Mechanistic arbitration between candidate dimensions of psychopathology

Celine A Fox[1,2*], Vanessa Teckentrup[1,2], Kelly R Donegan[1,2], Tricia XF Seow[3], Christopher SY Benwell[4], Brenden Tervo-Clemmens[5], Claire M Gillan[1,2]

[1]School of Psychology, Trinity College Dublin, Dublin, Ireland
[2]Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland
[3]Wellcome Centre for Human Neuroimaging, University College London, London, UK
[4]Division of Psychology, School of Humanities, Social Sciences and Law, University of Dundee, Dundee, UK
[5]Department of Psychiatry & Behavioral Sciences, University of Minnesota, Minnesota, USA

Funding: This project was funded by an ERC starting grant (ERC-H2020-HABIT) and a project grant from a Science Foundation Ireland (19/FFP/6418) awarded to Claire M. Gillan. Celine A Fox is supported by an Government of Ireland Postgraduate Scholarship (GOIPG/2020/662). Vanessa Teckentrup is supported by a Government of Ireland Postdoctoral Fellowship (GOIPD/2023/1238).

* = Celine A Fox is the corresponding author

Abstract

Transdiagnostic and dimensional alternatives to the Diagnostic and Statistical Manual of Mental Disorders have gained prominence in recent years. A key critique of these approaches, however, is that they are typically based on symptom correlations and as such are purely descriptive and therefore unlikely to have mechanistic grounding. We tested this idea empirically, conducting a large-scale comparison of competing factor solutions that allowed us to determine if the latent structure underlying the covariation of psychiatric symptoms has robust and specific cognitive correlates. In nine independent datasets, comprising N=7565 individuals including patients, healthy individuals, paid and unpaid participants, a broad set of age ranges and cognitive task variants that measured model-based planning and metacognition, we found that this was the case. Across datasets, the factors with the best fit to cognition were those derived from a first-order factor analysis on the maximal number of theoretically informed self-report symptoms available. These factors ('Compulsivity and Intrusive Thought' and 'Anxious-depression') performed better than thousands of engineered alternatives and performed twice as well as traditional questionnaire total scores. Crucially, this unsupervised approached based on symptom correlation only performed on-par with a partial least squares analysis, a supervised approach to deriving factors based on cognition. These results provide evidence that unsupervised factor analysis of psychiatric symptoms is a viable method for rethinking how we define mental health and illness, affording clear opportunities for enhancing our understanding of specific underlying mechanistic processes.

INTRODUCTION

Dimensional and transdiagnostic approaches to psychiatric classification have gained traction over the last decade[1,2], positioned as an alternative to the Diagnostic and Statistical Manual of Mental Disorders[3]. Their advantages are that they incorporate the lack of clear boundary between healthy and unhealthy states[4,5] and their existence accounts for the heterogeneity of symptom presentation within diagnostic categories[6,7], the explicit overlap in symptoms[8] and high rates of co-morbidity[9,10] across disorders. These challenges in defining valid and reliable clinical phenotypes have translated into a lack of progress in developing a mechanistic understanding of mental health[11–13]. Differences in cognitive functioning between healthy individuals and patients are small and disorder non-specific [14,15], as are genetic[16] and brain-behaviour associations[17–19]. To achieve progress in our mechanistic understanding of psychiatric conditions, there is a growing need for an empirically derived, valid and reliable, set of transdiagnostic dimensional clinical phenotypes for research. Without this, a ceiling is placed on the size and specificity of our scientific observations that cannot be addressed by recent initiatives to gather larger and more diverse datasets, or indeed advances in neuroimaging techniques[20], computational modelling[21] and artificial intelligence[22].

Although transdiagnostic, dimensional approaches to psychopathology have many proponents[2,23] and are central to major funding initiatives[24], there remains a lack of clarity around how candidate dimensions should be identified and validated. One increasingly popular approach is to derive dimensions by examining the natural covariation of symptoms between-subjects in the population using factor analysis[25,26]. A key conceptual critique of this method, however, is that it operates at the symptom level only[27]. Descriptive taxonomies based on symptom covariance are limited, as symptoms that hang together do not necessarily have a shared mechanism[28] or etiology[29]. Rather, a single dimension of related symptoms across DSM disorders can be reached via multiple mechanistic pathways[28]. Symptoms can cause one-another[30], the same symptoms can be arrived at from different causes and conversely, a single underlying cause can produce different symptoms in different individuals for example depending on early childhood experiences[31]. For these reasons, it is unclear if a covariation-based framework can lead us to etiologically or mechanistically useful clinical phenotypes. Emerging evidence suggests that, despite this potential limitation, psychiatric dimensions derived from factor analysis of symptoms nonetheless outperform classic

