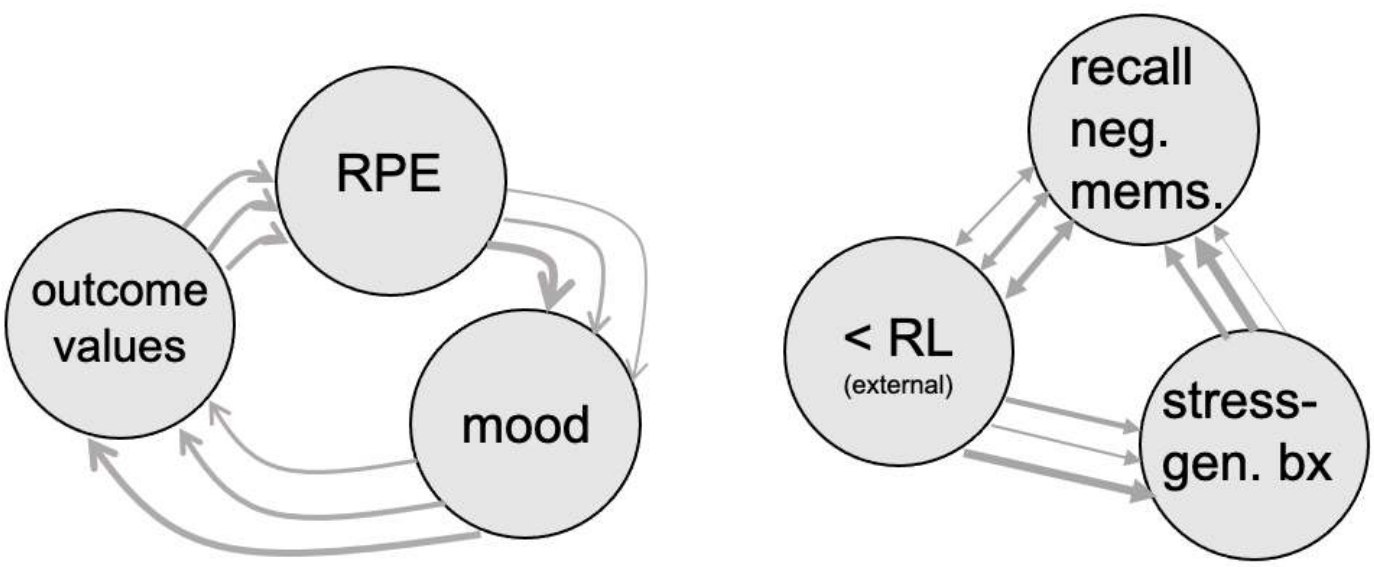


Figure 5. Theories from recent computational accounts that predict temporal and contextual dynamics in the real world.

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) to test and iteratively refine the theories. Arrows show theorized direction and arrow width the hypothetical strength of relations for different individuals. This reflects that specific element relationships may vary among people, such as one person showing an especially strong or weak sensitivity to reward prediction error.

Left: 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).

Right: Based on empirical literature on rumination and stress-generating 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 it takes place at the same time as a potentially important external learning experience, it impairs reinforcement learning (RL) about the contingencies. This concurrent process may at once increase the future likelihood of negative memory recollection and stress-generating behavior.

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 remediative 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, a longitudinal investigation by Konova and colleagues found that opioid use could be predicted by a 1-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 (Konova et al., 2020; see also, Guegen et al., 2021 for discussion of this result).

Hierarchical modeling (including frequentist mixed-effects models and hierarchical Bayesian models; see Box 2) offer a statistically principled approach to modeling between and within-subject effects and enjoy 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 has allowed researchers to corroborate clinical insights, such as the idea that, among neurotic individuals, worry strengthens the duration of 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 in order to overcome problems inherent to the investigation of suitcase constructs, such as worry, which may encompass so many primitive processes that their relationships to other items is confounded (Eronen, 2020b).

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 concurrently extract idiographic patterns (Beltz & Gates, 2017).

An exciting future avenue is to connect these person-specific methods 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 psychotherapy drawn from basic (e.g., computer and decision) sciences (Moutoussis et al., 2018; Niv, Hitchcock, Berwian, & Schoen, in press). It is worth noting that there are natural complementarities between the functional-analytic tradition in behavior therapy, which seeks to understand why behavior occurs in 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, 2020); 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; Simon, 1990).

Formalizing Mental Health Systems. A landmark development toward modeling time and context is the recent development by Robinaugh and colleagues 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 (Robinaugh et al., 2019). 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 then be applied to describe individual differences between people; these are exciting future avenues. 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 for even low-essentiality problems.

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 the above 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 to contextualize observed differences. For instance, do individuals with this problem show differences in trial-and-error learning full stop or are the differences limited to 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 they may take a longer time to make progress on than simpler problems. Computational psychiatry needs time as in temporal dynamics, but it needs time as in patience, too.

6. Algorithmic modeling has a special place in connecting 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 (Box 2).

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.

Disclosures

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

Acknowledgements

Thank you to Zachary Cohen, Romy Frömer, Ivan Grahek, Louis Gularte, Alexander Kline, and Amrita Lamba for feedback on parts of this manuscript and 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. PH was supported by NIMH grant F32 MH123055. MF was supported by NIMH grants P50 MH119467 and R01 MH084840-08A.

