Identifying transdiagnostic mechanisms in mental health using computational factor modeling

Toby Wise, Oliver Robinson, Claire Gillan

PII: S0006-3223(22)01661-4
DOI: https://doi.org/10.1016/j.biopsych.2022.09.034
Reference: BPS 15015

To appear in: Biological Psychiatry

Received Date: 10 June 2022
Revised Date: 9 September 2022
Accepted Date: 27 September 2022

Please cite this article as: Wise T., Robinson O. & Gillan C., Identifying transdiagnostic mechanisms in mental health using computational factor modeling, Biological Psychiatry (2022), doi: https://doi.org/10.1016/j.biopsych.2022.09.034.

© 2022 Published by Elsevier Inc on behalf of Society of Biological Psychiatry.
Special issue: Biological Psychiatry’s upcoming special issue on Emerging Topics in Computational Psychiatric Research.

Title: Identifying transdiagnostic mechanisms in mental health using computational factor modeling

Short title: Computational Factor Modeling

Keywords: Transdiagnostic, Computational Modeling, Factor Analysis, RDoC, Cognition, Computational Factor Modeling

Authors: Toby Wise¹, Oliver Robinson²,³, Claire Gillan⁴,⁵,⁶

Author Affiliations
1. Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King’s College London
2. Neuroscience and Mental Health group, Institute of Cognitive Neuroscience, UCL, London
3. Research department of Clinical Education and Health Psychology, UCL, London
4. School of Psychology, Trinity College Dublin, Dublin 2
5. Global Brain Health Institute, Trinity College Dublin, Dublin 2
6. Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin 2
Abstract

Most psychiatric disorders do not occur in isolation and most psychiatric symptom dimensions are not uniquely expressed within a single diagnostic category. Our current treatments fail to work for around 25-40% of individuals, perhaps due, at least in part, to an over-reliance on diagnostic categories in treatment development and allocation. This review will describe ongoing efforts in the field to surmount these challenges and precisely characterise 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 modelling 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 Modelling (CFM). We describe in detail some recent applications of this method to understand transdiagnostic cognitive processes including model-based planning, metacognition, appetitive processing, and uncertainty estimation. In this context, we highlight how the method has been used to identify specific associations between cognition and symptom dimensions and reveal previously obscured relationships, how findings generalise to smaller in-person clinical and non-clinical samples, and the method is being adapted and optimised 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.
Introduction

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 RDoC and HiTop (9,10). Although advances have been made within these frameworks, they continue to depend on traditional research formulae in psychiatry. That is, a focus on small, diagnosed patient samples, or the interrogation of cognitive mechanisms after symptom-level phenomena are defined, rather than defining them both in concert. Here, we introduce a novel combination of interdisciplinary methods called ‘computational factor modeling’ (CFM; Figure 1; Box 1), which we believe can accelerate transdiagnostic research in psychiatry. In CFM, candidate transdiagnostic symptom dimensions are identified not in patients, but in unselected online samples that 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 data-driven dimensionality reduction of self-report questionnaire data and theory-driven computational modeling of behaviour that allow us to precisely characterise the cognitive processes that characterise a given dimension (11). In this paper, we will 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 endeavour, moving from descriptive summaries of behaviour, with multiple potential mechanistic accounts, to detailed, falsifiable and precise accounts. We will show how advances in these areas, though still in the early stages, have augmented our understanding of mental illness, yielding putative mechanisms underlying transdiagnostic symptom dimensions that are precise and mechanistically plausible. We will discuss how CFM can support new frameworks like RDoC and HiTop and might drive innovations in treatment development and allocation (12,13).

Model-Based Planning

A number of case-control studies observed altered goal-directed (“model-based”, Box 1) behaviour in OCD, which leaves patients vulnerable to rigid habitual behaviours (Figure 1A,1B) (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 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’s Syndrome (28) - also showed deficits relative to controls. This suggested two possibilities. Either alterations in goal-directed control are a general feature of psychopathology. Or non-specific links between mechanisms and clinical phenotypes arise from problems with the validity/dissociability of diagnoses. One of the challenges in resolving this debate is that 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 and colleagues eschewed the traditional case-control framework and recruited members of the public (29). Over 1400 individuals completed an online assessment of self-report clinical assessments and a behavioural task that allowed researchers to use computational modelling to parse model-based planning from more reflexive learning styles.
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, 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’, which cut through existing diagnostic rules and explained the blurring of model-based deficits across diagnoses. This association was specific; ‘Anxious-Depression’ and ‘Social Withdrawal’ were unrelated to these deficits (Figure 2D). This finding replicates online (30), in-person (31) and, critically, in patients with diagnoses (32), where 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 the specific and precise underlying mechanisms of transdiagnostic symptoms that play a role in multiple disorders, but are experienced to different degrees by individuals.