frameworks in their association with cognitive-computational processes. For example, model-based planning, the tendency to use prospective mental maps to guide behaviour[32], has a stronger and more specific association to a transdiagnostic dimension 'Compulsivity and Intrusive Thought' than traditional disorder categories[33–36]. Metacognitive bias - the confidence one has in their own performance[37] – tends to be decreased in a non-specific way across many DSM disorder categories and questionnaires[38], but in fact has a highly specific and bi-directional association with two different transdiagnostic dimensions. Individuals high in 'Anxious-depression' are underconfident, whereas those with 'Compulsivity and Intrusive Thought' express overconfidence[39–44].

These examples illustrate that covariance-based clinical phenotypes may have more mechanistic validity compared with traditional diagnoses[26]. However, while conceptually symptom-based latent factors are hypothesised to give rise to co-occurence among disorders, factor solutions are fundamentally constrained by the specific dataset at hand. As such, they change depending on the inclusion or omission of symptoms or diagnoses[45]. It is possible that clinical dimensions that map even more closely to cognition exist, but do not correspond to the most substantial axes of variation in a given questionnaire set. A related issue is that factor modelling involves multiple degrees of analytic freedom[29], for example regarding the number of factors retained[46], the inclusion of a higher- order general factor[47], and whether dimensions of psychopathology should be orthogonal or partially correlated[48]. There is currently no consensus on these issues and problematically, specification decisions can produce a large, possibly infinite, set of alternative and well-fitting models[49]. Finally, and as articulated above, these models are designed to find the best solution for the covariance matrix of symptoms provided, these need not necessarily be optimal for or differentially correspond to distinct etiologies or mechanisms.

This paper aimed to resolve these issues by using data from over 7000 individuals from nine independent datasets[33–36,39,40,42,43,50], with multiple variations of cognitive tasks measuring model-based planning and metacognition, various age ranges, data-collection methods, and representing a spread of general, crowd-sourced and clinical populations. To generate an analytic consensus on the structure of psychopathology, we generated thousands of competing factor solutions, spanning variations in the number of factors retained, a multiverse of expanded and contracted questionnaire sets, and examining bifactor, orthogonal and oblique rotations and tested for the maximal association with the cognitive-computational

capacity of interest. We benchmarked performance against a baseline of established transdiagnostic dimensions (i.e., 'Anxious-depression' and 'Compulsivity and Intrusive Thought'), which were derived from a 3-factor solution, an oblique rotation, and based on the full set of items. Finally, we compared the mechanistic validity and identity of factors derived from the covariance of symptoms alone – as is the case in prominent dimensional frameworks such as the Hierarchical Taxonomy of Psychopathology (HiTOP)[25], versus a fully supervised method (partial least squares 'PLS' regression) that uses cognitive information to inform the derivation of factors.

RESULTS

Generating a transdiagnostic factor multiverse

In a factor discovery dataset of N=1413 individuals gathered from Amazon's Mechanical Turk (AMT)[34], we generated thousands of possible factor solutions based on a core dataset of 209 questionnaire items. These measured symptoms of obsessive compulsive disorder (OCD), eating disorders, impulsivity, schizotypy, apathy, alcohol addiction, depression, trait anxiety and social anxiety[34]. Using this set, we examined how variations in the number of factors retained per model (1- to 10-factor solutions, n=55 factors; Figure 1B(i)) affects the association between resulting dimensions and model-based planning and metacognition. We then tested the impact of the size and composition of the questionnaire set selected for analysis on the resulting factor solutions. To do this, we generated every possible combination of the nine self-report clinical scales (N=511) and factor-analysed the item-level responses, specifying an oblique ('oblimin') rotation and considering solutions retaining up to 5 factors (15 possible factors per model), resulting in 7665 candidate dimensions (Figure 1B(ii)). We followed this with a comparison of factors derived from oblique, orthogonal rotations and a bifactor model (Figure 1B(iii), n=10 factors). Mechanistic validity was defined as the magnitude of the effect size of a given dimensions in predicting a given cognitive capacity, controlling for age, gender and education/IQ. This was calculated across several datasets and summarised in a weighted average.