Reference Annotations

Andrews, P. W., & Thomson, J. A., Jr. (2009). The bright side of being blue: depression as an adaptation for analyzing complex problems. Psychological Review, 116(3), 620–654. https://doi.org/10.1037/a0016242

Makes the case that depressed mood and rumination facilitate recovery and elicit needed social support.

Eldar, E., & Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. Nature Communications, 6, 6149. https://doi.org/10.1038/ncomms7149

Shows a coupling between affect and appraisal and links individual differences therein to hypomanic symptoms.

Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science, 306(5703), 1940–1943. https://doi.org/10.1126/science.1102941

Shows a mechanistic learning difference in the high-essentiality problem Parkinson's Disease.

Gillan, C. M., & Rutledge, R. B. (2021). Smartphones and the Neuroscience of Mental Health. Annual Review of Neuroscience. https://doi.org/10.1146/annurev-neuro-101220-014053

Two computational psychiatrists make the case for the importance of smartphones in advancing mental-health research.

Hitchcock, P., Forman, E., Rothstein, N. J., Zhang, F., Kounios, J., Niv, Y., & Sims, C. R. (2021). RuminationDerails Reinforcement Learning with Possible Implications for Ineffective Behavior. [PsyArxiv preprint] https://doi.org/10.31234/osf.io/7c2e8

Argues that rumination may impede effective behavior in specific situations by impairing trial-and-error learning.

Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. (2019). Motivation and cognitive control in depression. Neuroscience & Biobehavioral Reviews, 102, 371-381.

Analyzes the sources of apparent cognitive-control impairments in depression.

Kendler, K. S., Hettema, J. M., Butera, F., Gardner, C. O., & Prescott, C. A. (2003). Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. Archives of General Psychiatry, 60(8), 789–796. https://doi.org/10.1001/archpsyc.60.8.789

Shows some specificity in the life events precipitating MDD and GAD (humiliation and endangerment, respectively).

Konova, A. B., Lopez-Guzman, S., Urmanche, A., Ross, S., Louie, K., Rotrosen, J., & Glimcher, P. W. (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), 368–377. https://doi.org/10.1001/jamapsychiatry.2019.4013

Remarkable study showing variation in the propensity to take ambiguous risk predicts prospective drug use.

Nolen-Hoeksema, S., & Watkins, E. R. (2011). A Heuristic for Developing Transdiagnostic Models of Psychopathology: Explaining Multifinality and Divergent Trajectories. Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 6(6), 589–609. https://doi.org/10.1177/1745691611419672

How to think about multifinality by distinguishing the factors that contribute to mental health problems.

Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., McNally, R. J., van Nes, E. H., Scheffer, M., Kendler, K. S., & Others. (2019). Advancing the network theory of mental disorders: A computational model of panic disorder. https://psyarxiv.com/km37w/

Empirically informed, quantitative representation of panic disorder as a network of interacting elements.

References

Agrawal, M., Mattar, M. G., Cohen, J. D., & Daw, N. D. (2020). The temporal dynamics of opportunity costs: A normative account of cognitive fatigue and boredom. bioRxiv. doi: https://doi.org/10.1101/2020.09.08.287276

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.

Amundson, R. (2000). Against normal function. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 31(1), 33–53.

Andrews, P. W., & Thomson, J. A., Jr. (2009). The bright side of being blue: depression as an adaptation for analyzing complex problems. Psychological Review, 116(3), 620–654. https://doi.org/10.1037/a0016242

Armour, C., Műllerová, J., & Elhai, J. D. (2015). A systematic literature review of PTSD's latent structure in the Diagnostic and Statistical Manual of Mental Disorders: DSM-IV to DSM-5. Clinical Psychology Review, 44, 60–74. https://doi.org/10.1016/j.cpr.2015.12.003

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. Nature Human Behaviour, 1(4), 1–10.

Barlow, D. H., & Hersen, M. (1973). Single-case experimental designs. Uses in applied clinical research. Archives of General Psychiatry, 29(3), 319–325. https://doi.org/10.1001/archpsyc.1973.04200030017003

Beeler, J. A., Frank, M. J., McDaid, J., Alexander, E., Turkson, S., Bernardez Sarria, M. S., McGehee, D. S., & Zhuang, X. (2012). A role for dopamine-mediated learning in the pathophysiology and treatment of Parkinson's disease. Cell Reports, 2(6), 1747–1761. https://doi.org/10.1016/j.celrep.2012.11.014

Beevers, C. G., Mullarkey, M. C., Dainer-Best, J., Stewart, R. A., Labrada, J., Allen, J. J. B., McGeary, J. E., & Shumake, J. (2019). Association between negative cognitive bias and depression: A symptom-level approach. Journal of Abnormal Psychology, 128(3), 212–227. https://doi.org/10.1037/abn0000405

Beltz, A. M., & Gates, K. M. (2017). Network Mapping with GIMME. Multivariate Behavioral Research, 52(6), 789–804. https://doi.org/10.1080/00273171.2017.1373014

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. Journal of Clinical Psychology, 64(9), 1089–1108. https://doi.org/10.1002/jclp.20503