**Metacognition**

Another area where the CFM approach has had impact is in the study of metacognition, the ability to accurately reflect on one’s own thoughts, feelings and behaviours. Metacognition plays a vital role in adaptive decision-making and can be modelled using signal detection theory (Box 1; Figure 3A)(33). Alterations in metacognition have been observed in depression, where patients tend to think they perform worse than other people, despite comparable performance (34). Case-control studies suggest this effect is non-specific and have found it in anxiety, OCD, and schizophrenia too (35) (Figure 3B). Given the high rates of comorbidity across these conditions, it is possible that a symptom common to all these conditions is responsible. To test this, Rouault et al. (36) used CFM in two large online, unselected samples (Figure 3C) examining the same transdiagnostic factors from the first CFM study (29) alongside a perceptual decision-making task. The latent factors were highly consistent across the studies, with correlation of loadings of r=.87-.97. They found that ‘compulsive behaviour and intrusive thought’ was linked to positive metacognitive bias (i.e. over-confidence), while ‘anxious-depression’ was associated with negative metacognitive bias confidence (i.e. under-confidence).

These bi-directional associations were replicated and extended in a further study in an online, unselected sample using a learning task (37) (Figure 3D), suggesting that these opposing metacognitive deficits are generalised and pervade many aspects of self-reflection. Interestingly, this study also showed that these deficits likely stem from dissociable mechanisms. Individuals high in compulsivity, but not anxious-depression, showed a reduction in the extent to which their confidence assessments were updated based on evidence (37). Seow and colleagues (38) suggested that these distinct metacognitive biases are due to two different mechanisms; reduced confidence in depression may stem from global self-beliefs (e.g. self-esteem (39)) while over-confidence in compulsivity may relate to difficulties in
building a mental model of one’s performance (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 conceptualised as the clinical symptom of anhedonia and features prominently in depression, but also 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 fMRI, reduced neural signatures of reward prediction errors were also seen in both depression (42,43) and schizophrenia (42). Another paper with careful computational modelling of behavioural data in 69 patients with MDD showed that anhedonia was linked to both learning rate and outcome sensitivity biases (44). However, an fMRI 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 one single task (46) suggested that anhedonia was associated with reduced reward sensitivity in clinical samples (i.e. ‘consummatory pleasure’) but not reward learning per se (47). More dramatically, a study with both fMRI from a small case-control sample and behavioural data from >1800 general population users of a smartphone app found no neural or behavioural deficits in reward processing in the case control sample, but a relationship with depression symptoms in the unselected sample that was opposite to 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 anxious-depression factor in a healthy sample (49), but, perhaps critically, did not isolate depression and anxiety. In a different type of task translated from animal work, where biases in reward learning are examined by testing if learned reward values generalise to ambiguous cues (50), individuals with mood and anxiety disorders were more likely to ‘pessimistically’ assume that neutral cues will lead to low (rather than high) rewards, driven potentially by lower evidence accumulation for high rewards (Box 1)(51,52). Critically, CFM in 990 general population participants, 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,4D).
Uncertainty

Changes in uncertainty processing are thought to play a major role in anxiety disorders (54), where individuals report both feeling more uncertain (55), finding uncertainty 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 using 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, which revealed that a transdiagnostic anxiety factor was specifically associated with enhanced generalisation 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 trait, marker of anxiety that emerges in those exhibiting acute symptoms. In line with this account, one study in a large, unselected sample found 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 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 comprising clinically diagnosed patients with MDD and GAD, 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 (Figure 5C,5D,5E). These tasks assess how people respond to objective uncertainty, but recent work has shown that computational modelling can also be used to infer and quantify individual-level subjective uncertainty (67). Wise & 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, which incorporated a combination of behavioural data and self-report in the identification of transdiagnostic factors, constitutes an intriguing progression of the CFM approach and may assist in the more data-driven identification of dimensions of pathology going forward.