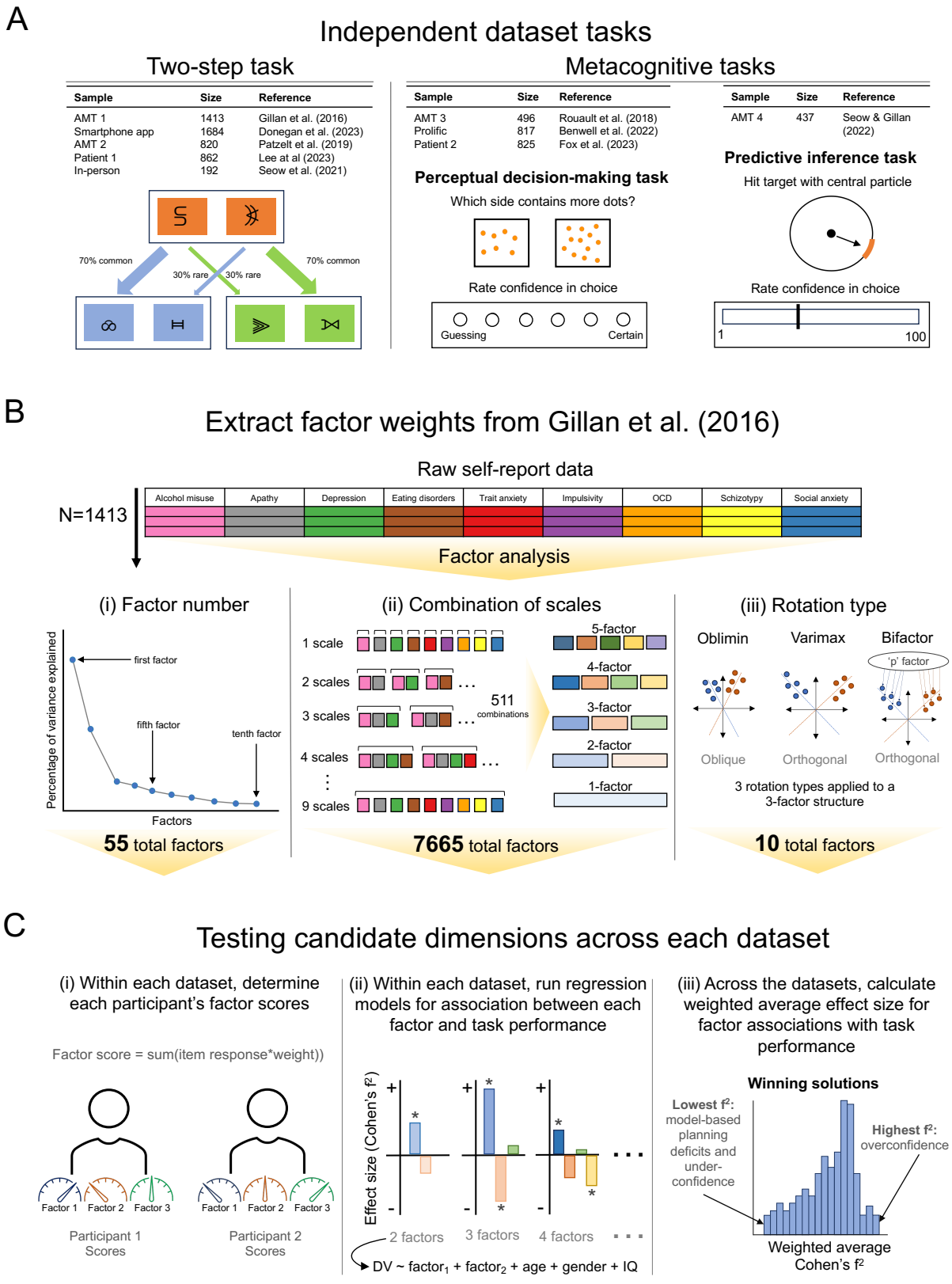


Figure 1. General Methodology. AMT = Amazon's Mechanical Turk. The four AMT and two patient datasets are numerically labelled to help distinguish samples (e.g., AMT 1 = Gillan et al. (2016) dataset (N=1413)). **(A)** Five datasets included alternative versions of the two-step task to assess model-based planning abilities. Four datasets included metacognitive tasks: Three included alternative versions of a visuo-perceptual decision-making task, and one dataset used a predictive inference task. **(B)** Factor weights were extracted from a discovery dataset [34], following separate manipulations of (i) the number of factors retained (ranging from a single- to 10-factor structure, resulting in 55 total factors), (ii) the combinations of scales (ranging from a single