Boyd, R. (1991). Realism, anti-foundationalism and the enthusiasm for natural kinds. Philosophical Studies, 61(1-2), 127–148. https://doi.org/10.1007/bf00385837

Brick, C., Hood, B., Ekroll, V., & de-Wit, L. (2020). Illusory essences: A bias holding back theorizing in psychological science. https://doi.org/10.31234/osf.io/eqma4

Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model: Reconsidering the network approach to psychopathology. Psychological Review, 125(4), 606–615. https://doi.org/10.1037/rev0000108

Bringmann, L. F., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, F., Borsboom, D., & Tuerlinckx, F. (2013). A network approach to psychopathology: new insights into clinical longitudinal data. PloS One, 8(4), e60188. https://doi.org/10.1371/journal.pone.0060188

Browning, M., Carter, C. S., Chatham, C., Den Ouden, H., Gillan, C. M., Baker, J. T., Chekroud, A. M., Cools, R., Dayan, P., Gold, J., Goldstein, R. Z., Hartley, C. A., Kepecs, A., Lawson, R. P., Mourao-Miranda, J., Phillips, M. L., Pizzagalli, D. A., Powers, A., Rindskopf, D., … Paulus, M. (2020). Realizing the Clinical Potential of Computational Psychiatry: Report From the Banbury Center Meeting, February 2019. Biological Psychiatry, 88(2), e5–e10. https://doi.org/10.1016/j.biopsych.2019.12.026

Brown, V. M., Zhu, L., Wang, J. M., Frueh, B. C., King-Casas, B., & Chiu, P. H. (2018). Associability-modulated loss learning is increased in posttraumatic stress disorder. eLife, 7. https://doi.org/10.7554/eLife.30150

Burger, J., van der Veen, D. C., Robinaugh, D. J., Quax, R., Riese, H., Schoevers, R. A., & Epskamp, S. (2020). Bridging the gap between complexity science and clinical practice by formalizing idiographic theories: a computational model of functional analysis. BMC Medicine, 18(1), 99. https://doi.org/10.1186/s12916-020-01558-1

Cai, N., Choi, K. W., & Fried, E. I. (2020). Reviewing the genetics of heterogeneity in depression: operationalizations, manifestations and etiologies. Human molecular genetics, 29 (R1), R10-R18.

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nature Neuroscience, 14(11), 1462–1467. https://doi.org/10.1038/nn.2925

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. Development and Psychopathology, 8(4), 597–600. https://doi.org/10.1017/S0954579400007318

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychological Review, 99(1), 45–77. https://doi.org/10.1037/0033-295x.99.1.45

Cohen, R. T., & Kahana, M. J. (2020). A memory-based theory of emotional disorders. bioRxiv. https://doi.org/10.1101/817486

Collins, A. G. E., Albrecht, M. A., Waltz, J. A., Gold, J. M., & Frank, M. J. (2017). Interactions Among Working Memory, Reinforcement Learning, and Effort in Value-Based Choice: A New Paradigm and Selective Deficits in Schizophrenia. Biological Psychiatry, 82(6), 431–439. https://doi.org/10.1016/j.biopsych.2017.05.017

Collins, A. G. E., & Frank, M. J. (2014). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychological Review, 121(3), 337–366. https://doi.org/10.1037/a0037015

Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. Cerebral Cortex , 11(12), 1136–1143. https://doi.org/10.1093/cercor/11.12.1136

Dasgupta, I., & Gershman, S. J. (2021). Memory as a Computational Resource. Trends in Cognitive Sciences, 25(3), 240–251. https://doi.org/10.1016/j.tics.2020.12.008

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron, 69(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience, 8(12), 1704–1711. https://doi.org/10.1038/nn1560

Dayan, P., & Huys, Q. J. M. (2008). Serotonin, inhibition, and negative mood. PLoS Computational Biology, 4(2), e4. https://doi.org/10.1371/journal.pcbi.0040004

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016). From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. Psychological Science, 27(6), 848–858. https://doi.org/10.1177/0956797616639301

Dowd, E. C., Frank, M. J., Collins, A., Gold, J. M., & Barch, D. M. (2016). Probabilistic Reinforcement Learning in Patients With Schizophrenia: Relationships to Anhedonia and Avolition. Biological Psychiatry. Cognitive Neuroscience and Neuroimaging, 1(5), 460–473. https://doi.org/10.1016/j.bpsc.2016.05.005

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., ... & Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nature medicine, 23(1), 28-38.

Durstewitz, D., Huys, Q. J. M., & Koppe, G. (2020). Psychiatric Illnesses as Disorders of Network Dynamics. Biological Psychiatry. Cognitive Neuroscience and Neuroimaging. https://doi.org/10.1016/j.bpsc.2020.01.001

Durstewitz, D., & Seamans, J. K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. Biological Psychiatry, 64(9), 739–749. https://doi.org/10.1016/j.biopsych.2008.05.015

Egerton, A., Valmaggia, L. R., Howes, O. D., Day, F., Chaddock, C. A., Allen, P., Winton-Brown, T. T., Bloomfield, M. A. P., Bhattacharyya, S., Chilcott, J., Lappin, J. M., Murray, R. M., & McGuire, P. (2016). Adversity in childhood linked to elevated striatal dopamine function in adulthood. Schizophrenia Research, 176(2-3), 171–176. https://doi.org/10.1016/j.schres.2016.06.005

Eldar, E., & Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. Nature Communications, 6, 6149. https://doi.org/10.1038/ncomms7149

Eldar, E., Roth, C., Dayan, P., & Dolan, R. J. (2018). Decodability of Reward Learning Signals Predicts Mood Fluctuations. Current Biology: CB, 28(9), 1433–1439.e7. https://doi.org/10.1016/j.cub.2018.03.038

Eronen, M. I. (2019). The levels problem in psychopathology. Psychological Medicine, 1–7. https://doi.org/10.1017/S0033291719002514

Eronen, M. I. (2020a). Interventionism for the intentional stance: True believers and their brains. Topoi. An International Review of Philosophy, 39(1), 45–55.