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 will describe in detail in the forthcoming section with reference to concrete examples that have begun to realise this
promise. For example, mechanistic insights can help us understand if and how existing treatments can change key neurocomputational processes, such as model-based planning or metacognition. This can inform the development of new treatments that can target these processes more effectively or selectively. There is additionally much hope that this method can help us deliver treatments more precisely, based on an individual’s specific transdiagnostic, and mechanistically defined, profile. This work is still in its infancy, and an important task for research in the coming years will be to realise this potential. In the following sections, we review the strides already 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, or if they could still 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 CBT for OCD (70), even in individuals who respond extremely well to treatment. 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 (cTBS) (71). The focus of this stimulation was to reduce left orbitofrontal cortex activation, building on prior knowledge of the role the OFC plays in both habit and compulsive behaviours (72,73). This treatment acutely decreased compulsive behaviour 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 itself (74). Further work is needed to determine if activation of habit circuits is necessary for patients to achieve this benefit from cTBS. If it does, this might provide 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, might 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 a way 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 how metacognition is defined in the field of computational psychiatry. One study of this sort randomised healthy individuals to receive training on their metacognitive assessments (77) and found that metacognitive performance improved and generalised 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 behaviours 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 face. An important next question for this area is whether metacognitive training can be delivered in a more personalised manner, based on what CFM has taught us about the dissociable correlates of metacognition, anxious-depression and compulsivity.
Reward

Attenuated reward processing is thought to contribute to ‘negative schemata’, which manifest in poorer emotional recognition of happy faces (79), attentional biases towards negative information (80–82), and biases for negative memories (83) in depression. The clinical relevance of these biases are perhaps one of the most long-considered in cognitive models of depression and as such are key targets for CBT (84). Indeed both therapy (85) and SSRIs (86) have been shown to increase striatal response to reward, and increased computationally-modelled pre-treatment reward responses are associated with a greater symptom improvement (87). Several recent computational modelling studies in clinical samples have made notable strides in this area. One study showed that the reduced reward learning rates associated with anhedonia normalised following CBT in MDD (44) (Figure 6C). Another found that relapse following discontinuation of SSRIs was predicted by reduced baseline effort expenditure to gain rewards (88). A third 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 outcomes in a subsequent clinical trial (90). Although the lack of generalisation is discouraging, this methodology is in many ways exemplary, 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 it reduced punishment learning). This suggests this adaptation is 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 behavioural assay of uncertainty processing) decreases after CBT, but not SSRI treatment (93). This dissociation is important as 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 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 reliably estimate if uncertainty processing can be addressed clinically, and if 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 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 bringing insights from CFM back to study mechanistically in a smaller in-person sample; Seow et al., examined the electrophysiological correlates of model-based planning in ~200 students who varied in their levels of compulsivity and intrusive thought. 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 (31). This converges with recent findings from general population samples suggesting failures in goal-directed control in impulsivity 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 should be included earlier in the discovery process and integrated with CFM approaches. Two potential extensions which ask slightly different questions are 1) identifying factors using CFM as reviewed above 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 markers of treatment-response (Figure 6E). One of the key challenges with this work is achieving the sample sizes necessary to develop and validate neurocomputational markers of treatment response. Similar issues have been faced by chronically under-powered machine learning research in the area of treatment response (99). Online methods can help here too. Lee and colleagues partnered with a digital CBT provider to recruit, assess (using CFM) and follow hundreds of patients through treatment in a short space of time (100). 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 study tightly constrained lab-models (analogues) of psychological therapy in large, unselected samples to understand how they work on CFM dimensions. Dercon and colleagues (101) took such an 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 two examples illustrate how CFM approaches can be integrated more directly to the study of how treatments work on well-defined computational processes, and how internet-based methods allow researchers to do this at an unprecedented 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 is messy, crowdsourcing platforms are changing all the time, and concerns about data quality are mounting. For example, inattentive responding to questionnaire items and behavioural tasks can induce spurious correlations between (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 harmonise 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 optimised to detect 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 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 prediction of mechanisms onto symptoms or vice versa, 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 most thoroughly evaluated. More broadly though, CFM has been used to study information seeking (116), deliberation (117), value-free random exploration (118), credit assignment (119), language use (120), foraging (120,121), mental effort avoidance (30), choice stochasticity (122), error-related negativity (123), and the inter-relation of symptom dimensions (124). The approach has been extended to other areas of psychology also, including the study of chronic pain (125), social interactions, learning and evaluations (126–128), and political leanings (129). 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 output of CFM studies. Wise and colleagues used machine learning to identify a battery of questions capable of capturing CIT, AD and SW using just 20 items each (68). While we think this is an important endpoint 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 how questions are framed and how responses are recorded (130). Factor structures may also differ depending on characteristics of the sample being studied, an issue that is especially pertinent when considering clinical syndromes. It is therefore imperative to confirm the robustness and reliability of these structures. It is for this reason some studies repeatedly interrogate the same factor structure across studies (e.g. AD, CIT and SW: (30,36,40,68,101,116,120,123)), 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 in focusing too narrowly on a single dimensional structure; factors, like disorders, may get reified as novel questionnaires and 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 we take steps forward with each new study, in much the same way as the field of psychometrics carefully balances evaluation of existing measures with iteratively refining the measurement of psychological constructs (131).

