

Identifying Transdiagnostic Mechanisms in Mental Health Using Computational Factor Modeling

Toby Wise, Oliver J. Robinson, and Claire M. Gillan

ABSTRACT

Most psychiatric disorders do not occur in isolation, and most psychiatric symptom dimensions are not uniquely expressed within a single diagnostic category. Current treatments fail to work for around 25% to 40% of individuals, perhaps due at least in part to an overreliance on diagnostic categories in treatment development and allocation. In this review, we describe ongoing efforts in the field to surmount these challenges and precisely characterize psychiatric symptom dimensions using large-scale studies of unselected samples via remote, online, and “citizen science” efforts that take a dimensional, mechanistic approach. We discuss the importance that efforts to identify meaningful psychiatric dimensions be coupled with careful computational modeling to formally specify, test, and potentially falsify candidate mechanisms that underlie transdiagnostic symptom dimensions. We refer to this approach, i.e., where symptom dimensions are identified and validated against computationally well-defined neurocognitive processes, as computational factor modeling. We describe in detail some recent applications of this method to understand transdiagnostic cognitive processes that include model-based planning, metacognition, appetitive processing, and uncertainty estimation. In this context, we highlight how computational factor modeling has been used to identify specific associations between cognition and symptom dimensions and reveal previously obscured relationships, how findings generalize to smaller in-person clinical and nonclinical samples, and how the method is being adapted and optimized beyond its original instantiation. Crucially, we discuss next steps for this area of research, highlighting the value of more direct investigations of treatment response that bridge the gap between basic research and the clinic.

<https://doi.org/10.1016/j.biopsych.2022.09.034>

A shift away from a categorical view on mental health is well underway across psychiatry research (1–3). This is in response to well-documented issues with diagnostic frameworks in terms of comorbidity (4), reliability (5), heterogeneity (6), and binarization of a continuous mental health space (7,8). Numerous promising alternatives to the existing diagnostic rubric are in development, such as Research Domain Criteria and Hierarchical Taxonomy of Psychopathology (9,10). Although advances have been made within these frameworks, they continue to depend on traditional research formulas in psychiatry; that is, they focus on small, diagnosed patient samples or the interrogation of cognitive mechanisms after symptom-level phenomena have been defined, rather than defining them both in concert. Here we introduce a novel combination of interdisciplinary methods called computational factor modeling (CFM) (Figure 1 and Box 1) which we believe can accelerate transdiagnostic research in psychiatry. In CFM, candidate transdiagnostic symptom dimensions are identified not in patients, but rather in unselected online samples of individuals who experience a range of psychopathology and can be gathered at the scale required to support robust exploration and replication approaches. Transdiagnostic symptom dimensions in CFM are defined using a combination of a data-driven dimensionality reduction of self-report questionnaire

responses and theory-driven computational modeling of behavior that allows us to precisely characterize the cognitive processes that characterize a given dimension (11). In this paper, we discuss the genesis of CFM and describe a range of recent applications. We highlight the importance of computational modeling as a central part of this endeavor, moving from descriptive summaries of behavior with multiple potential mechanistic accounts toward detailed, falsifiable, and precise theories. We show how CFM, although still in the early stages, has already augmented our understanding of mental illness, yielding putative mechanisms underlying several transdiagnostic symptom dimensions. Finally, we discuss how CFM can support new frameworks like Research Domain Criteria and Hierarchical Taxonomy of Psychopathology and may drive innovations in treatment development and allocation (12,13).

MODEL-BASED PLANNING

A number of case-control studies have observed altered goal-directed (model-based) (Box 1) behavior in obsessive-compulsive disorder (OCD), which leaves patients vulnerable to rigid habitual behaviors (Figure 1A, B) (16–18). These findings were subsequently extended to addiction and binge-eating disorder (19–22), leading researchers to posit that impaired goal-directed control over habits was a neurocomputational

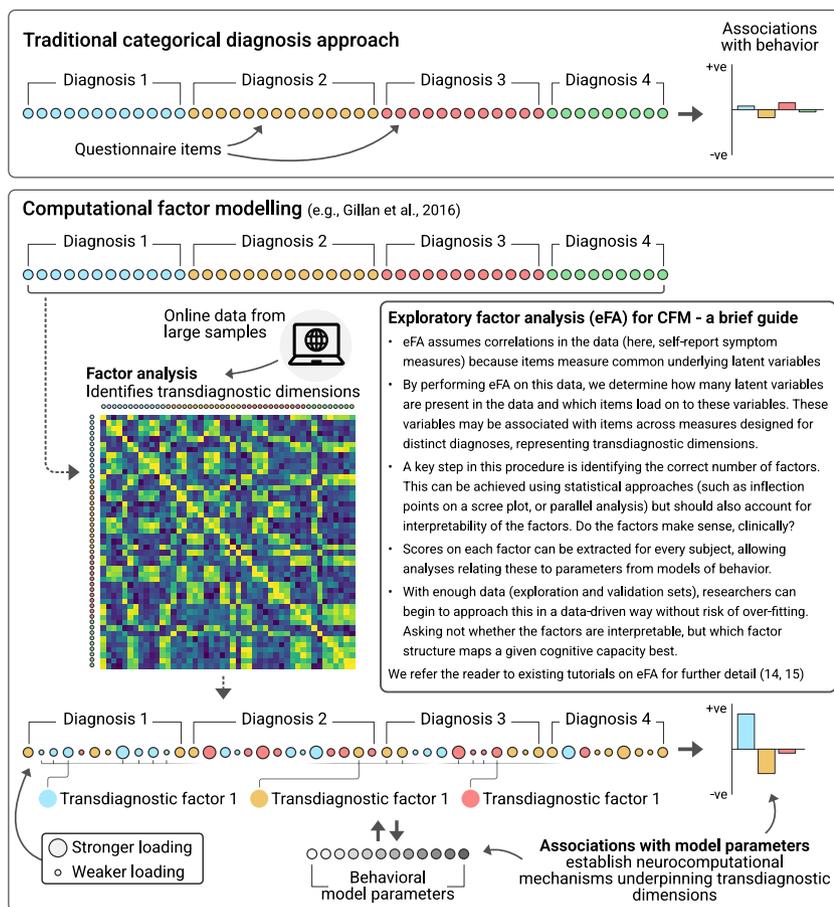


Figure 1. Computational factor modeling (CFM). CFM aims to identify transdiagnostic symptom dimensions that are associated with precise neurocomputational mechanisms. The method looks under the hood of cognitive processes using computational modeling and links their component parts to the symptoms that individuals experience transdiagnostically. Unsupervised dimensionality reduction like exploratory factor analysis or principal component analysis (14,15) are used to identify crosscutting, data-driven latent symptom dimensions (e.g., compulsivity and intrusive thought) in large unselected samples, typically using data gathered online. Computational models are then fit to participants' behavior to identify theory-driven latent behavioral dimensions (e.g., learning rate). The relationship between these 2 sets of latent factors is then examined and can be iteratively and bidirectionally refined.

feature of compulsivity more generally. But a problem for this theory soon followed; other conditions with less characteristic compulsive features—social anxiety (23,24), autism (24,25), schizophrenia (26,27), and Tourette syndrome (28)—also showed deficits relative to controls. This suggested 2 possibilities: either alterations in goal-directed control are a general feature of psychopathology, or nonspecific links between model-based planning and clinical phenotypes arise from problems with the validity/dissociability of diagnoses. One of the challenges in resolving this debate is that in order to test whether specific transdiagnostic mechanisms exist, we need to measure multiple aspects of psychopathology in the same individuals at-scale.

To resolve this, Gillan *et al.* (29) eschewed the traditional case-control framework and recruited members of the public. Over 1400 individuals completed an online assessment that included self-report clinical assessments and a behavioral task that allowed researchers to use computational modeling to parse model-based planning from more reflexive learning styles (model-free learning). They found that the clinical correlates of model-based planning were indeed broader than the symptoms of a single disorder (associated with eating disorder, impulsivity, OCD, and addiction symptoms) but also showed some specificity (e.g., with schizotypy, depression,

apathy, and trait and social anxiety) (Figure 2C). A factor analytic approach was used to identify a transdiagnostic symptom dimension that could explain this pattern. This identified one dimension, compulsivity and intrusive thought (CIT), that cut through existing diagnostic rules and explained the blurring of model-based deficits across diagnoses. This association was specific; the dimensions of anxious-depression (AD) and social withdrawal (SW) were unrelated to these deficits (Figure 2D). This finding was replicated online (30), in-person (31), and, critically, in patients with diagnoses (32), wherein it was found that model-based planning deficits do not distinguish between diagnostic labels very well; rather, they track individual differences in compulsivity irrespective of diagnosis (Figure 2E). This finding underscores the value of CFM. Diagnostic groups are heterogeneous and overlapping, and without large samples, we cannot unpack the clinical complexity and robustly identify the specific symptom dimensions that are driving effects that otherwise appear common across psychiatry. We posit that in this respect, CFM is an important new complement to patient studies, allowing us to identify specific and precise underlying mechanisms of transdiagnostic symptoms that play a role in multiple disorders but are experienced to different degrees by individuals.