Eronen, M. I. (2020b). Causal discovery and the problem of psychological interventions. New Ideas in Psychology, 59, 100785. https://doi.org/10.1016/j.newideapsych.2020.100785

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2020). Likelihood Approximation Networks (LANs) for Fast Inference of Simulation Models in Cognitive Neuroscience. In Cold Spring Harbor Laboratory (p. 2020.11.20.392274). https://doi.org/10.1101/2020.11.20.392274

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. Journal of Cognitive Neuroscience, 17(1), 51–72. https://doi.org/10.1162/0898929052880093

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. Psychological Review, 113(2), 300–326. https://doi.org/10.1037/0033-295X.113.2.300

Frank, M. J., Gagne, C., Nyhus, E., Masters, S., Wiecki, T. V., Cavanagh, J. F., & Badre, D. (2015). fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. Journal of Neuroscience, 35(2), 485-494.

Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science, 318(5854), 1309–1312. https://doi.org/10.1126/science.1146157

Frank, M. J., Seeberger, L. C., & O'reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science, 306(5703), 1940–1943. https://doi.org/10.1126/science.1102941

Frey, A.-L., Frank, M. J., & McCabe, C. (2019). Social reinforcement learning as a predictor of real-life experiences in individuals with high and low depressive symptomatology. Psychological Medicine, 1–8. https://doi.org/10.1017/S0033291719003222

Frey, A.-L., & McCabe, C. (2020a). Impaired social learning predicts reduced real-life motivation in individuals with depression: A computational fMRI study. Journal of Affective Disorders, 263, 698–706. https://doi.org/10.1016/j.jad.2019.11.049

Frey, A.-L., & McCabe, C. (2020b). Effects of serotonin and dopamine depletion on neural prediction computations during social learning. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 45(9), 1431–1437. https://doi.org/10.1038/s41386-020-0678-z

Fried, E. I., & Cramer, A. O. J. (2017). Moving Forward: Challenges and Directions for Psychopathological Network Theory and Methodology. Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 12(6), 999–1020. https://doi.org/10.1177/1745691617705892

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. BMC Medicine, 13, 72. https://doi.org/10.1186/s12916-015-0325-4

Geana, A., Barch, D., Gold, J., Carter, C., MacDonald, A., Ragland, D., Silverstein, S., & Frank, M. (in press). Using Computational Modelling to Capture Schizophrenia-Specific Reinforcement Learning Differences and Their Implications on Patient Classification.

Gelman, A., & Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Gelman, S. A. (2004). Psychological essentialism in children. Trends in Cognitive Sciences, 8(9), 404–409. https://doi.org/10.1016/j.tics.2004.07.001

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. Science, 349(6245), 273–278.

https://doi.org/10.1126/science.aac6076

Gillan, C. M., Fineberg, N. A., & Robbins, T. W. (2017). A trans-diagnostic perspective on obsessive-compulsive disorder. Psychological Medicine, 47(9), 1528–1548. https://doi.org/10.1017/S0033291716002786

Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A., & Daw, N. D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife, 5. https://doi.org/10.7554/eLife.11305

Gillan, C. M., Otto, A. R., Phelps, E. A., & Daw, N. D. (2015). Model-based learning protects against forming habits. Cognitive, Affective & Behavioral Neuroscience, 15(3), 523–536. https://doi.org/10.3758/s13415-015-0347-6

Gillan, C. M., & Rutledge, R. B. (2021). Smartphones and the Neuroscience of Mental Health. Annual Review of Neuroscience. https://doi.org/10.1146/annurev-neuro-101220-014053

Gillan, C. M., & Seow, T. X. F. (2020). Carving Out New Transdiagnostic Dimensions for Research in Mental Health. Biological Psychiatry. Cognitive Neuroscience and Neuroimaging, 5(10), 932–934. https://doi.org/10.1016/j.bpsc.2020.04.013

Gold, J. M., Waltz, J. A., & Frank, M. J. (2015). Effort cost computation in schizophrenia: a commentary on the recent literature. Biological Psychiatry, 78(11), 747–753. https://doi.org/10.1016/j.biopsych.2015.05.005

Gold, J. M., Waltz, J. A., Matveeva, T. M., Kasanova, Z., Strauss, G. P., Herbener, E. S., Collins, A. G. E., & Frank, M. J. (2012). Negative symptoms and the failure to represent the expected reward value of actions: behavioral and computational modeling evidence. Archives of General Psychiatry, 69(2), 129–138. https://doi.org/10.1001/archgenpsychiatry.2011.1269

Goldstein, B. L., & Klein, D. N. (2014). A review of selected candidate endophenotypes for depression. Clinical Psychology Review, 34(5), 417–427. https://doi.org/10.1016/j.cpr.2014.06.003

Grahek, I., Shenhav, A., Musslick, S., Krebs, R. M., & Koster, E. H. (2019). Motivation and cognitive control in depression. Neuroscience & Biobehavioral Reviews, 102, 371-381.