questionnaire to all nine questionnaires considered, with each combination subsequently analysed as a single- to 5-factor structure, resulting in 7665 total factors), and (iii) the rotation type (oblique 'oblimin', orthogonal 'varimax' or 'bifactor' rotations, each as a three-factor structure (in addition to a hierarchical factor for the bifactor rotation), resulting in 10 total factor solutions). **(C)** (i) The performance of each dimension was assessed by examining how within dataset factor scores (ii) were associated with model-based planning/metacognitive bias, controlling for age, gender and IQ/education, using linear regression analyses to extract the effect size (Cohen's f^2 value of each factor). (iii) The weighted average Cohen's f^2 values across datasets was used to determine which factor had the largest 'winning' effect size for predicting individual differences in model-based planning (lowest Cohen's f^2 value), overconfidence (highest Cohen's f^2 value), and underconfidence (lowest Cohen's f^2 value) respectively.

Comparing transdiagnostic dimensions to questionnaire total scores

As a preliminary step, we calculated the weighted average effect sizes for total scores across the nine clinical questionnaire and the established transdiagnostic dimensions, for associations with model-based planning and metacognitive bias (Figure 2).

To compare the performance of total scores on the nine clinical scales to the three established transdiagnostic dimensions, we conducted separate regression analyses per cognitive task within each dataset, with the questionnaires/dimensions as an independent variable, along with age, gender and education/IQ. Regression analyses with clinical questionnaires included each questionnaire in a separate model (e.g., metacognitive bias ~ depression scale score + age + gender + IQ), while regression analyses for dimensions included the three dimensions in the same model (e.g., metacognitive bias ~ Anxious-depression + Compulsivity and Intrusive Thought + Social Withdrawal + age + gender + IQ). Examining the weighted average effect sizes (Cohen's f^2) across datasets for each questionnaire and dimension, transdiagnostic dimensions outperformed clinical questionnaires across all cognitive facets (Figure 2). For deficits in model-based planning, 'Compulsivity and Intrusive Thought' was the top performing factor (Cohen's $f^2 = -0.014$), performing better than all the clinical questionnaires (Figure 2A). Similar for metacognitive bias, 'Compulsivity and Intrusive Thought' had the largest positive effect size for individual differences in overconfidence (Cohen's $f^2 = 0.038$), and 'Anxious-depression' had the largest negative effect size for individual differences in underconfidence (Cohen's $f^2 = -0.030$) (Figure 2B). The effect sizes of 'Compulsivity and Intrusive Thought' and 'Anxious-depression' for metacognitive bias were twice as large as the effect sizes of the next best performing questionnaires (the Obsessive-Compulsive Inventory-Revised for overconfidence with a Cohen's f^2=0.016, and the Apathy Evaluation Scale with a Cohen's $f^2 = -0.015$ for underconfidence) (Figure 2B).