Box 1. Glossary of Computational Mechanisms Commonly Identified From Cognitive Task Behavior

- **Model-Based Planning.** Model-based planning and goal-directed learning are often used synonymously. They refer to the use of cognitive maps or models of the world to guide behavior in a prospective fashion (133–135) (Figure 2). Rather than relying on direct experience of reward, our model-based faculties allow us to simulate future states (136,137), integrate information from various sources (e.g. experience, observation, interoception), and rapidly update our action plans without requiring direct experience of the outcome of a new action. Failures in model-based decision-making lead people to rely on more automatic behaviors called habits (138) that appear rigid and outside intentional control. The first empirical studies testing these ideas trained patients with obsessive-compulsive disorder to perform responses to stimuli to gain rewards and then subsequently reduced the value of those rewards (an outcome devaluation procedure) and tested whether behavior ceased. In a range of experimental preparations, patients with obsessive-compulsive disorder were found to persist in responding, indicating reduced goal-directed control over behavior (16–18). Later, a more sophisticated two-step task used reinforcement learning to characterize the computational mechanism of these goal-directed lapses, coining the term model-based planning. Model-based planning in this task refers to the extent to which individuals use a high-level understanding of task structure (models) to learn not just from experience, but to update the value of actions not taken and prevent incorrect assignment of value to actions taken. Model-based planning is linked to the ventromedial prefrontal cortex activity (133) and requires the hippocampus (139), highlighting these as potential targets for investigation with regard to compulsivity.
- **Metacognition.** Recent years have seen a proliferation of novel tasks and analysis approaches (140), enabling more precise estimates of metacognition that go beyond self-report and control tightly for potential confounds (Figure 3). Tasks that measure metacognition typically have participants complete a perceptual decision-making task, such as estimating which side of a screen has more dots displayed. Staircase procedures are employed so that task difficulty adapts to each person, and they can be held at consistent levels of performance (e.g. 70% correct), thereby removing type 1 performance confounds (real differences in accuracy). These kinds of tests allow researchers to derive 2 components using signal detection theory models, metacognitive bias (i.e. over- or underconfidence in your own performance), and metacognitive sensitivity (i.e., how well confidence discriminates correct vs. incorrect responses) (140). Metacognition involves the lateral prefrontal cortex and dorsal anterior cingulate cortex (141,142), suggesting that these areas may be relevant for anxious-depression and compulsivity (36).
- **Reward Processing.** Reward processing is often studied in the context of reinforcement learning tasks, where computational models provide a framework for understanding how people update their expectations about future events based on new evidence. A core concept in reinforcement learning models is prediction error (143), which is defined as the difference between what we expect to happen and what actually happens. Animals use prediction errors to update new expectations via a learning rate, which is a parameter that governs how much we update our existing expectations based on new information (Figure 4). A related concept is reward sensitivity, which is defined as the consummatory pleasure one obtains from a reward. Recent work suggests that a more sensitive (or potentially distinct) measure of this can be gleaned from studying how values that we learn to associate with cues spread or “generalize” to other similar cues (50,144) and change the way that evidence is accumulated [as modeled by drift diffusion modeling of reaction times (145)]. The affective bias task commonly used to study affective bias and drift rate in depression (53) was adapted from a task used in rodents, providing additional potential for neurobiological investigations; specifically, administration of a GABA_A inverse agonist induces a negative bias and lower drift rate (50), suggesting that GABA may play a role in this symptom dimension.
- **Uncertainty.** Gambling tasks are commonly used to assess decision making under uncertainty, where subjects must choose between certain (e.g., 50 points guaranteed) and risky options (e.g., 50/50 chance of winning nothing or 100 points) or ambiguous options where information is obscured (unknown probability of winning 0 or 100 points). Performance on these tasks can be modeled using Prospect Theory models (146) to isolate behavioral tendencies including risk aversion (avoiding uncertain outcomes), ambiguity aversion (avoiding unknown outcome probabilities or magnitudes), reward maximization (choosing higher expected values), and loss aversion (overweighting potential losses relative to gains). Other kinds of tasks have been used to look at how people learn under conditions of uncertainty. Browning *et al.* (65) examined this using a task in which participants learn to choose between 2 options with different probabilities and magnitudes of punishment (Figure 5). These decisions take place in two states, one in which the correct choice is stable and another in which it is volatile and the correct choice switches frequently. To avoid punishment in this task, learning rates should increase in volatile states so that recent outcomes are prioritized over old outcomes. Individuals high in trait anxiety failed to update their learning rate accordingly, suggesting an impairment in uncertainty processing. Adaptation of learning in response to volatility is linked to noradrenaline (147), suggesting that this neuromodulator could play a role in internalizing symptoms.

METACOGNITION

Another area where the CFM approach has had an impact is the study of metacognition, the ability to accurately reflect on one's own thoughts, feelings, and behaviors. Metacognition plays a vital role in adaptive decision making and can be modeled using signal detection theory (Box 1; Figure 3A) (33). Alterations in metacognition have been observed in depression, with patients tending to think they perform worse than other people despite comparable performance (34). Case-control studies suggest that this effect is nonspecific and have found it in anxiety, OCD, and schizophrenia as well (35) (Figure 3B). Given the high rates of comorbidity across these conditions, it is possible, though, that a symptom common to all these conditions is responsible. To test this, Rouault *et al.*

(36) used CFM in 2 large, online, unselected samples (Figure 3C) to examine the same transdiagnostic factors from the first CFM study (29) alongside a perceptual decision-making task. The latent clinical dimensions of CIT and AD were highly consistent across the studies, with correlation of loadings of $r = 0.87$ to 0.97 . Strikingly, CIT was linked to positive metacognitive bias (i.e., overconfidence), while AD was associated with negative metacognitive bias (i.e., underconfidence).

These bidirectional associations were replicated and extended in another study in an online, unselected sample using a reinforcement learning task (37) (Figure 3D), which allowed a deeper analysis of how evidence informs confidence assessments. They found that confidence in

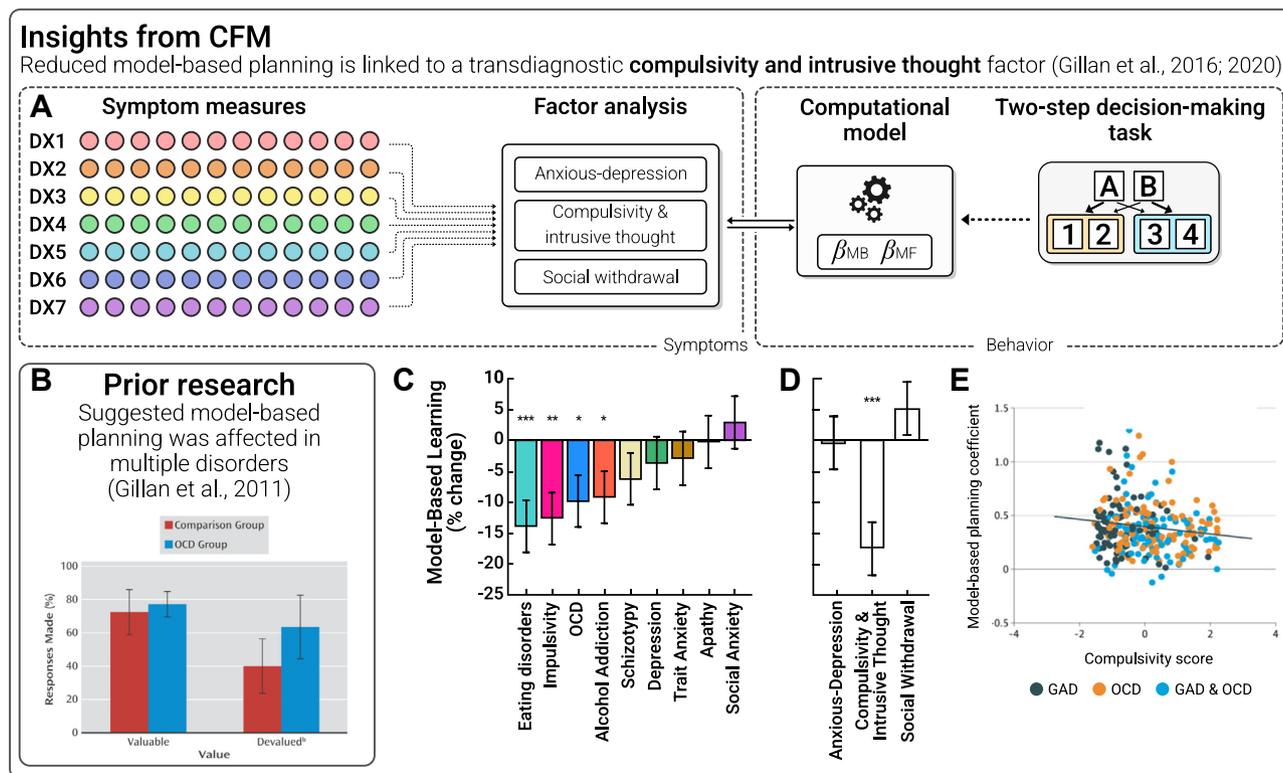


Figure 2. Model-based planning. **(A)** Computational factor modeling (CFM) has been used to identify a transdiagnostic psychiatric dimension related to deficits in model-based planning (see [Box 1](#) for a detailed definition). Individual items (circles) from a range of questionnaires relating to traditional diagnoses (DX, colors) were subjected to factor analysis. Three dimensions resulted: anxious-depression, compulsivity and intrusive thought, and social withdrawal. Behavioral data on a 2-step decision-making task were fit using a computational model that extracted individual estimates of model-based planning that the model can separate from a range of alternatives such as choice perseveration, randomness, or model-free learning. The authors tested for associations between computational parameters and transdiagnostic dimensions (controlling for age, gender, and IQ). **(B)** Prior work suggested that the balance between goal-directed behavior and habit is linked to obsessive-compulsive disorder (OCD), but it was unclear what specific aspect of psychopathology drove this effect and what precise mechanism explained this imbalance (16). **(C)** Mirroring smaller patient studies, in a large unselected sample of $N = 1413$, the symptoms of many conditions correlated with model-based planning deficits (29). **(D)** CFM revealed that this apparent blurring of model-based planning deficits across questionnaires was explained by the compulsivity and intrusive thought dimension. **(E)** These results were replicated in diagnosed patients and, moreover, effects were stronger when measuring individual differences in compulsivity compared to diagnostic status (OCD or no OCD) (32). * $p < .05$; ** $p < .01$; *** $p < .001$. GAD, generalized anxiety disorder.

individuals high in CIT is high overall, but less informed by evidence from the external world (e.g., hits and misses in the task). In contrast, as levels of AD increased in the sample, there was no decrease in the use of evidence to inform confidence judgments (37). Seow *et al.* (38) suggested that this is evidence of dissociable mechanisms underlying the confidence biases in compulsivity and AD; reduced confidence in depression may stem from a lower setpoint of confidence relating to global self-beliefs [e.g., low self-esteem (39)], while overconfidence in CIT may be caused by more specific learning difficulties, for example, problems in building an accurate mental model of one's performance based on experience (37). Hoven *et al.* (40) recently tested this directly using CFM by studying the association between AD and CIT and various levels of confidence along a hierarchy in 489 individuals from the general population. They found that the association between local confidence and AD was explained by reduced confidence in their general abilities (i.e., self-beliefs). Importantly, this was not the case for

CIT; in fact, there was a marked decoupling of local and global confidence as CIT severity increased. This suggests that the bidirectional associations with metacognition in AD and CIT may have their origin at different levels of the self-confidence hierarchy. More broadly, it underscores the advantage of the transdiagnostic factor approach in disentangling specific disease mechanisms that may be impossible to study using case-control frameworks.