Gueguen, M. C. M., Schweitzer, E. M., & Konova, A. B. (2021). Computational theory-driven studies of reinforcement learning and decision-making in addiction: what have we learned? Current Opinion in Behavioral Sciences, 38, 40–48. https://doi.org/10.1016/j.cobeha.2020.08.007

Hammen, C. (2005). Stress and depression. Annual Review of Clinical Psychology, 1, 293–319. https://doi.org/10.1146/annurev.clinpsy.1.102803.143938

Harlé, K. M., Guo, D., Zhang, S., Paulus, M. P., & Yu, A. J. (2017). Anhedonia and anxiety underlying depressive symptomatology have distinct effects on reward-based decision-making. PloS One, 12(10), e0186473. https://doi.org/10.1371/journal.pone.0186473

Haslbeck, J. M. B., Borsboom, D., & Waldorp, L. J. (2019). Moderated Network Models. Multivariate Behavioral Research, 1–32. https://doi.org/10.1080/00273171.2019.1677207

Haslbeck, J. M. B., & Fried, E. I. (2017). How predictable are symptoms in psychopathological networks? A reanalysis of 18 published datasets. Psychological Medicine, 47(16), 2767–2776. https://doi.org/10.1017/S0033291717001258

Haslbeck, J. M. B., Ryan, O., Robinaugh, D., Waldorp, L., & Borsboom, D. (2019). Modeling Psychopathology: From Data Models to Formal Theories. https://doi.org/10.31234/osf.io/jgm7f

Henry, T. R., Robinaugh, D., & Fried, E. I. (2020). On the control of psychological networks. https://doi.org/10.31234/osf.io/7vpz2

Herz, D. M., Zavala, B. A., Bogacz, R., & Brown, P. (2016). Neural Correlates of Decision Thresholds in the Human Subthalamic Nucleus. Current Biology: CB, 26(7), 916–920. https://doi.org/10.1016/j.cub.2016.01.051

Hitchcock, P., Radulescu, A., Niv, Y., Sims, C. (2017). Translating a reinforcement learning task into a computational psychiatry assay: Challenges and strategies. Proceedings of the Cognitive Science Society, 39, 2217-2222.

Hitchcock, P., Forman, E., Rothstein, N. J., Zhang, F., Kounios, J., Niv, Y., & Sims, C. R. (2021). Rumination Derails Reinforcement Learning with Possible Implications for Ineffective Behavior. [PsyArxiv preprint] https://doi.org/10.31234/osf.io/7c2e8

Hofmann, S. G., & Hayes, S. C. (2019). The Future of Intervention Science: Process-Based Therapy. Clinical Psychological Science, 7(1), 37–50. https://doi.org/10.1177/2167702618772296

Hunter, L. E., Meer, E. A., Gillan, C. M., Hsu, M., & Daw, N. D. (2019). Excessive deliberation in social anxiety. In Cold Spring Harbor Laboratory (p. 522433). https://doi.org/10.1101/522433

Huys, Q. J. M., Browning, M., Paulus, M. P., & Frank, M. J. (2021). Advances in the computational understanding of mental illness. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 46(1), 3–19. https://doi.org/10.1038/s41386-020-0746-4

Huys, Q. J. M., Daw, N. D., & Dayan, P. (2015). Depression: a decision-theoretic analysis. Annual Review of Neuroscience, 38, 1–23. https://doi.org/10.1146/annurev-neuro-071714-033928

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. Nature Neuroscience, 19(3), 404–413. https://doi.org/10.1038/nn.4238

Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., & Pizzagalli, D. A. (2015). Large-Scale Network Dysfunction in Major Depressive Disorder: A Meta-analysis of Resting-State Functional Connectivity. JAMA Psychiatry, 72(6), 603–611. https://doi.org/10.1001/jamapsychiatry.2015.0071

Kapur, S. (2003). Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. The American Journal of Psychiatry, 160(1), 13–23. https://doi.org/10.1176/appi.ajp.160.1.13

Kendler, K. S. (2005). Toward a philosophical structure for psychiatry. The American Journal of Psychiatry, 162(3), 433–440. https://doi.org/10.1176/appi.ajp.162.3.433

Kendler, K. S., Gardner, C. O., Gatz, M., & Pedersen, N. L. (2007). The sources of co-morbidity between major depression and generalized anxiety disorder in a Swedish national twin sample. Psychological Medicine, 37(3), 453–462. https://doi.org/10.1017/S0033291706009135

Kendler, K. S., Hettema, J. M., Butera, F., Gardner, C. O., & Prescott, C. A. (2003). Life event dimensions of loss, humiliation, entrapment, and danger in the prediction of onsets of major depression and generalized anxiety. Archives of General Psychiatry, 60(8), 789–796. https://doi.org/10.1001/archpsyc.60.8.789