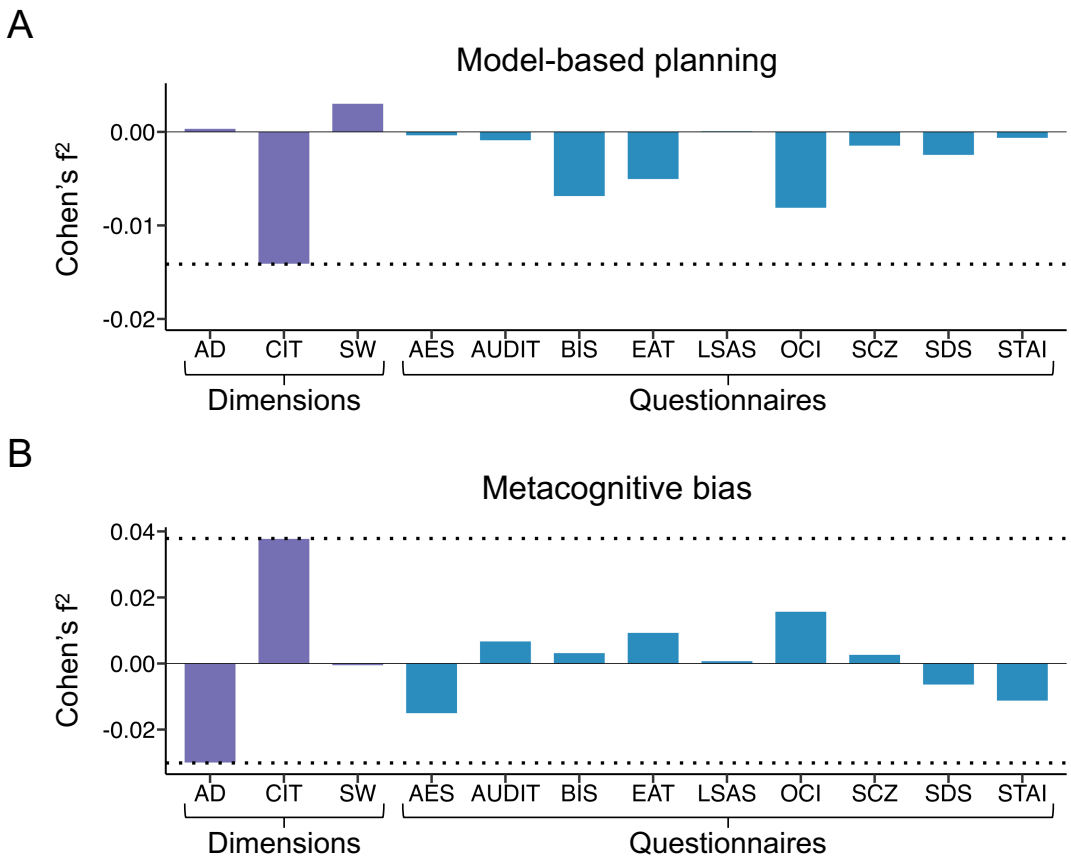
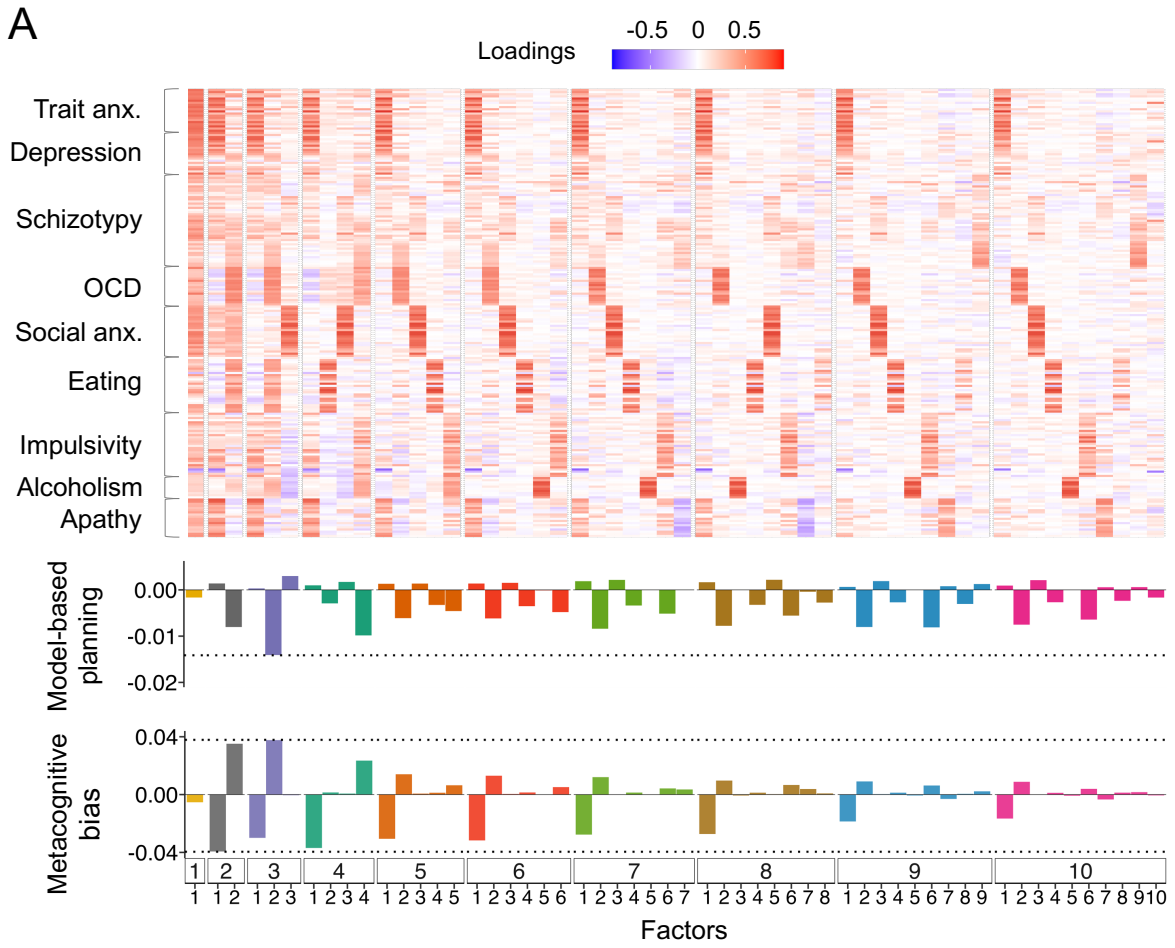