REWARD PROCESSING

Altered processing of reward is conceptualized as the clinical symptom of anhedonia and features prominently in depression, but also in schizophrenia and other disorders. One of the earliest papers linking anhedonia to components of reward processing (41) demonstrated reduced reward learning rates ([Box 1](#); [Figure 4A](#)) with increasing anhedonia across healthy and depressed individuals (irrespective of diagnosis). Using functional magnetic resonance imaging, reduced neural

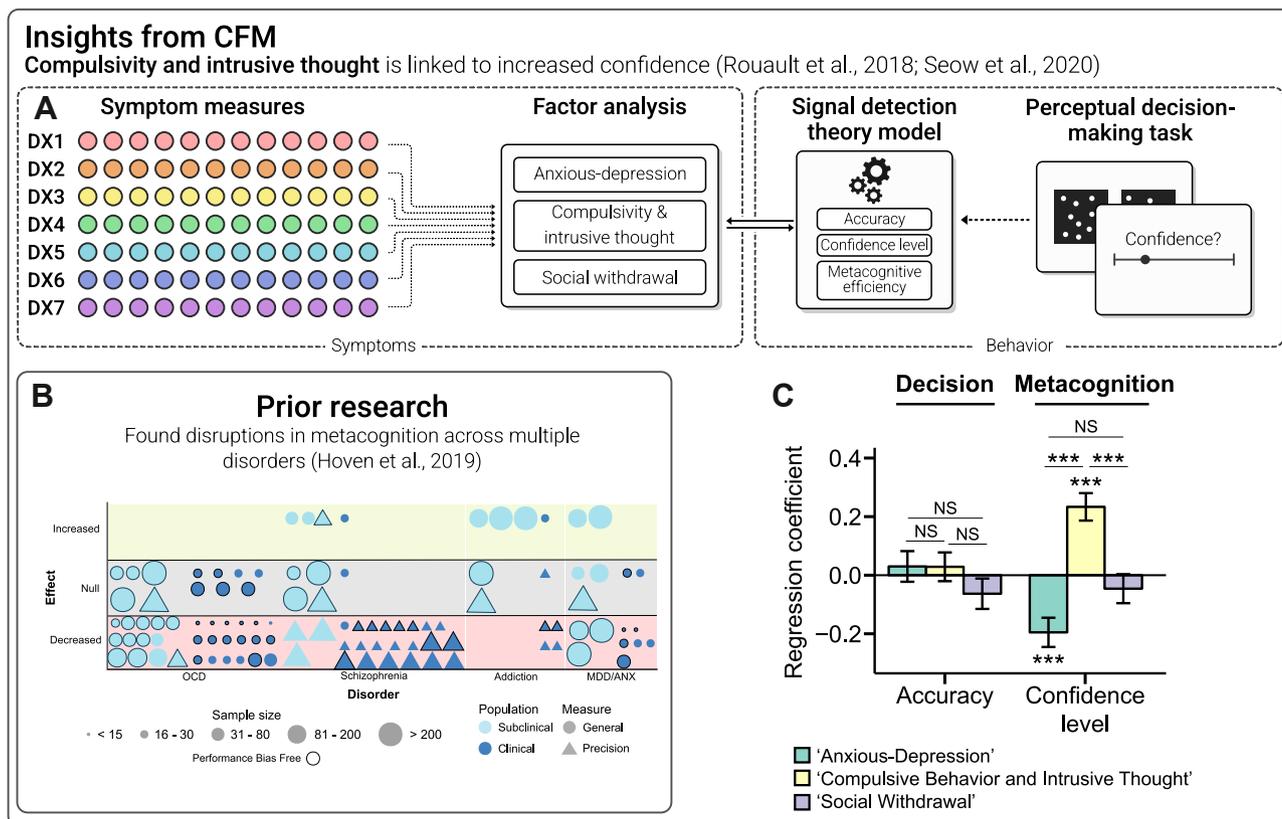


Figure 3. Metacognition. **(A)** Computational factor modeling (CFM) applied to the study of metacognition (see Box 1 for detailed definition). The same set of questionnaires used in (29) were subjected to a factor analysis that yielded the same structure and highly correlated loadings as the original paper (all $r_s > 0.87$). This time, transdiagnostic factors were related to metacognitive bias, a person's tendency to over- or underestimate their own performance on a perceptual decision-making task (where objective performance differences are tightly controlled). **(B)** A great deal of prior work has been carried out in this area in both clinical and nonclinical samples. As for model-based planning, patterns of association blur across diagnostic lines, showing fairly consistent reductions in metacognitive bias (that is, confidence) (35). **(C)** Using CFM, Rouault et al. (36) showed that in fact a bidirectional association exists wherein anxious-depression is linked to decreased confidence in performance, while compulsivity and intrusive thought is characterized by increased confidence. This illustrates how traditional methods using heterogeneous disorder categories may average out specific and transdiagnostic processes. *** $p < .001$. ANX, anxiety; MDD, major depressive disorder; NS, nonsignificant; OCD, obsessive-compulsive disorder.

signatures of reward prediction errors were also seen in both depression (42,43) and schizophrenia (42). Another paper with careful computational modeling of behavioral data in 69 patients with major depressive disorder (MDD) showed that anhedonia was linked to both learning rate and outcome sensitivity biases (44). However, a functional magnetic resonance imaging study of 148 patients with MDD and 31 controls (45) found no case-control differences in reward prediction errors, and other recent, larger-scale work has also yielded mixed results. A mega-analysis of a single task (46) suggested that anhedonia was associated with reduced reward sensitivity in clinical samples (i.e., consummatory pleasure) but not with reward learning per se (47). More dramatically, a study with both functional magnetic resonance imaging from a small case-control sample and behavioral data from a general population sample of more than 1800 users of a smartphone app found no neural or behavioral deficits in reward processing in the case-control sample but did find a relationship with depression symptoms in the unselected sample that was the

opposite of what was expected (i.e., increased consummatory reward response) (48) (Figure 4B). This inconsistency across studies may be due to the very high comorbidity between MDD and anxiety disorders in case-control studies. Indeed, a recent CFM study showed no association between reward learning deficits and a single AD factor in a healthy sample (49), but, perhaps critically, did not isolate depression from anxiety. In a different type of task translated from animal work, where biases in reward learning are examined by testing whether learned reward values generalize to ambiguous cues (50), individuals with mood and anxiety disorders were more likely to pessimistically assume that neutral cues would lead to low (rather than high) rewards, driven potentially by lower evidence accumulation for high rewards (Box 1) (51,52). Critically, CFM in 990 participants from the general population showed that performance correlated with depression but not anxiety (psychosis or compulsivity) (53), suggesting a need to tease apart depression and anxiety symptomatology in reward-processing studies (Figure 4C, D).

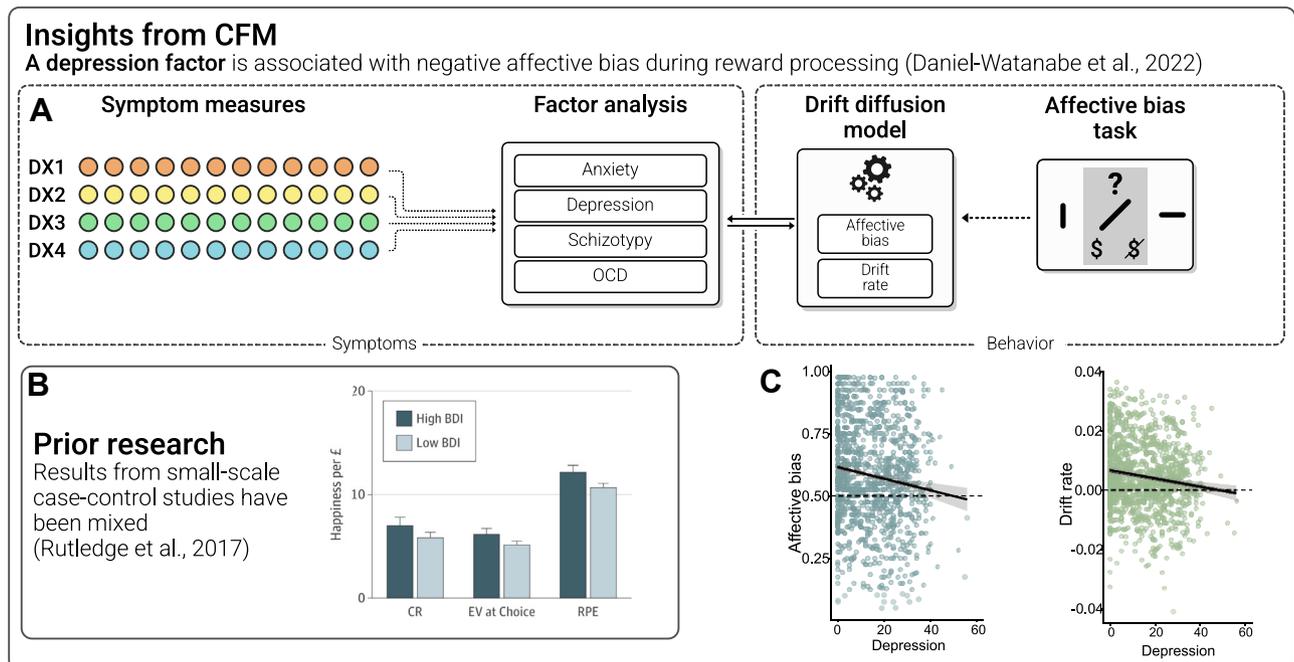


Figure 4. Reward processing. **(A)** Newer computational factor modeling (CFM) studies have used different sets of questionnaire items to derive new transdiagnostic factors. One study (52) took this approach recently to study reward processing biases in the commonly co-occurring clinical symptoms of depression and anxiety. Factor analyses of a set of 4 questionnaires recapitulated a similar structure to that of the original questionnaires, demonstrating that anxiety and depression do not always occur together. **(B)** Reward processing has been studied in great detail in psychiatry using small-scale case-control designs focusing on depression, but the results have been mixed, with prominent failures to replicate in large samples (48). This may be due to comorbidities between depression and anxiety making it challenging to isolate the specific symptoms that are linked to reward biases. **(C)** A recent CFM study (53) used a large unselected sample to show how negative reward-related affective biases and drift rate (the rate at which evidence is accumulated to make a decision) are linked specifically to a factor representing depressive symptoms but not anxiety symptoms. BDI, Beck Depression Inventory; CR, certain rewards; EV, expected value; OCD, obsessive-compulsive disorder; RPE, reward prediction error.