Kendler, K. S., Thornton, L. M., & Gardner, C. O. (2000). Stressful Life Events and Previous Episodes in the Etiology of Major Depression in Women: An Evaluation of the “Kindling” Hypothesis. American Journal of Psychiatry, 157(8), 1243–1251. https://doi.org/10.1176/appi.ajp.157.8.1243

Kendler, K. S., Zachar, P., & Craver, C. (2011). What kinds of things are psychiatric disorders? Psychological Medicine, 41(6), 1143–1150. https://doi.org/10.1017/S0033291710001844

Keramati, M., & Smittenaar, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal–directed spectrum. Proceedings of the. https://www.pnas.org/content/113/45/12868.short

Keren, H., O’Callaghan, G., Vidal-Ribas, P., Buzzell, G. A., Brotman, M. A., Leibenluft, E., Pan, P. M., Meffert, L., Kaiser, A., Wolke, S., Pine, D. S., & Stringaris, A. (2018). Reward Processing in Depression: A Conceptual and Meta-Analytic Review Across fMRI and EEG Studies. The American Journal of Psychiatry, 175(11), 1111–1120. https://doi.org/10.1176/appi.ajp.2018.17101124

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson’s paradox in psychological science: a practical guide. Frontiers in Psychology, 4, 513. https://doi.org/10.3389/fpsyg.2013.00513

Konova, A. B., Lopez-Guzman, S., Urmanche, A., Ross, S., Louie, K., Rotrosen, J., & Glimcher, P. W. (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), 368–377. https://doi.org/10.1001/jamapsychiatry.2019.4013

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. Journal of Abnormal Psychology, 126(4), 454–477. https://doi.org/10.1037/abn0000258

Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH research domain criteria initiative: background, issues, and

pragmatics. Psychophysiology, 53(3), 286-297.

Lamba, A., Frank, M. J., & FeldmanHall, O. (2020). Anxiety Impedes Adaptive Social Learning Under Uncertainty. Psychological Science, 31(5), 592–603. https://doi.org/10.1177/0956797620910993

Leising, D., Burger, J., Zimmermann, J., Bäckström, M., Oltmanns, J. R., & Connelly, B. S. (2020). Why do items correlate with one another? A conceptual analysis with relevance for general factors and network models. https://doi.org/10.31234/osf.io/7c895

Lieder, F., Goodman, N. D., & Huys, Q. J. M. (2013). Learned helplessness and generalization. Proceedings of the Annual Meeting of the Cognitive Science Society, 35. https://escholarship.org/content/qt31362551/qt31362551.pdf

Litz, B. T., Stein, N., Delaney, E., Lebowitz, L., Nash, W. P., Silva, C., & Maguen, S. (2009). Moral injury and moral repair in war veterans: a preliminary model and intervention strategy. Clinical Psychology Review, 29(8), 695–706. https://doi.org/10.1016/j.cpr.2009.07.003

Lydon-Staley, D. M., Cornblath, E. J., Blevins, A. S., & Bassett, D. S. (2021). Modeling brain, symptom, and behavior in the winds of change. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 46(1), 20–32. https://doi.org/10.1038/s41386-020-00805-6

Lydon-Staley, D. M., Xia, M., Mak, H. W., & Fosco, G. M. (2019). Adolescent Emotion Network Dynamics in Daily Life and Implications for Depression. Journal of Abnormal Child Psychology, 47(4), 717–729. https://doi.org/10.1007/s10802-018-0474-y

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. Nature Neuroscience, 14(2), 154–162. https://doi.org/10.1038/nn.2723

Maia, T. V., & Frank, M. J. (2017). An Integrative Perspective on the Role of Dopamine in Schizophrenia. Biological Psychiatry, 81(1), 52–66. https://doi.org/10.1016/j.biopsych.2016.05.021

Mason, L., Eldar, E., & Rutledge, R. B. (2017). Mood Instability and Reward Dysregulation—A Neurocomputational Model of Bipolar Disorder. JAMA Psychiatry , 74(12), 1275–1276. https://doi.org/10.1001/jamapsychiatry.2017.3163

Matar, E., Shine, J. M., Gilat, M., Ehgoetz Martens, K. A., Ward, P. B., Frank, M. J., ... & Lewis, S. J. (2019). Identifying the neural correlates of doorway freezing in Parkinson's disease. Human brain mapping, 40(7), 2055-2064.

McCabe, C. (2018). Linking anhedonia symptoms with behavioural and neural reward responses in adolescent depression. Current Opinion in Behavioral Sciences, 22, 143–151. https://doi.org/10.1016/j.cobeha.2018.07.001

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate neural systems value immediate and delayed monetary rewards. Science, 306(5695), 503–507. https://doi.org/10.1126/science.1100907

McCutcheon, R. A., Marques, T. R., & Howes, O. D. (2020). Schizophrenia—An Overview. JAMA Psychiatry , 77(2), 201–210. https://doi.org/10.1001/jamapsychiatry.2019.3360

McNally, R. J. (2001). On Wakefield’s harmful dysfunction analysis of mental disorder [Review of On Wakefield’s harmful dysfunction analysis of mental disorder]. Behaviour Research and Therapy, 39(3), 309–314. https://doi.org/10.1016/s0005-7967(00)00068-1

McNally, R. J. (2011). What Is Mental Illness? Cambridge, MA: Harvard University Press.