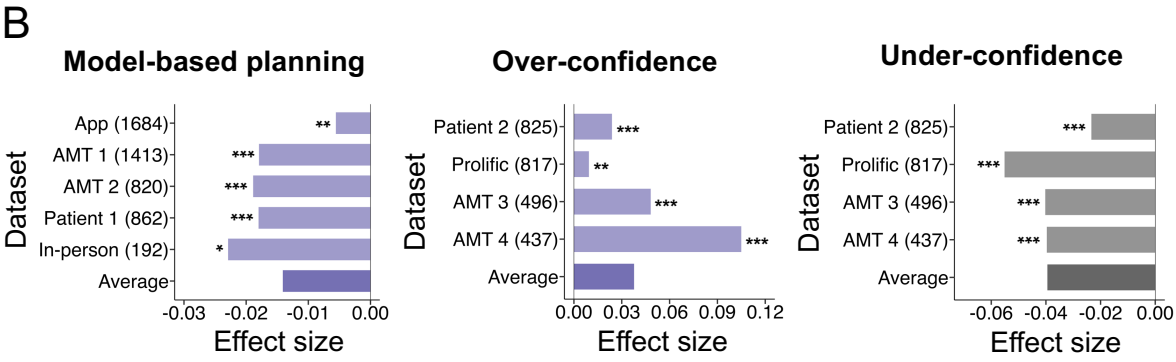
Figure 2. Weighted average effect sizes of total scores on clinical questionnaires and transdiagnostic dimensions. AD = Anxious-depression, CIT = Compulsivity and Intrusive Thought, SW = Social Withdrawal, AES = Apathy Evaluation Scale, AUDIT = Alcohol Use Disorders Identification Test, BIS = Barratt Impulsiveness Scale, EAT = Eating Attitudes Test, LSAS = Liebowitz Social Anxiety Scale, OCI = Obsessive-Compulsive Inventory-Revised, SCZ = Short Scales for Measuring Schizotypy, SDS = Zung Self-Rating Depression Scale, STAI = State Trait Anxiety Inventory. **(A)** Weighted effect sizes of dimensions and questionnaires from linear regression analyses predicting model-based planning, averaged across 5 datasets (N=4990). **(B)** Weighted effect sizes of dimensions and questionnaires from linear regression analyses predicting metacognitive bias, averaged across 4 datasets (N=2575).

Variation of the number of factors selected for retention

Factors that do not represent a significant axis of variation (based on our selection criteria, see below) in a questionnaire set may not be selected for retention; but they may nonetheless constitute an important and mechanistically homogenous dimension of psychopathology. Examining how variations in the number of factors retained per model (Figure 1B(i)), we generated a total of 55 factors, with the first from a 1-factor solution, the following 2 from a 2-factor solution, all the way to a 10-factor model (Figure 3A). For model-based planning, none of these factors outperformed the benchmark of 'Compulsivity and Intrusive Thought', which was the 2nd factor of a three-factor solution (Figure 3A), as suggested by the Cattell-Nelson-Gorsuch (CNG) indices[34], a mathematical formalisation of the scree plot method[51].