UNCERTAINTY

Changes in uncertainty processing are thought to play a major role in anxiety disorders (54), in which individuals report feeling more uncertain (55), report uncertainty as more aversive (54), and show elevated psychophysiological (e.g., startle) and neural responses during uncertain threats (56,57). A growing literature suggests that this intolerance of uncertainty (58) may represent a transdiagnostic construct. However, much of the early research relied on self-report assessments and used tasks that had difficulty isolating the components of uncertainty. In recent years, computational approaches have been adopted that can distinguish risk, loss, and ambiguity sensitivity (59) (Box 1). Using these methods, studies have shown that risk aversion is elevated in anxiety disorders (60) and that individual differences in trait anxiety correlate with ambiguity aversion (61). A key question that CFM has helped resolve is whether uncertainty-related processing is linked specifically to anxiety or to a more general negative affect dimension. One study investigated ambiguity aversion using CFM in an unselected sample and revealed that a transdiagnostic anxiety factor was specifically associated with enhanced generalization of aversive value, a mechanism through which ambiguity is reduced (62). However, another study in a large, online, unselected sample found no link between trait anxiety or depression and risk or ambiguity aversion (63). One possibility is that increases in risk and ambiguity aversion may be

a state rather than a trait marker of anxiety that emerges in individuals exhibiting acute symptoms. In line with this account, 1 study in a large, unselected sample found that heightened ambiguity aversion was linked to COVID-19-induced anxiety (64).

Another significant research area concerns uncertainty induced by environmental volatility (Figure 5A). This was investigated by Browning *et al.* (65), who found that healthy individuals who were high in trait anxiety failed to update their learning rate in response to changes in environmental volatility, suggesting an impairment in uncertainty processing (Figure 5B). In a larger follow-up study consisting of clinically diagnosed patients with MDD and generalized anxiety disorder and another unselected sample recruited from a crowdsourcing platform (66), the authors used a bifactor model approach to CFM to determine that the failure to adjust learning rates was best captured by a general factor representing combined anxiety and depressive features rather than anxiety or depression specifically (Figures 5C–E). These tasks assess how people respond to objective uncertainty, but recent work has shown that computational modeling can also be used to infer and quantify individual-level subjective uncertainty (67). Wise and Dolan (68) demonstrated that a factor including cognitive anxiety, depression, and intolerance of uncertainty was linked to heightened subjective uncertainty during a highly gamified aversive learning task in an unselected sample. This paper incorporated a combination of behavioral and self-

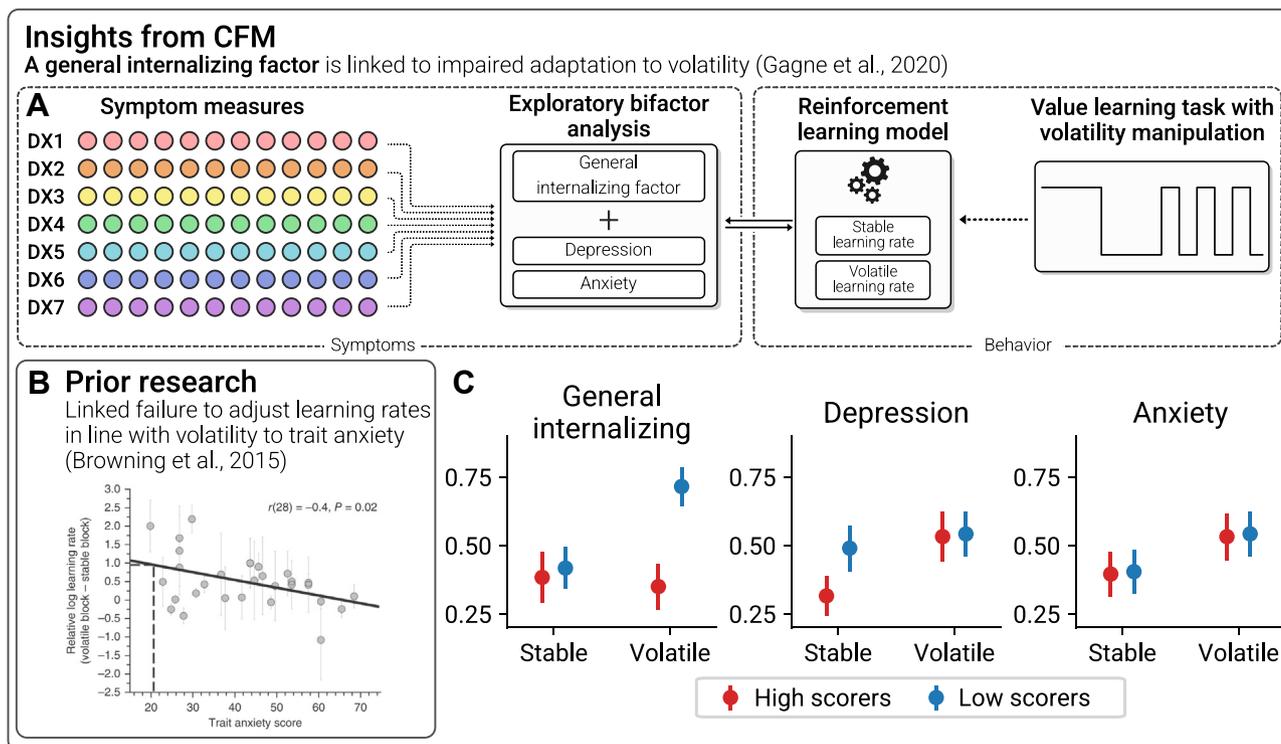


Figure 5. Uncertainty. **(A)** Computational factor modeling (CFM) studies have recently begun to adopt other approaches to dimensionality reduction. Gagne *et al.* (66) reduced a range of questionnaires into a general internalizing factor as well as 2 specific factors relating to depression and anxiety. They tested for associations with parameters from a computational model estimating how people adapt their learning rates (i.e., how quickly they learn from new evidence) in response to changes in environmental volatility. **(B)** Prior research found that individuals high in trait anxiety fail to adapt their learning rate (65). **(C)** Bifactor modeling using the CFM approach revealed that this failure to adapt learning rate was linked to the general internalizing factor rather than being specific to what distinguishes depression and anxiety from each another (66). DX, diagnosis.

report data as part of the identification of transdiagnostic factors, representing an intriguing progression of the CFM approach that may assist in the more data-driven identification of dimensions of psychopathology in the future.

IMPLICATIONS FOR TREATMENT

The framework we have outlined, focusing on transdiagnostic symptom dimensions with associated neurocomputational mechanisms, has significant potential for improving outcomes. This may occur through several pathways, which we describe in detail in the next section with reference to recent examples that have begun to realize this promise. In brief, we propose that mechanistic insights from CFM can help us understand whether and how existing treatments work, for example by changing key neurocomputational processes such as model-based planning or metacognition. This might inform the development of new treatments that target these processes more effectively or selectively. Additionally, CFM might be the key ingredient to help us deliver treatments more precisely, based on an individual's specific transdiagnostic and mechanistically defined profile rather than their diagnosis. This work is still in its infancy, and an important task for research in the coming years will be to focus on these real-world applications. In the following sections, we review the progress that has

already been made in this area and outline suggestions for future work (Figure 6).

Model-Based Planning

Key questions that emerge from the link between model-based planning and compulsivity are whether model-based planning can be changed using available or novel therapeutics and whether they could signal which treatment will work best for whom. The answer to the former question appears to be no; model-based planning does not improve following targeted training on tasks of this kind (69), nor does it improve following cognitive behavioral therapy (CBT) for OCD, even in individuals who respond extremely well to treatment (70). If model-based abilities cannot be easily changed, are there alternative ways that this mechanistic understanding of compulsivity might improve treatment? One study tested this by engaging model-based systems using a habit-override task during the administration of continuous theta burst stimulation (71). The focus of this stimulation was to reduce left orbitofrontal cortex activation, building on prior knowledge of the role that the orbitofrontal cortex plays in both habit and compulsive behaviors (72,73). This treatment acutely decreased compulsive behavior in individuals with compulsive disorders, with these beneficial effects persisting for 1 week (Figure 6A). As with CBT, however, the treatment had little effect on model-based planning

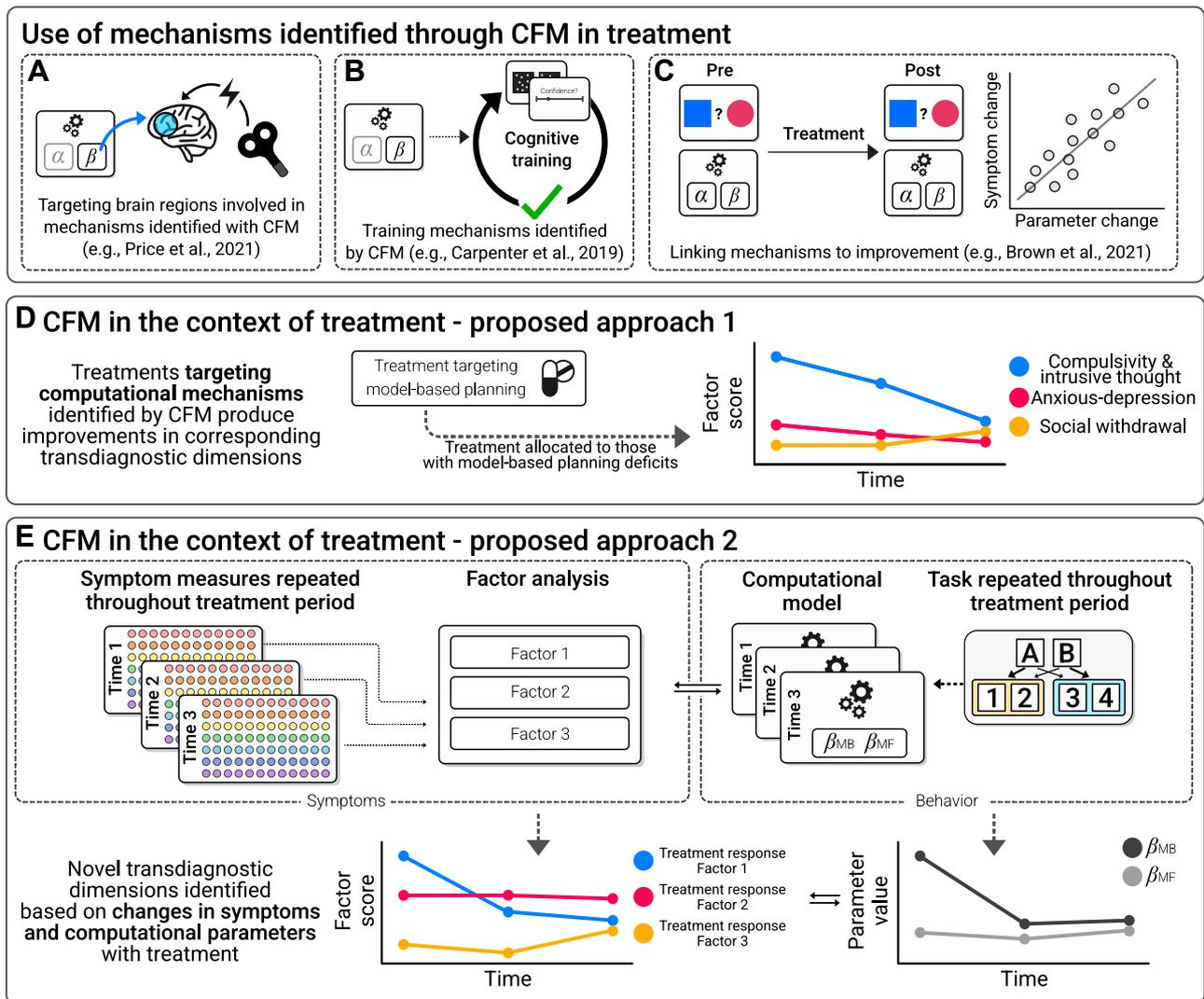


Figure 6. A focus on treatment. An important next step for computational factor modeling (CFM) research is to integrate it more directly with treatment. There are a number of promising examples, including using CFM-defined mechanisms to (A) optimize existing interventions, (B) develop entirely novel interventions that target specific computational processes, and (C) understand how existing treatments work. However, to date, there has been limited direct application of the full CFM approach (i.e., including both latent symptom dimensions and computational models of behavior) within treatment studies. There are at least 2 ways that this could be of value. First, (D) by directly testing the impact of treatment on previously identified transdiagnostic dimensions or second, (E) by using a data-driven approach on repeated measures data to identify latent dimensions that specifically predict treatment response. MB, model-based; MF, model-free.