McNally, R. J. (2020). Network Analysis of Psychopathology: Controversies and Challenges. Annual Review of Clinical Psychology. https://doi.org/10.1146/annurev-clinpsy-081219-092850

Michely, J., Eldar, E., Martin, I. M., & Dolan, R. J. (2020). A mechanistic account of serotonin’s impact on mood. Nature Communications, 11(1), 2335. https://doi.org/10.1038/s41467-020-16090-2

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

Mobbs, D., Headley, D. B., Ding, W., & Dayan, P. (2020). Space, Time, and Fear: Survival Computations along Defensive Circuits. Trends in Cognitive Sciences, 24(3), 228–241. https://doi.org/10.1016/j.tics.2019.12.016

Montague, P. R., Dolan, R. J., Friston, K. J., & Dayan, P. (2012). Computational psychiatry. Trends in Cognitive Sciences, 16(1), 72–80. https://doi.org/10.1016/j.tics.2011.11.018

Moutoussis, M., Eldar, E., & Dolan, R. J. (2017). Building a New Field of Computational Psychiatry. Biological Psychiatry, 82(6), 388–390. https://doi.org/10.1016/j.biopsych.2016.10.007

Moutoussis, M., Shahar, N., Hauser, T. U., & Dolan, R. J. (2018). Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies. Computational Psychiatry (Cambridge, Mass.), 2, 50–73. https://doi.org/10.1162/CPSY_a_00014

Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W., & Hartmann, J. A. (2017). Moving From Static to Dynamic Models of the Onset of Mental Disorder: A Review. JAMA Psychiatry , 74(5), 528–534. https://doi.org/10.1001/jamapsychiatry.2017.0001

Y Niv, P Hitchcock, IM Berwian & G Schoen (in press). Toward Precision Cognitive Behavioral Therapy via Reinforcement Learning Theory. In L.M. Williams & L.M. Hack (Eds.) Precision Psychiatry (Ch. 12). Washington, D.C.: American Psychiatric Association.

Nolen-Hoeksema, S., & Watkins, E. R. (2011). A Heuristic for Developing Transdiagnostic Models of Psychopathology: Explaining Multifinality and Divergent Trajectories. Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 6(6), 589–609. https://doi.org/10.1177/1745691611419672

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking Rumination. Perspectives on Psychological Science: A Journal of the Association for Psychological Science, 3(5), 400–424. https://doi.org/10.1111/j.1745-6924.2008.00088.x

Norman, S. B., Wilkins, K. C., Myers, U. S., & Allard, C. B. (2014). Trauma Informed Guilt Reduction Therapy With Combat Veterans. Cognitive and Behavioral Practice, 21(1), 78–88. https://doi.org/10.1016/j.cbpra.2013.08.001

Patzelt, E. H., Kool, W., Millner, A. J., & Gershman, S. J. (2019). Incentives Boost Model-Based Control Across a Range of Severity on Several Psychiatric Constructs. Biological Psychiatry, 85(5), 425–433. https://doi.org/10.1016/j.biopsych.2018.06.018

Perivoliotis, D., Morrison, A. P., Grant, P. M., French, P., & Beck, A. T. (2009). Negative performance beliefs and negative symptoms in individuals at ultra-high risk of psychosis: a preliminary study. Psychopathology, 42(6), 375–379. https://doi.org/10.1159/000236909

Piray, P., Dezfouli, A., Heskes, T., Frank, M. J., & Daw, N. D. (2019). Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. PLoS Computational Biology, 15(6), e1007043. https://doi.org/10.1371/journal.pcbi.1007043

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. Annual Review of Psychology, 63, 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Poldrack, R. A., & Yarkoni, T. (2016). From Brain Maps to Cognitive Ontologies: Informatics and the Search for Mental Structure. Annual Review of Psychology, 67, 587–612. https://doi.org/10.1146/annurev-psych-122414-033729

Pulcu, E., & Browning, M. (2017). Using Computational Psychiatry to Rule Out the Hidden Causes of Depression [Review of Using Computational Psychiatry to Rule Out the Hidden Causes of Depression]. JAMA Psychiatry, 74(8), 777–778.jamanetwork.com. https://doi.org/10.1001/jamapsychiatry.2017.1500

Radulescu, A., & Niv, Y. (2019). State representation in mental illness. *Current opinion in neurobiology*, *55*, 160-166.

Redish A.D., & Gordon J.A. (2016). Computational psychiatry: New perspectives on mental illness. Cambridge, MA: MIT Press.

Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., McNally, R. J., van Nes, E. H., Scheffer, M., Kendler, K. S., & Others. (2019). Advancing the network theory of mental disorders: A computational model of panic disorder. https://psyarxiv.com/km37w/download

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biological psychiatry, 84(6), 443-451.