We repeated this analysis for metacognitive bias, which has two poles of clinical association – 'Compulsivity and Intrusive Thought' is associated with higher confidence and 'Anxious-depression' with lower confidence. We found that once again no factor outperformed the benchmark of 'Compulsivity and Intrusive Thought' in their positive association with metacognitive bias (i.e. higher confidence) (Figure 3A). However, the first factor from a 2-factor solution had a nominally stronger negative association with metacognitive bias than 'Anxious-depression' (Figure 3A). The top factors had statistically significant effects on model-based planning and metacognitive bias (all $p<0.05$) within each individual dataset (Figure 3B). Comparing the top performing factor with 'Anxious-depression', they were effectively identical; scores from participants in the discovery dataset (N=1413) were correlated at $r(207)=.99$, $p<0.001$ (Figure 3C). Likewise, the loadings were correlated at $r(207)=.94$, $p<0.001$ (Figure 3D). Additionally, scores and loadings from the second factor from the 2-factor solution was highly correlated ($r_{both}>0.80$) with 'Compulsivity and Intrusive Thought' (Figure S2).





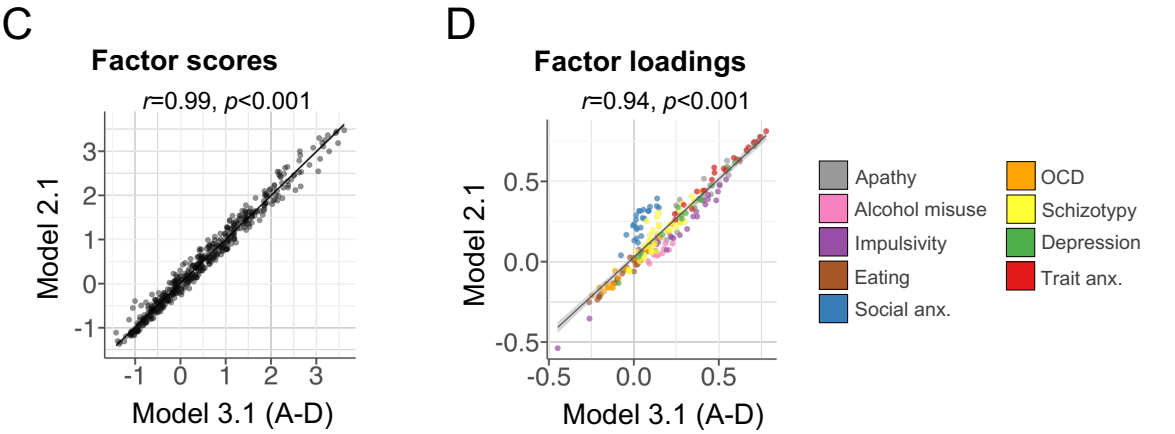


Figure 3. Variation of the number of factors selected for retention. AMT = Amazon's Mechanical Turk. A-D = Anxious-depression, OCD = Obsessive compulsive disorder, Social anx = Social anxiety, Trait anx = Trait anxiety, r= Pearson correlation coefficient, p = p-value. **(A)** Fifty-five factors were

generated from solutions retaining 1-factor, 2-factors, …, 10-factors. Heatmap indicates the loading of individual items onto each resulting factor (top panel). Weighted effect sizes from 10 linear regression analyses predicting model-based planning, averaged across 5 datasets (N=4990) (middle panel) and metacognition in 4 datasets (N=2575) (bottom panel). Factor loadings and effect sizes are aligned along the same x-axis. **(B)** Top-performing factors' effect sizes within each dataset and the weighted averaged effects across datasets. The top-performing factors had a statistically significant effects on model-based planning overconfidence and underconfidence within each dataset (all $p<0.05$). *=p<0.05, **=p<0.01, ***=p<0.001. The top performing factor for model-based planning and overconfidence was extracted from the 3-factor model (in purple) and the top performing factor for underconfidence came from the 2-factor model (in grey). **(C)** Correlation between scores on the first factor in a 3-factor solution (model 3.1 i.e., 'Anxious-depression') and the first factor from a 2-factor solution (model 2.1) in the discovery dataset. **(D)** Correlation between factor loadings of 'Anxious-depression' and model 2.1 in the discovery dataset.