itself (74). Further work is needed to determine whether activation of habit circuits is necessary for patients to achieve this benefit from continuous theta burst stimulation. If it does, this may provide a further basis for exploring innovative psychological therapies, as well as stimulation techniques, that can increase model-based planning (75).

Metacognition

Recent work suggests that metacognition, unlike model-based planning, may be a trainable cognitive capacity and/or a target for treatment (Figure 6B). In clinical settings, metacognitive therapy has been used to treat depression and includes components such as attention training and detached

mindfulness as ways to alter how people respond to negative thoughts (76). Recently, researchers have attempted to study analogues of these treatments in lab settings using tightly controlled tasks, bridging real-world interventions to the way metacognition is defined in the field of computational psychiatry. One such study randomized healthy individuals to receive training on their metacognitive assessments (77) and found that metacognitive performance improved and generalized to new tasks. A second study also demonstrated improvements following metacognitive training in healthy individuals but found that this did not have more general impacts on real-world behaviors like cognitive offloading (78). This translation to real world function outside the confines of contrived laboratory settings is crucial and a challenge that many cognitive training interventions

Computational Factor Modeling

face. An important next question for this area is whether meta-cognitive training can be delivered in a more personalized manner based on what CFM has taught us about the dissociable correlates of metacognition, AD, and compulsivity.

Reward

Attenuated reward processing is thought to be partly responsible for a host of cognitive phenomena observed in depression, for example, reduced emotional recognition of happy faces (79), attentional biases toward negative information (80–82), and biases for negative memories (83). These biases have high face validity for the negative schemata that are central to cognitive models of depression and are key targets for CBT (84). Both therapy (85) and selective serotonin reuptake inhibitors (SSRIs) (86) have been shown to increase striatal response to reward, and increased computationally modeled pretreatment reward responses are associated with a greater symptom improvement (87). Several recent computational modeling studies in clinical samples have made notable strides in this area. One study showed that the reduced reward learning rates associated with anhedonia normalize following CBT in MDD (44) (Figure 6C). Another study found that relapse following discontinuation of SSRIs was predicted by reduced baseline effort expenditure to gain rewards (88). A third study trained an algorithm to predict treatment response based on a combination of symptom and negative bias changes 1 week after starting antidepressants (89). Although this algorithm performed above chance in the discovery study, it failed to improve pre-registered outcomes in a subsequent clinical trial (90). Although the lack of generalization is discouraging, this methodology is exemplary in many ways and has great potential if employed with appropriately powered samples.

Uncertainty Processing

The apparent state-dependence of uncertainty-guided decision-making strategies in anxiety (64) raises the possibility that these may represent causal or maintaining factors that could be targeted through intervention. Indeed, asking healthy subjects to adopt different cognitive strategies to regulate emotional responses has been shown to influence risk aversion (91). A placebo-controlled study of the antihypertensive drug losartan (92) found no evidence that it improved learning rate adaptation to uncertainty in healthy individuals (instead finding that it reduced punishment learning). This suggests that adaptation may be relatively difficult to change, but this awaits confirmation using a more conventional anxiolytic intervention. In contrast, recent work has shown that elevated startle responses to unpredictable threats (another behavioral assay of uncertainty processing) decrease after CBT but not SSRI treatment (93). This dissociation is important because it may suggest differential mechanisms of action of these treatments, which could aid in precision allocation. However, another study demonstrated that SSRIs did reduce startle responses to unpredictable shock, this time in healthy volunteers (94). This indicates that the ways in which people respond to uncertainty are malleable, but more work is needed to test how and for whom. This is an important target for future work using CFM in large samples that can be used to reliably estimate whether

uncertainty processing can be addressed clinically and whether there is scope for stratification based on individual differences.

DISCUSSION

Bringing It Back to Neuroscience

Online methods have been crucial for CFM studies to achieve large samples, but it is not envisioned that research should remain exclusively in the online space. Brain imaging, physiology, pharmacology, and animal studies are necessary to elaborate on underlying mechanisms. There are numerous examples of overlapping neurobiological changes across psychiatric conditions, for example reduced medial prefrontal cortex volume (95) or altered default mode network function (96). One possibility is that these reflect neurobiological substrates of a transdiagnostic mechanism that CFM can help illuminate. One study has already taken the approach of elaborating mechanistically on insights from CFM in a smaller in-person sample. Seow *et al.* (31) examined the electrophysiological correlates of model-based planning in approximately 200 students who varied in their levels of CIT. The authors bridged directly from earlier work by applying the exact factor weights derived from an unselected online sample to the in-person student sample. They found that deficits in model-based planning linked to this symptom dimension were associated with diminished neural representations of task structure. This converges with recent findings from general population samples suggesting that failures in goal-directed control in compulsivity are driven by problems with building and maintaining accurate and high-level maps of the world (97,98). As more studies adopt CFM methods in large online samples, this back-translation will be crucial to test many of the causal predictions made by the models.

A Focus on Treatment, From the Start

Research aiming to correlate symptoms with neurocomputational mechanisms can only take us so far. Treatment-oriented work is an essential next step, and we argue that it should be included earlier in the discovery process and integrated with CFM approaches. Two potential extensions that ask slightly different questions are 1) identifying factors using CFM as reviewed earlier and then assessing whether they are impacted by treatment (Figure 6D) or 2) using the CFM approach on the treatment-related change in performance to identify transdiagnostic predictors of treatment response (Figure 6E). One of the key challenges with this work is achieving the sample sizes that are necessary to develop and validate neurocomputational markers of treatment response. Similar issues have been faced by chronically underpowered machine learning research in the area of treatment response (99). Online methods can help here, too. Lee *et al.* (100) partnered with a digital CBT provider to recruit, assess (using CFM), and follow hundreds of patients through treatment in a short space of time. This illustrates how collaboration between the digital health industry and academia could radically transform research in this area. Another interesting approach (similar to Figure 6D) is to develop tightly constrained lab models (analogues) of psychological therapy and study them in large, unselected samples to understand

how they affect CFM dimensions. Dercon *et al.* (101) took this approach in a large online sample of healthy individuals and found that a cognitive distancing intervention increased participants' learning from negative events and integration of previous choice values. These 2 examples illustrate how CFM approaches can be integrated more directly into the study of how treatments work on well-defined computational processes and how internet-based methods allow researchers to do this at scale. We must acknowledge, however, that clinical impact is still speculative; the field is new, and the utility of CFM in informing treatment has yet to be evaluated.

Challenges

Online research can be messy, crowdsourcing platforms are changing all the time, and concerns about data quality are mounting. For example, inattentive responding to questionnaire items and behavioral tasks can induce spurious correlations between variables (102), while the presence of bots on certain services can threaten validity (103). Proposals to remedy this include a renewed focus on aligning incentives in online studies (i.e., considering what motivates people to participate and redesigning tasks to reflect that) (68), involving participants in design (104), and implementing more checks and balances (105). In tandem, there has been renewed focus on the reliability of the tasks we employ (106–109) and efforts to harmonize tasks across labs and species (110–112). Model-based planning, although far from a perfect assay, serves as an example of how advances in model-fitting have improved reliability (106,107) and how task design can be optimized to best capture individual differences (113). To date, there has been an overemphasis on snap-shot cross-sectional designs throughout computational psychiatry. While bridging more directly to treatment is the most important next step, we suggest that there are intermediate approaches that can already help the field move from correlation to causation. The next phase of research in this area should adopt richer, repeated within-subject designs that can establish temporal precedence of cognitive change and symptom change, thus helping to understand causality (114,115). Finally, an assumption of CFM is that the constructs under investigation are dimensional, following a linear progression from subclinical to clinical. While existing evidence suggests that this is a reasonable assumption in many cases, this may not hold for all aspects of psychiatry (8).

Outlook

CFM approaches have gained popularity in a variety of areas, and we have focused on those evaluated most thoroughly. CFM has also been used to study information seeking (116), deliberation (117), value-free random exploration (118), credit assignment (119), language use (120), foraging (121), mental effort avoidance (30,122), choice stochasticity (123), error-related negativity (124), and the interrelation of symptom dimensions (125). The approach has been extended to other areas of psychology also including the study of chronic pain (126), social interactions, learning and evaluations (127–129), and political leanings (130). A key challenge associated with the proliferation of studies is how to integrate knowledge across them. One

approach is to develop new questionnaires based on the CIT, AD, and SW output of CFM studies. Wise *et al.* (68) used machine learning to identify a battery of questions capable of capturing CIT, AD, and SW using just 20 items each. While we think this is an important end point for well-developed transdiagnostic dimensions, we also urge some caution. Factor analysis seeks to explain the data it is provided, which means that the choice of questionnaires included in each analysis can dramatically influence the factors that emerge. For example, studies with more specific and abundant anxiety-relevant items are less likely to merge anxiety and depression in a single factor (53). Moreover, the emergent factors are only as meaningful/relevant as the data fed into them and can be influenced by symptom-irrelevant features such as how questions are framed and how responses are recorded (131). Factor structures may also differ depending on characteristics of the sample being studied, an issue that is especially pertinent when considering clinical syndromes. Therefore, it is imperative to confirm the robustness and reliability of these structures. It is for this reason that some studies repeatedly interrogate the same factor structure across studies [e.g., AD, CIT, and SW: (30,36,46,68,101,116,120,124)], establishing that the association between dimensions and cognitive measures is replicable [e.g., (30,40)] and that results extend to diagnosed patients (32). While this is vital work, there are risks too in focusing narrowly on a single dimensional structure; specific factors, like disorders, may get reified as novel questionnaires and become difficult to change. If this occurs, we may miss the opportunity for incremental gain and refinement of measures or fail to see hidden hierarchical structures (or confounds) that influence our interpretations. To avoid this, we propose that researchers make modifications that can be systematically compared to ensure that we take steps forward with each new study, in much the same way that the field of psychometrics carefully balances evaluation of existing measures with iteratively refining the measurement of psychological constructs (132).