Rutledge, R. B., Chekroud, A. M., & Huys, Q. J. (2019). Machine learning and big data in psychiatry: toward clinical

applications. Current Opinion in Neurobiology, 55, 152–159. https://doi.org/10.1016/j.conb.2019.02.006

Salmi, J., Ritakallio, L., Fellman, D., Ellfolk, U., Rinne, J. O., & Laine, M. Disentangling the role of working memory in Parkinson's disease. Frontiers in Aging Neuroscience. 2020; 12: 1-17.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron, 79(2), 217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Simon, H. A. (1990). Invariants of human behavior. Annual Review of Psychology, 41, 1–19. https://doi.org/10.1146/annurev.ps.41.020190.000245

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. Psychological Bulletin, 139(1), 81–132. https://doi.org/10.1037/a0028727

Snyder, H. R., & Hankin, B. L. (2016). Spiraling out of control: Stress generation and subsequent rumination mediate the link between poorer cognitive control and internalizing psychopathology. Clinical Psychological Science, 4(6), 1047–1064. https://doi.org/10.1177/2167702616633157

Spitzer, R. L., First, M. B., & Wakefield, J. C. (2007). Saving PTSD from itself in DSM-V. Journal of Anxiety Disorders, 21(2), 233–241. https://doi.org/10.1016/j.janxdis.2006.09.006

Webb, C., Trivedi, M., Bruder, G., & Pizzagalli, D. A. (2016). Neural correlates of three promising endophenotypes of depression: evidence from the EMBARC study. https://www.nature.com/articles/npp2015165

Valton, V., Romaniuk, L., Douglas Steele, J., Lawrie, S., & Seriès, P. (2017). Comprehensive review: Computational modelling of schizophrenia. Neuroscience and Biobehavioral Reviews, 83, 631–646. https://doi.org/10.1016/j.neubiorev.2017.08.022

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., van Nes, E. H., Viechtbauer, W., Giltay, E. J., Aggen, S. H., Derom, C., Jacobs, N., Kendler, K. S., van der Maas, H. L. J., Neale, M. C., Peeters, F., Thiery, E., Zachar, P., & Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. Proceedings of the National Academy of Sciences of the United States of America, 111(1), 87–92. https://doi.org/10.1073/pnas.1312114110

Van Vugt, M. K., Hitchcock, P., & Shahar, B. (2012). The effects of mindfulness-based cognitive therapy on affective memory recall dynamics in depression: a mechanistic model of rumination. Frontiers in Human Neuroscience.

Wachbroit, R. (1994). Normality as a Biological Concept. Philosophy of Science, 61(4), 579–591. https://doi.org/10.1086/289823

Wakefield, J. C. (1992a). The concept of mental disorder. On the boundary between biological facts and social values. The American Psychologist, 47(3), 373–388. https://doi.org/10.1037//0003-066x.47.3.373

Wakefield, J. C. (1992b). Disorder as harmful dysfunction: a conceptual critique of DSM-III-R's definition of mental disorder. Psychological Review, 99(2), 232–247. https://doi.org/10.1037/0033-295x.99.2.232

Wang, X.-J., & Krystal, J. H. (2014). Computational psychiatry. Neuron, 84(3), 638–654. https://doi.org/10.1016/j.neuron.2014.10.018

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. Psychological Bulletin, 134(2), 163–206. https://doi.org/10.1037/0033-2909.134.2.163

Watts, A. L., Boness, C. L., Loeffelman, J. E., & Steinley, D. (2020). Does crude measurement contribute to observed unidimensionality of psychological constructs? An example with DSM-5 alcohol use disorder. Psyarxiv. 10.31234/osf.io/paxd4

Weiner, W. J. (2008). There Is No Parkinson Disease. Archives of Neurology, 65(6), 705–708. https://doi.org/10.1001/archneur.65.6.705

Whitmer, A. J., Frank, M. J., & Gotlib, I. H. (2012). Sensitivity to reward and punishment in major depressive disorder: effects of rumination and of single versus multiple experiences. Cognition & Emotion, 26(8), 1475–1485. https://doi.org/10.1080/02699931.2012.682973

Wiecki, T. V., Antoniades, C. A., Stevenson, A., Kennard, C., Borowsky, B., Owen, G., Leavitt, B., Roos, R., Durr, A.,

Tabrizi, S. J., & Frank, M. J. (2016). A Computational Cognitive Biomarker for Early-Stage Huntington's Disease. PloS One, 11(2), e0148409. https://doi.org/10.1371/journal.pone.0148409

Wiecki, T. V., & Frank, M. J. (2010). Neurocomputational models of motor and cognitive deficits in Parkinson's disease. Progress in Brain Research, 183, 275–297. https://doi.org/10.1016/S0079-6123(10)83014-6

Wiecki, T. V., Poland, J., & Frank, M. J. (2015). Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry: Clustering and Classification. Clinical Psychological Science, 3(3), 378–399. https://doi.org/10.1177/2167702614565359

Williams, L. M. (2016). Precision psychiatry: a neural circuit taxonomy for depression and anxiety. The Lancet. Psychiatry, 3(5), 472–480. https://doi.org/10.1016/S2215-0366(15)00579-9

Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. eLife, 8. https://doi.org/10.7554/eLife.49547

Wright, A. G. C., & Woods, W. C. (2020). Personalized Models of Psychopathology. Annual Review of Clinical Psychology, 16, 49–74. https://doi.org/10.1146/annurev-clinpsy-102419-125032

Zorowitz, S., Momennejad, I., & Daw, N. D. (2020). Anxiety, avoidance, and sequential evaluation. Computational Psychiatry, 4, 1-17.