Most of the work we covered uses exploratory factor analysis, but there is no reason at CFM be confined to this approach. Recent work with bifactor modelling (66) illustrated how this hierarchical approach might provide the best solution for certain mechanisms of psychopathology. We have focused here on dimensionality reduction within self-report data, but there is no reason why this approach could not also be used to reveal latent dimensions within behaviour 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 considering canonical correlation analyses and cross-validation to identify novel and robust dimensions.

Conclusion

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 non-specific findings seen across disorders, and revealing bidirectional effects that are hidden within a diagnosis. CFM complements traditional in-lab methods and diagnosis-led research, it speeds up and scales up research and we hope it can inform the development of interventions that are precisely targeted at a neurocomputational level.

Funding Statement
Clare M Gillan is funded by a fellowship from MQ: transforming mental health (MQ16IP13), a project award from Science Foundation Ireland’s Frontiers for the Future Scheme (19/FFP/6418), and a European Research Council (ERC) Starting Grant (ERC-H2020-HABIT). Oliver J Robinson is funded by a Medical Research Council Senior Non Clinical Fellowship (MR/R020817/1). Toby Wise is supported by a fellowship from the Anthony and Elizabeth Mellows Foundation.

Conflicts of Interest
Claire M Gillan holds industry partnership funding from SilverCloud Health, confounded from the Irish Research Council. Oliver J Robinson 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 sits on the committee of the British Association of Psychopharmacology. Toby Wise reports no biomedical financial interests or potential conflicts of interest.
Figure Legends

Figure 1. Computational Factor Modeling

Computational Factor Modeling (CFM) aims to identify transdiagnostic symptom dimensions associated with precise neurocomputational mechanisms. The method looks ‘under the hood’ of cognitive processes with 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 cross-cutting data-driven ‘latent’ symptom dimensions (e.g. ‘Compulsivity and Intrusive Thought’) in large unselected samples, typically gathered online. Computational models are then fit to participants’ behaviour, to identify theory-driven latent behavioural dimensions (e.g. ‘Learning Rate’). The relationship between these two sets of latent factors is then examined and can be iteratively and bi-directionally refined.

Figure 2. Model-Based Planning

(A) CFM has been used to identify a transdiagnostic psychiatric dimension related to deficits in model-based planning (see Box 1 for detailed definition). Individual items (circles) from a range of questionnaires relating to traditional diagnoses (DX, colours) were subjected to factor analysis. Three dimensions resulted: Anxious-Depression, Compulsivity and Intrusive Thought, and Social Withdrawal. Behavioural data on a two-step decision making task were fit using a computational model that extracted individual estimates of model-based planning, which 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 the balance between goal-directed behaviour 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 CIT dimension.

(E) These results replicated in diagnosed patients, and moreover, effects were stronger when measuring individual differences in compulsivity compared to diagnostic status (OCD or not) (32).

Figure 3 Metacognition

(A) 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, yielding the same structure and highly correlated loadings to the original paper (all r>.87). This time, transdiagnostic factors were related to metacognitive bias,
a person’s tendency to over or under-estimate their own performance at 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 non-clinical samples. As for model-based planning, patterns of association blur across diagnostic lines, showing fairly consistent reductions in metacognitive bias (aka confidence) (35).

(C) Using CFM, Rouault et al., 2018 (36) showed that in fact a bi-directional association exists, where anxious-depression is linked to decreased confidence in performance, while compulsivity and intrusive thought is characterised by increased confidence. This illustrates how traditional methods using heterogeneous disorder categories may ‘average out’ specific and transdiagnostic processes.

Figure 4. Reward Processing

(A) Newer 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 the original questionnaires, demonstrating that anxiety and depression do not always lie together.

(B) Reward processing has been studied in great detail in psychiatry using small-scale case-control designs focusing on depression. But 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.