Most of the work we covered uses exploratory factor analysis, but there is no reason that CFM should be confined to this approach. Recent work with bifactor modeling (66) illustrated how this hierarchical approach may provide the best solution for certain mechanisms of psychopathology. We have focused on dimensionality reduction within self-report data, but there is no reason this approach could not also be used to reveal latent dimensions within behavior, too. For example, partial least squares regression has shown promise for more fully integrating the selection of factors with their underlying mechanisms (68). Future work should continue to expand the repertoire of CFM, for example by considering canonical correlation analyses and crossvalidation to identify novel and robust dimensions.

CONCLUSIONS

CFM is a new method that can help advance transdiagnostic, mechanistic research in psychiatry using large and unselected samples. The approach has identified new psychiatric dimensions with specific neurocomputational correlates, resolving seemingly nonspecific findings seen across disorders and revealing bidirectional effects that are hidden within a diagnosis.

Computational Factor Modeling

CFM complements traditional in-lab methods and diagnosis-led research; it speeds up and scales up research, and we hope that it can inform the development of interventions that are precisely targeted at a neurocomputational level.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was funded by a fellowship from MQ: Transforming Mental Health (Grant No. MQ16IP13 [to CMG]), a project award from Science Foundation Ireland's Frontiers for the Future Scheme (Grant No. 19/FFP/6418 [to CMG]), a European Research Council Starting Grant (Grant No. ERC-H2020-HABIT [to CMG]), Medical Research Council Senior Non Clinical Fellowship (Grant No. MR/R020817/1 [to OJR]), and a fellowship from the Anthony and Elizabeth Mellows Foundation (to TW).

CG holds industry partnership funding from SilverCloud Health, cofunded by the Irish Research Council. OR held a Medical Research Council Proximity-to-Discovery award with Roche, who have provided in-kind contributions regarding work on heart rate variability and anxiety. He is running an investigator-initiated trial with medication donated by Lundbeck (escitalopram and placebo; no financial contribution); has completed consultancy work on affective bias modification for Peak and on randomized clinical trials for anxiety for Roche; he also sat on the committee of the British Association of Psychopharmacology until 2022. TW reports no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom (TW); Neuroscience and Mental Health Group, Institute of Cognitive Neuroscience, University College London, London, United Kingdom (OJR); Research Department of Clinical Education and Health Psychology, University College London, London, United Kingdom (OJR); School of Psychology, Trinity College Dublin, Dublin 2, Ireland (CMG); Global Brain Health Institute, Trinity College Dublin, Dublin 2, Ireland (CMG); and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland (CMG).

Address correspondence to Claire Gillan, Ph.D., at claire.gillan@gmail.com or gillancl@tcd.ie.

Received Jun 10, 2022; revised Sep 9, 2022; accepted Sep 27, 2022.

REFERENCES

- Dalgleish T, Black M, Johnston D, Bevan A (2020): Transdiagnostic approaches to mental health problems: Current status and future directions. *J Consult Clin Psychol* 88:179–195.
- Robbins TW, Gillan CM, Smith DG, de Wit S, Ersche KD (2012): Neurocognitive endophenotypes of impulsivity and compulsivity: Towards dimensional psychiatry. *Trends Cogn Sci* 16:81–91.
- Hyman SE (2021): Psychiatric disorders: Grounded in human biology but not natural kinds. *Perspect Biol Med* 64:6–28.
- Maj M (2005): 'Psychiatric comorbidity': An artefact of current diagnostic systems? *Br J Psychiatry* 186:182–184.
- Freedman R, Lewis DA, Michels R, Pine DS, Schultz SK, Tamminga CA, et al. (2013): The initial field trials of DSM-5: New blooms and old thorns. *Am J Psychiatry* 170:1–5.
- Fried EI (2015): Problematic assumptions have slowed down depression research: Why symptoms, not syndromes are the way forward. *Front Psychol* 6:309.
- Markon KE, Chmielewski M, Miller CJ (2011): The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychol Bull* 137:856–879.
- Haslam N (2003): Categorical versus dimensional models of mental disorder: The taxometric evidence. *Aust N Z J Psychiatry* 37:696–704.
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. (2010): Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am J Psychiatry* 167:748–751.
- Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. (2017): The hierarchical taxonomy of psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol* 126:454–477.
- Gillan CM, Seow TXF (2020): Carving out new transdiagnostic dimensions for research in mental health. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5:932–934.
- Reiter AM, Atiya NA, Berwian IM, Huys QJ (2021): Neuro-cognitive processes as mediators of psychological treatment effects. *Curr Opin Behav Sci* 38:103–109.
- Paulus MP, Huys QJM, Maia TV (2016): A roadmap for the development of applied computational psychiatry. *Biol Psychiatry Cogn Neurosci Neuroimaging* 1:386–392.
- Yong AG, Pearce S (2013): A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *TQMP* 9:79–94.
- Schmitt TA (2011): Current methodological considerations in exploratory and confirmatory factor analysis. *J Psychoeducational Assess* 29:304–321.
- Gillan CM, Pappmeyer M, Morein-Zamir S, Sahakian BJ, Fineberg NA, Robbins TW, de Wit S (2011): Disruption in the balance between goal-directed behavior and habit learning in obsessive-compulsive disorder. *Am J Psychiatry* 168:718–726.
- Gillan CM, Apergis-Schoute AM, Morein-Zamir S, Urcelay GP, Sule A, Fineberg NA, et al. (2015): Functional neuroimaging of avoidance habits in obsessive-compulsive disorder. *Am J Psychiatry* 172:284–293.
- Gillan CM, Morein-Zamir S, Urcelay GP, Sule A, Voon V, Apergis-Schoute AM, et al. (2014): Enhanced avoidance habits in obsessive-compulsive disorder. *Biol Psychiatry* 75:631–638.
- Voon V, Derbyshire K, Rück C, Irvine MA, Worbe Y, Enander J, et al. (2015): Disorders of compulsivity: A common bias towards learning habits [No. 3]. *Mol Psychiatry* 20:345–352.
- Ersche KD, Gillan CM, Jones PS, Williams GB, Ward LHE, Luijten M, et al. (2016): Carrots and sticks fail to change behavior in cocaine addiction. *Science* 352:1468–1471.
- Foerde K, Daw ND, Ruffin T, Walsh BT, Shohamy D, Steinglass JE (2021): Deficient goal-directed control in a population characterized by extreme goal pursuit. *J Cogn Neurosci* 33:463–481.
- Sebold M, Nebe S, Garbusow M, Guggenmos M, Schad DJ, Beck A, et al. (2017): When habits are dangerous: Alcohol expectancies and habitual decision making predict relapse in alcohol dependence. *Biol Psychiatry* 82:847–856.
- Alvares GA, Balleine BW, Guastella AJ (2014): Impairments in goal-directed actions predict treatment response to cognitive-behavioral therapy in social anxiety disorder. *PLoS One* 9:e94778.
- Alvares GA, Balleine BW, Whittle L, Guastella AJ (2016): Reduced goal-directed action control in autism spectrum disorder. *Autism Res* 9:1285–1293.
- Geurts HM, de Wit S (2014): Goal-directed action control in children with autism spectrum disorders. *Autism* 18:409–418.
- Morris RW, Quail S, Griffiths KR, Green MJ, Balleine BW (2015): Corticostriatal control of goal-directed action is impaired in schizophrenia. *Biol Psychiatry* 77:187–195.
- Culbreth AJ, Westbrook A, Daw ND, Botvinick M, Barch DM (2016): Reduced model-based decision-making in schizophrenia. *J Abnorm Psychol* 125:777–787.
- Delorme C, Salvador A, Valabrègue R, Roze E, Palminteri S, Vidailhet M, et al. (2016): Enhanced habit formation in Gilles de la Tourette syndrome. *Brain* 139:605–615.
- Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016): Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* 5:e11305.
- Patzelt EH, Kool W, Millner AJ, Gershman SJ (2019): Incentives boost model-based control across a range of severity on several psychiatric constructs. *Biol Psychiatry* 85:425–433.
- Seow TXF, Benoit E, Dempsey C, Jennings M, Maxwell A, O'Connell R, Gillan CM (2021): Model-based planning deficits in compulsivity are linked to faulty neural representations of task structure. *J Neurosci* 41:6539–6550.

32. Gillan CM, Kalanthroff E, Evans M, Weingarden HM, Jacoby RJ, Gershkovich M, *et al.* (2020): Comparison of the association between goal-directed planning and self-reported compulsivity vs obsessive-compulsive disorder diagnosis. *JAMA Psychiatry* 77:77–85.
33. Fleming SM, Dolan RJ, Frith CD (2012): Metacognition: Computation, biology and function. *Philos Trans R Soc Lond B Biol Sci* 367:1280–1286.
34. Fu T, Koutstaal W, Fu CHY, Poon L, Cleare AJ (2005): Depression, confidence, and decision: Evidence against depressive realism. *J Psychopathol Behav Assess* 27:243–252.
35. Hoven M, Lebreton M, Engemann JB, Denys D, Luigjes J, van Holst RJ (2019): Abnormalities of confidence in psychiatry: an overview and future perspectives. *Transl Psychiatry* 9:1–18.
36. Rouault M, Seow T, Gillan CM, Fleming SM (2018): Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biol Psychiatry* 84:443–451.
37. Seow TXF, Gillan CM (2020): Transdiagnostic phenotyping reveals a host of metacognitive deficits implicated in compulsivity. *Sci Rep* 10:2883.
38. Seow TXF, Rouault M, Gillan CM, Fleming SM (2021): How local and global metacognition shape mental health. *Biol Psychiatry* 90:436–446.
39. Moses-Payne ME, Rollwage M, Fleming SM, Roiser JP (2019): Postdecision evidence integration and depressive symptoms. *Front Psychiatry* 10:639.
40. Hoven M, Denys D, Rouault M, Luigjes J, van Holst R (2022): How do confidence and self-beliefs relate in psychopathology: A transdiagnostic approach. *PsyArXiv*. <https://doi.org/10.31234/osf.io/d45gn>.
41. Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW (2010): Approach and avoidance learning in patients with major depression and healthy controls: Relation to anhedonia. *Psychol Med* 40:433–440.
42. Gradin VB, Kumar P, Waiter G, Ahearn T, Stickler C, Milders M, *et al.* (2011): Expected value and prediction error abnormalities in depression and schizophrenia. *Brain* 134:1751–1764.
43. Robinson OJ, Cools R, Carlisi CO, Sahakian BJ, Drevets WC (2012): Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. *Am J Psychiatry* 169:152–159.
44. Brown VM, Zhu L, Solway A, Wang JM, McCurry KL, King-Casas B, Chiu PH (2021): Reinforcement learning disruptions in individuals with depression and sensitivity to symptom change following cognitive behavioral therapy. *JAMA Psychiatry* 78:1113–1122.
45. Greenberg T, Chase HW, Almeida JR, Stiffler R, Zevallos CR, Aslam HA, *et al.* (2015): Moderation of the relationship between reward expectancy and prediction error-related ventral striatal reactivity by anhedonia in unmedicated major depressive disorder: Findings from the EMBARC study. *Am J Psychiatry* 172:881–891.
46. Pizzagalli DA, Jahn AL, O’Shea JP (2005): Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. *Biol Psychiatry* 57:319–327.
47. Huys QJ, Pizzagalli DA, Bogdan R, Dayan P (2013): Mapping anhedonia onto reinforcement learning: A behavioural meta-analysis. *Biol Mood Anxiety Disord* 3:12.
48. Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T, Hrynkiewicz L, *et al.* (2017): Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA Psychiatry* 74:790–797.
49. Suzuki S, Yamashita Y, Katahira K (2021): Psychiatric symptoms influence reward-seeking and loss-avoidance decision-making through common and distinct computational processes. *Psychiatry Clin Neurosci* 75:277–285.
50. Hales CA, Robinson ESJ, Houghton CJ (2016): Diffusion modelling reveals the decision making processes underlying negative judgement bias in rats. *PLoS One* 11:e0152592.
51. Locke SM, Robinson OJ (2021): Affective Bias through the Lens of Signal Detection Theory [No. 1]. *Comput Psychiatr* 5:4–20.
52. Love J, Robinson OJ (2020): “Bigger” or “better”: The roles of magnitude and valence in “affective bias. *Cogn Emot* 34:633–642.
53. Daniel-Watanabe L, McLaughlin M, Gormley S, Robinson OJ (2022): Association between a directly translated cognitive measure of negative bias and self-reported psychiatric symptoms. *Biol Psychiatry Cogn Neurosci Neuroimaging* 7:201–209.
54. Grupe DW, Nitschke JB (2013): Uncertainty and anticipation in anxiety: An integrated neurobiological and psychological perspective. *Nat Rev Neurosci* 14:488–501.
55. Barlow DH (2000): Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory. *Am Psychol* 55:1247–1263.
56. Robinson OJ, Pike AC, Cornwell B, Grillon C (2019): The translational neural circuitry of anxiety. *J Neurol Neurosurg Psychiatry* 90:1353–1360.
57. Chavanne AV, Robinson OJ (2021): The overlapping neurobiology of induced and pathological anxiety: A meta-analysis of functional neural activation. *Am J Psychiatry* 178:156–164.
58. Morris J, Christakou A, van Reekum CM (2016): Nothing is safe: Intolerance of uncertainty is associated with compromised fear extinction learning. *Biol Psychol* 121:187–193.
59. Hartley CA, Phelps EA (2012): Anxiety and decision-making. *Biol Psychiatry* 72:113–118.
60. Charpentier CJ, Aylward J, Roiser JP, Robinson OJ (2017): Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. *Biol Psychiatry* 81:1014–1022.
61. Lawrance EL, Gagne CR, O’Reilly JX, Bijsterbosch J, Bishop SJ (2022): The computational and neural substrates of ambiguity avoidance in anxiety [No. 1]. *Comput Psychiatr* 6:8–33.
62. Norbury A, Robbins TW, Seymour B (2018): Value generalization in human avoidance learning. *eLife* Lee D, editor 7:e34779.
63. Zbozinek TD, Charpentier CJ, Qi S, Mobbs D (2021): Economic decisions with ambiguous outcome magnitudes vary with low and high stakes but not trait anxiety or depression. *Comp Psychiatry* 5:119–139.
64. Wise T, Zbozinek TD, Charpentier CJ, Michelini G, Hagan CC, Mobbs D (2022): Computationally-defined markers of uncertainty aversion predict emotional responses during a global pandemic [published online ahead of print Jun 6]. *Emotion*.
65. Browning M, Behrens TE, Jochem G, O’Reilly JX, Bishop SJ (2015): Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* 18:590–596.
66. Gagne C, Zika O, Dayan P, Bishop SJ (2020): Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife* 9:e61387.
67. Wise T, Michely J, Dayan P, Dolan RJ (2019): A computational account of threat-related attentional bias. *PLoS Comput Biol* 15:e1007341.
68. Wise T, Dolan RJ (2020): Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample [No. 1]. *Nat Commun* 11:4179.
69. Grosskurth ED, Bach DR, Economides M, Huys QJM, Holper L (2019): No substantial change in the balance between model-free and model-based control via training on the two-step task. *PLoS Comput Biol* 15:e1007443.
70. Wheaton MG, Gillan CM, Simpson HB (2019): Does cognitive-behavioral therapy affect goal-directed planning in obsessive-compulsive disorder? *Psychiatry Res* 273:94–99.
71. Price RB, Gillan CM, Hanlon C, Ferrarelli F, Kim T, Karim HT, *et al.* (2021): Effect of experimental manipulation of the orbitofrontal cortex on short-term markers of compulsive behavior: A theta burst stimulation study. *Am J Psychiatry* 178:459–468.
72. Ahmari SE, Dougherty DD (2015): Dissecting Ocd circuits: From animal models to targeted treatments. *Depress Anxiety* 32:550–562.
73. Evans DW, Lewis MD, Iobst E (2004): The role of the orbitofrontal cortex in normally developing compulsive-like behaviors and obsessive-compulsive disorder. *Brain Cogn* 55:220–234.
74. Brown VM, Gillan CM, Renard M, Kaskie R, Degutis M, Wears A, *et al.* (2022): A double-blind study assessing the impact of orbitofrontal

Computational Factor Modeling

- theta burst stimulation on goal-directed behavior. *J Psychopathol Clin Sci* 131:287–300.
75. Schoenbaum G, Chang CY, Lucantonio F, Takahashi YK (2016): Thinking outside the box: Orbitofrontal cortex, imagination, and how we can treat addiction [No. 13]. *Neuropsychopharmacology* 41:2966–2976.
 76. Wells A, Fisher P, Myers S, Wheatley J, Patel T, Brewin CR (2012): Metacognitive therapy in treatment-resistant depression: A platform trial. *Behav Res Ther* 50:367–373.
 77. Carpenter J, Sherman MT, Kievit RA, Seth AK, Lau H, Fleming SM (2019): Domain-general enhancements of metacognitive ability through adaptive training. *J Exp Psychol Gen* 148:51–64.
 78. Engeler NC, Gilbert SJ (2020): The effect of metacognitive training on confidence and strategic reminder setting. *PLoS One* 15:e0240858.
 79. Harmer CJ, O'Sullivan U, Favaron E, Massey-Chase R, Ayres R, Reinecke A, *et al.* (2009): Effect of acute antidepressant administration on negative affective bias in depressed patients. *AJP* 166:1178–1184.
 80. Armstrong T, Olatunji BO (2012): Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clin Psychol Rev* 32:704–723.
 81. Bar-Haim Y, Lamy D, Pergamin L, van Bakermans-Kranenburg MJ, van IJzendoorn MH (2007): Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. *Psychol Bull* 133:1–24.
 82. Peckham AD, McHugh RK, Otto MW (2010): A meta-analysis of the magnitude of biased attention in depression. *Depress Anxiety* 27:1135–1142.
 83. Duyser FA, van Eijndhoven PFP, Bergman MA, Collard RM, Schene AH, Tendolker I, Vrijzen JN (2020): Negative memory bias as a transdiagnostic cognitive marker for depression symptom severity. *J Affect Disord* 274:1165–1172.
 84. Beck AT (1979): *Cognitive Therapy and the Emotional Disorders*. London: Penguin.
 85. Dichter GS, Felder JN, Petty C, Bizzell J, Ernst M, Smoski MJ (2009): The effects of psychotherapy on neural responses to rewards in major depression. *Biol Psychiatry* 66:886–897.
 86. Heller AS, Johnstone T, Light SN, Peterson MJ, Kolden GG, Kalin NH, Davidson RJ (2013): Relationships between changes in sustained fronto-striatal connectivity and positive affect in major depression resulting from antidepressant treatment. *Am J Psychiatry* 170:197–206.
 87. Webb CA, Auerbach RP, Bondy E, Stanton CH, Appleman L, Pizzagalli DA (2021): Reward-related neural predictors and mechanisms of symptom change in cognitive behavioral therapy for depressed adolescent girls. *Biol Psychiatry Cogn Neurosci Neuroimaging* 6:39–49.
 88. Berwian IM, Wenzel JG, Collins AGE, Seifritz E, Stephan KE, Walter H, Huys QJM (2020): Computational mechanisms of effort and reward decisions in patients with depression and their association with relapse after antidepressant discontinuation. *JAMA Psychiatry* 77:513–522.
 89. Browning M, Kingslake J, Dourish CT, Goodwin GM, Harmer CJ, Dawson GR (2019): Predicting treatment response to antidepressant medication using early changes in emotional processing. *Eur Neuropsychopharmacol* 29:66–75.
 90. Browning M, Bilderbeck AC, Dias R, Dourish CT, Kingslake J, Deckert J, *et al.* (2021): The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (PRE-DicT): An open-label, randomised controlled trial [No. 7]. *Neuropsychopharmacology* 46:1307–1314.
 91. Szasz PL, Hofmann SG, Heilman RM, Curtiss J (2016): Effect of regulating anger and sadness on decision-making. *Cogn Behav Ther* 45:479–495.
 92. Pulcu E, Shkrelil L, Holst CG, Woud ML, Craske MG, Browning M, Reinecke A (2019): The effects of the angiotensin II receptor antagonist losartan on appetitive versus aversive learning: A randomized controlled trial. *Biol Psychiatry* 86:397–404.
 93. Gorka SM, Lieberman L, Klumpp H, Kinney KL, Kennedy AE, Ajilore O, *et al.* (2017): Reactivity to unpredictable threat as a treatment target for fear-based anxiety disorders. *Psychol Med* 47:2450–2460.
 94. Grillon C, Chavis C, Covington MF, Pine DS (2009): Two-week treatment with the selective serotonin reuptake inhibitor citalopram reduces contextual anxiety but not cued fear in healthy volunteers: A fear-potentiated startle study [No. 4]. *Neuropsychopharmacol* 34:964–971.
 95. Goodkind M, Eickhoff SB, Oathes DJ, Jiang Y, Chang A, Jones-Hagata LB, *et al.* (2015): Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* 72:305–315.
 96. Doucet GE, Janiri D, Howard R, O'Brien M, Andrews-Hanna JR, Frangou S (2020): Transdiagnostic and disease-specific abnormalities in the default-mode network hubs in psychiatric disorders: A meta-analysis of resting-state functional imaging studies. *Eur Psychiatry* 63:e57.
 97. Sharp PB, Dolan RJ, Eldar E (2021): Disrupted state transition learning as a computational marker of compulsivity. *Psychol Med* 1–11.
 98. Castro-Rodrigues P, Akam T, Snorasson I, Camacho MM, Paixão V, Barahona-Corrêa JB, *et al.* (2022): Explicit knowledge of Task Structure Is the Primary Determinant of Human Model-Based Action. *Nat Hum Behav* 6:1126–1141.
 99. Sajjadian M, Lam RW, Milev R, Rotzinger S, Frey BN, Soares CN, *et al.* (2021): Machine learning in the prediction of depression treatment outcomes: A systematic review and meta-analysis. *Psychol Med* 51:2742–2751.
 100. Lee CT, Palacios J, Richards D, Hanlon AK, Lynch K, Harty S, *et al.* (in press): The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation. *BMC Psychiatry*.
 101. Dercon Q, Mehrhof SZ, Sandhu T, Hitchcock C, Lawson R, Pizzagalli DA, *et al.* (2022): A core component of psychological therapy causes adaptive changes in computational learning mechanisms. *PsyArXiv*. <https://doi.org/10.31234/osf.io/jmnek>.
 102. Zorowitz S, Niv Y, Bennett D (2021): Inattentive responding can induce spurious associations between task behavior and symptom measures. *PsyArXiv*. <https://doi.org/10.31234/osf.io/rynhk>.
 103. Burnette CB, Luzier JL, Bennett BL, Weisenmuller CM, Kerr P, Martin S, *et al.* (2022): Concerns and recommendations for using Amazon MTurk for eating disorder research. *Int J Eat Disord* 55:263–272.
 104. Crocker JC, Ricci-Cabello I, Parker A, Hirst JA, Chant A, Petit-Zeman S, *et al.* (2018): Impact of patient and public involvement on enrolment and retention in clinical trials: Systematic review and meta-analysis. *BMJ* 363:k4738.
 105. Donegan KR, Gillan CM (2022): New principles and new paths needed for online research in mental health: Commentary on Burnette *et al.* (2021). *Int J Eat Disord* 55:278–281.
 106. Shahar N, Hauser TU, Moutoussis M, Moran R, Keramati M, NSPN consortium, Dolan RJ (2019): Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. *PLoS Comput Biol* 15:e1006803.
 107. Brown VM, Chen J, Gillan CM, Price RB (2020): Improving the reliability of computational analyses: Model-based planning and its relationship with compulsivity. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5:601–609.
 108. Mkrтчian A, Valtou V, Roiser JP (2021): Reliability of decision-making and reinforcement learning computational parameters. *bioRxiv*. <https://doi.org/10.1101/2021.06.30.450026>.
 109. Pike AC, Tan K, Ansari HJ, Wing M, Robinson OJ (2022): Test-retest reliability of affective bias tasks. *PsyArXiv*. <https://doi.org/10.31234/osf.io/n2fkh>.
 110. Pike AC, Lowther M, Robinson OJ (2021): The importance of common currency tasks in translational psychiatry. *Curr Behav Neurosci Rep* 8:1–10.
 111. Bach DR (2022): Cross-species anxiety tests in psychiatry: Pitfalls and promises [No. 1]. *Mol Psychiatry* 27:154–163.