Figure 5. Uncertainty

(A) CFM studies have recently begun to adopt other approaches to dimensionality reduction. One paper by Gagne et al. (66) reduced a range of questionnaires into a general internalising factor as well as two specific factors relating to depression and anxiety. They tested for association 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 internalising factor, rather than being specific to what distinguishes depression and anxiety from one-another (66).
An important next step for CFM research is to integrate it more directly with treatment. There are a number of promising examples, including using CFM-defined mechanisms to (A) optimise 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 behaviour) within treatment studies. There are at least two ways this could be of value. Firstly, (D) directly testing the impact of treatment on previously identified transdiagnostic dimensions. Or secondly, (E) using a data driven approach on repeated measures data to identify latent dimensions that specifically predict treatment response.
References


https://doi.org/10.3389/fpsyt.2019.00639


70. Wheaton MG, Gillan CM, Simpson HB (2019): Does cognitive-behavioral therapy affect


108. Zorowitz S, Niv Y, Bennett D (2021, April 12): Inattentive responding can induce
spurious associations between task behavior and symptom measures. PsyArXiv. https://doi.org/10.31234/osf.io/rynhk


Traditional categorical diagnosis approach

- Diagnosis 1
- Diagnosis 2
- Diagnosis 3
- Diagnosis 4

Questionnaire items

Computational factor modelling (e.g., Gillan et al., 2016)

- Diagnosis 1
- Diagnosis 2
- Diagnosis 3
- Diagnosis 4

Online data from large samples

Factor analysis
Identifies transdiagnostic dimensions

Exploratory factor analysis (eFA) for CFM - a brief guide
- eFA assumes correlations in the data (here, self-report symptom measures) because items measure common underlying latent variables.
- By performing eFA on this data, we determine how many latent variables are present in the data and which items load on to these variables. These variables may be associated with items across measures designed for distinct diagnoses, representing transdiagnostic dimensions.
- A key step in this procedure is identifying the correct number of factors. This can be achieved using statistical approaches (such as inflection points on a scree plot, or parallel analysis) but should also account for interpretability of the factors. Do the factors make sense, clinically?
- Scores on each factor can be extracted for every subject, allowing analyses relating these to parameters from models of behavior.
- With enough data (exploration and validation sets), researchers can begin to approach this in a data-driven way without risk of over-fitting. Asking not whether the factors are interpretable, but which factor structure maps a given cognitive capacity best.

We refer the reader to existing tutorials on eFA for further detail (14, 15)

Associations with model parameters
Establish neurocomputational mechanisms underpinning transdiagnostic dimensions

Stronger loading
- Weaker loading

Behavioral model parameters
Insights from CFM
Reduced model-based planning is linked to a transdiagnostic compulsivity and intrusive thought factor (Gillan et al., 2016; 2020)

A. Symptom measures
- DX1, DX2, DX3, DX4, DX5, DX6, DX7

B. Prior research
Suggested model-based planning was affected in multiple disorders (Gillan et al., 2011)

C. Model-Based Learning (% change)
- Eating disorders
- Impulsivity
- OCD
- Alcohol Addiction
- Schizophrenia
- Depression
- Trait Anxiety
- Apathy
- Social Anxiety

D. Computational model
- Anxious-depression
- Compulsivity & intrusive thought
- Social withdrawal

E. Two-step decision-making task
- A: 1, 2, 3, 4

Model-based planning coefficient
- GAD
- OCD
- GAD & OCD
Insights from CFM
Compulsivity and intrusive thought is linked to increased confidence (Rouault et al., 2018; Seow et al., 2020)

A
Symptom measures

Factor analysis

Anxious-depression
Compulsivity & intrusive thought
Social withdrawal

Signal detection theory model

Accuracy
Confidence level
Metacognitive efficiency

Perceptual decision-making task

Confidence?

B
Prior research
Found disruptions in metacognition across multiple disorders (Hoven et al., 2019)

C

<table>
<thead>
<tr>
<th>Decision</th>
<th>Metacognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>NS</td>
</tr>
<tr>
<td>Confidence level</td>
<td>*** NS ***</td>
</tr>
</tbody>
</table>

Regression coefficient

- ‘Anxious-Depression’
- ‘Compulsive Behavior and Intrusive Thought’
- ‘Social Withdrawal’
Insights from CFM
A depression factor is associated with negative affective bias during reward processing (Daniel-Watanabe et al., 2022)

A
Symptom measures

Factor analysis
Anxiety
Depression
Schizotypy
OCD

Drift diffusion model
Affective bias
Drift rate

Affective bias task

B
Prior research
Results from small-scale case-control studies have been mixed (Rutledge et al., 2017)

C

![Image of graph showing relationship between depression and affective bias](image)

![Image of graph showing relationship between depression and drift rate](image)
Insights from CFM
A general internalizing factor is linked to impaired adaptation to volatility (Gagne et al., 2020)

A
Symptom measures
DX1
DX2
DX3
DX4
DX5
DX6
DX7

Exploratory bifactor analysis
General internalizing factor
+ Depression
Anxiety

Reinforcement learning model
Stable learning rate
Volatile learning rate

Value learning task with volatility manipulation

Behavior

B
Prior research
Linked failure to adjust learning rates in line with volatility to trait anxiety (Browning et al., 2015)

C
General internalizing
Depression
Anxiety

Stable
Volatile
Stable
Volatile
Stable
Volatile
High scorers
Low scorers
Use of mechanisms identified through CFM in treatment

A - Targeting brain regions involved in mechanisms identified with CFM (e.g., Price et al., 2021)
B - Training mechanisms identified by CFM (e.g., Carpenter et al., 2019)
C - Linking mechanisms to improvement (e.g., Brown et al., 2021)

D - CFM in the context of treatment - proposed approach 1
- Treatments targeting computational mechanisms identified by CFM produce improvements in corresponding transdiagnostic dimensions
- Treatment targeting model-based planning
- Treatment allocated to those with model-based planning deficits

E - CFM in the context of treatment - proposed approach 2
- Symptom measures repeated throughout treatment period
- Novel transdiagnostic dimensions identified based on changes in symptoms and computational parameters with treatment

Factor analysis:
- Factor 1
- Factor 2
- Factor 3

Computational model:
- Task repeated throughout treatment period
- Parameter value

Behavior:
- Compulsivity & intrusive thought
- Anxious-depression
- Social withdrawal

Parameter change

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

- **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 behaviour in a prospective fashion (132–134) (Figure 2). Rather than relying on direct experience of reward, our model-based faculties allow us to simulate future states (135,136), to 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 behaviours called habits (137) that appear rigid and outside intentional control. The first empirical studies testing these ideas trained OCD patients to perform responses to stimuli to gain rewards, then subsequently reduced the value of those rewards (an “outcome devaluation” procedure) and tested if behaviour ceased. In a range of experimental preparations, OCD patients were found to persist in responding (16–18). Later, more sophisticated ‘two-step’ tasks used reinforcement learning to characterise 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 vmPFC activity (132) and requires the hippocampus (138), highlighting these as potential targets for investigation with regard to compulsivity.

- **Metacognition.** Recent years have seen a proliferation of novel tasks and analysis approaches (139), enabling more precise estimates of metacognition than purely self-report approaches (Figure 3). Tasks that measure it typically focus on having participants complete a perceptual decision making task, such as estimating which side of a screen has more dots displayed. Staircase procedures can be 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 sorts of tests allow researchers to derive two components using signal detection theory models: metacognitive bias (i.e. over- or under-confidence in your own performance) and metacognitive sensitivity (i.e. how well confidence discriminates correct vs incorrect responses) (139). Metacognition involves the lateral PFC and dACC (140,141), suggesting that these areas may be relevant for anxious-depression and compulsivity (36).

- **Reward Processing.** Reward (appetitive) 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’ (142), 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 a more sensitive (or potential distinct) measure of this can be gleaned from studying how values that we learn to associate with cues both 1) spread or ‘generalise’ to other similar cues (50,143) as well as 2) drive reaction time changes
through changes in the way that evidence is accumulated (as modelled by drift diffusion modelling (143). The affective bias task used to link depressive symptoms to affective bias and drift rate (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-centred 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 0 or 100), or “ambiguous” options where information is obscured (unknown probability of 0 or 100). Performance on these tasks can be modelled using Prospect Theory models (144) to isolate behavioural tendencies including risk aversion (avoiding uncertain outcomes), ambiguity aversion (avoiding unknown outcome probabilities or magnitudes), reward maximisation (choosing higher expected values) and loss aversion (overweighting potential losses relative to gains). Other sorts of tasks have looked at how people learn under conditions of uncertainty. Browning et al. (65) examined this using a task where participants learn to choose between two options with different probabilities and magnitudes of punishment (Figure 5). These decisions take place in two states, one where the correct choice is stable and another where it switches frequently, inducing volatility. To avoid punishment, learning rates as modelled using reinforcement learning should increase in volatile states so that recent outcomes are prioritised over old. 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 (145), suggesting that this neuromodulator could play a role in internalising symptoms.