112. Eisenberg IW, Bissett PG, Zeynep Enkavi A, Li J, MacKinnon DP, Marsch LA, Poldrack RA (2019): Uncovering the structure of self-regulation through data-driven ontology discovery [No. 1]. *Nat Commun* 10:2319.
113. Kool W, Cushman FA, Gershman SJ (2016): When does model-based control pay off? *PLoS Comput Biol* 12:e1005090.
114. Neuser MP, Kraeutlein F, Kuehnel A, Teckentrup V, Svaldi J, Kroemer NB (2021): 04.27.441601: Influenza: A gamified assessment of value-based decision-making for longitudinal studies. *bioRxiv*. <https://doi.org/10.1101/2021.04.27.441601>.
115. Gillan CM, Rutledge RB (2021): Smartphones and the neuroscience of mental health. *Annu Rev Neurosci* 44:129–151.
116. Kelly CA, Sharot T (2021): Individual differences in information-seeking [No. 1]. *Nat Commun* 12:7062.
117. Hunter LE, Meer EA, Gillan CM, Hsu M, Daw ND (2022): Increased and biased deliberation in social anxiety. *Nat Hum Behav* 6:146–154.
118. Dubois M, Hauser TU (2022): Value-free random exploration is linked to impulsivity [No. 1]. *Nat Commun* 13:4542.
119. Shahar N, Hauser TU, Moran R, Moutoussis M, Bullmore ET, Dolan RJ (2021): Assigning the right credit to the wrong action: compulsivity in the general population is associated with augmented outcome-irrelevant value-based learning. *Transl Psychiatry* 11:1–9.
120. Kelley SW, Mhaonaigh CN, Burke L, Whelan R, Gillan CM (2022): Machine learning of language use on Twitter reveals weak and non-specific predictions. *npj Digit Med* 5:1–13.
121. Fan H, Gershman SJ, Phelps EA (2021): Trait somatic anxiety is associated with reduced directed exploration and underestimation of uncertainty [published online ahead of print Oct 3]. *Nat Hum Behav*.
122. Scholl J, Trier HA, Rushworth MFS, Kolling N (2022): The effect of apathy and compulsivity on planning and stopping in sequential decision-making. *PLoS Biol* 20:e3001566.
123. Suzuki S, Yamashita Y, Katahira K (2019): Exploration-related strategy mediates negative coupling between decision-making performance and psychiatric symptoms *bioRxiv* <https://doi.org/10.1101/730614>.
124. Seow TXF, Benoit E, Dempsey C, Jennings M, Maxwell A, McDonough M, Gillan CM (2020): A dimensional investigation of error-related negativity (ERN) and self-reported psychiatric symptoms. *Int J Psychophysiol* 158:340–348.
125. Petitot P, Scholl J, Attaallah B, Drew D, Manohar S, Husain M (2021): The relationship between apathy and impulsivity in large population samples [No. 1]. *Sci Rep* 11:4830.
126. Rogers AH, Garey L, Allan NP, Zvolensky MJ (2021): Exploring transdiagnostic processes for chronic pain and opioid misuse among two studies of adults with chronic pain. *Behav Res Ther* 136:103786.
127. Yu H, Lu C, Gao X, Shen B, Liu K, Li W, *et al.* (2022): Explaining individual differences in advantageous inequity aversion by social-affective trait dimensions and family environment. *Soc Psychol Pers Sci* 13:626–637.
128. Will G-J, Moutoussis M, Womack PM, Bullmore ET, Goodyer IM, Fonagy P, *et al.* (2020): Neurocomputational mechanisms underpinning aberrant social learning in young adults with low self-esteem. *Transl Psychiatry* 10:1–14.
129. Yu H, Cao R, Lin C, Wang S (2022): Distinct neurocognitive bases for social trait judgments of faces in autism spectrum disorder. *Transl Psychiatry* 12:1–13.
130. Schulz L, Rollwage M, Dolan RJ, Fleming SM (2020): Dogmatism manifests in lowered information search under uncertainty. *Proc Natl Acad Sci U S A* 117:31527–31534.
131. American Educational Research Association and American Psychological Association (1999): National Council on Measurement. In: *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
132. Smith GT, McCarthy DM (1995): Methodological considerations in the refinement of clinical assessment instruments. *Psychol Assess* 7:300–308.
133. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011): Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69:1204–1215.
134. Dickinson A, Weiskrantz L (1985): Actions and habits: The development of behavioural autonomy. *Phil Trans R Soc Lond B* 308:67–78.
135. Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018): What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* 100:490–509.
136. Doll BB, Duncan KD, Simon DA, Shohamy D, Daw ND (2015): Model-based choices involve prospective neural activity. *Nat Neurosci* 18:767–772.
137. Wise T, Liu Y, Chowdhury F, Dolan RJ (2021): Model-based aversive learning in humans is supported by preferential task state reactivation. *Sci Adv* 7:eabf9616.
138. Gillan CM, Otto AR, Phelps EA, Daw ND (2015): Model-based learning protects against forming habits. *Cogn Affect Behav Neurosci* 15:523–536.
139. Vikbladh OM, Meager MR, King J, Blackmon K, Devinsky O, Shohamy D, *et al.* (2019): Hippocampal contributions to model-based planning and spatial memory. *Neuron* 102:683–693.e4.
140. Fleming SM, Lau HC (2014): How to measure metacognition. *Front Hum Neurosci* 8:443.
141. Fleming SM, Dolan RJ (2012): The neural basis of metacognitive ability. *Philos Trans R Soc B Biol Sci* 367:1338–1349.
142. Rouault M, Fleming SM (2020): Formation of global self-beliefs in the human brain. *Proc Natl Acad Sci U S A* 117:27268–27276.
143. Schultz W, Dayan P, Montague PR (1997): A neural substrate of prediction and reward. *Science* 275:1593–1599.
144. Aylward J, Hales C, Robinson E, Robinson OJ (2020): Translating a rodent measure of negative bias into humans: The impact of induced anxiety and unmedicated mood and anxiety disorders. *Psychol Med* 50:237–246.
145. Ratcliff R, Smith PL, Brown SD, McKoon G (2016): Diffusion decision model: Current issues and history. *Trends Cogn Sci* 20:260–281.
146. Kahneman D, Tversky A (2013): Prospect theory: An analysis of decision under risk. In: MacLean LC, Ziemba WT, editors. *Handbook of the Fundamentals of Financial Decision Making*. World Scientific Handbook in Financial Economics Series: Volume 4. Vancouver: University of British Columbia, 99–127.
147. Lawson RP, Bisby J, Nord CL, Burgess N, Rees G (2021): The computational, pharmacological, and physiological determinants of sensory learning under uncertainty. *Curr Biol* 31:163–172.e4.