Variation in the selection of clinical symptoms

We conducted a multiverse analysis that systematically varied not just the number of factors retained, but the symptom feature set itself. Specifically, we examined all combinations of 1-9 questionnaires, considering factor solutions 1-5 for each, resulting in 7665 candidate dimensions. For their association with model-based planning, there was a large positive spike in effect sizes around 0, but the majority of factors had negative Cohen's f^2 effect sizes (81.03%, 6211/7665) (Figure 4A). 'Compulsivity and intrusive thought' was the 9th best factor for explaining individual differences in model-based planning deficits (Cohen's f^2=-0.011), performing better than 99.88% of all factors generated (Figure 4A). The top performing factor (Cohen's f^2=-0.012) and 'Compulsivity and intrusive thought' both had highest loadings for OCD items and the lowest loadings for apathy items, with the top factor being generated from a set of questionnaires that omitted the schizotypy and alcohol addiction questionnaires (Figure S3A). Scores across the factors were highly correlated ($r(1411)=0.85$, $p<0.001$) in the discovery dataset (N=1413) (Figure 4B). Within datasets, the top performing factor outperformed 'Compulsivity and Intrusive Thought' for the association with model-based planning in 2/5 cases (Figure 4C). The top performing factor (Cohen's f^2=-0.012) and 'Compulsivity and Intrusive Thought' (Cohen's f^2=-0.011) both had very small weighted average effect sizes (Cohen's $f^2<0.02$)[52] (Figure 4C). The mean difference between the weighted average effect size of the top performing factor and 'Compulsivity and Intrusive Thought' was not substantial in magnitude (difference in Cohen's f^2=0.001) (Figure 4C). A heatmap of the top 1500 dimensions illustrated the relative importance of each questionnaire for explaining individual differences in model-based planning (Figure S4A). Averaging item-level loadings across the top 100 dimensions (Figure S5A), items related to OCD (M=0.51,

SD=0.01), followed by eating disorders (M=0.20, SD=0.07), had the highest average loadings (Figure S5D).

'Compulsivity and Intrusive Thought' had a stronger association with overconfidence than 99.70% of factors (Cohen's f^2=0.039), making it the 23rd top factor across analyses controlling for age, gender, IQ/levels of education and levels of 'Anxious-depression' (Figure 4D). Relative to 'Compulsivity and Intrusive Thought', the top performing factor (Cohen's f^2=0.041) omitted the trait anxiety and eating disorder questionnaires (Figure S3B), and scores on both factors were positively correlated in the discovery dataset at $r(1441)=0.62$, $p<0.001$ (Figure 4E). While the effect size was nominally greater for the top performing factor across samples, this difference was miniscule (Cohen's f^2 difference = 0.002), and the top factor did not outperform 'Compulsivity and Intrusive Thought' within half (2/4) of the datasets (Figure 4F). Items related to OCD, schizotypy and impulsivity had strong positive loadings across the top 1500 performing factors (Figure S4B) and specifically, items from the OCD questionnaire contributed most predominantly to the top 100 dimensions associated with overconfidence (Figure S5B), and had the highest questionnaire-level average loadings (M=0.36, SD=0.02) (Figure S5D).

When examining negative associations with confidence bias, 'Anxious-depression' was the 731st best performing factor (better than 90.46% of factors) in terms of its association with confidence (Cohen's f^2=-0.036) controlling for age, gender, IQ/levels of education and levels of 'Compulsivity and Intrusive Thought' (Figure 4G).'Anxious-depression' scores were highly correlated at ($r(1411)=0.95$, $p<0.001$), with the top performing factor (Cohen's f^2=-0.041) (Figure 4H), which had high loadings for the trait anxiety items (Figure S3C). Despite having a larger negative effect size across samples, the top generated factor did not consistently outperform 'Anxious-depression', which had a larger effect size in the Prolific dataset (N=817) (Figure 4I). Overall, the differences in effect size between the top performing dimensions versus 'Anxious-depression' were negligible, reflecting a weighted effect size difference of r=0.005 for underconfidence (Figure 4I). Visualising the heatmap of item loadings across the top 1500 factors, trait anxiety contributed highly to explaining individual differences in underconfidence (Figure S4C), and had the highest average item- and questionnaire-level loadings across the top 100 dimensions (M=0.59, SD=0.02) (Figure S5C-D).

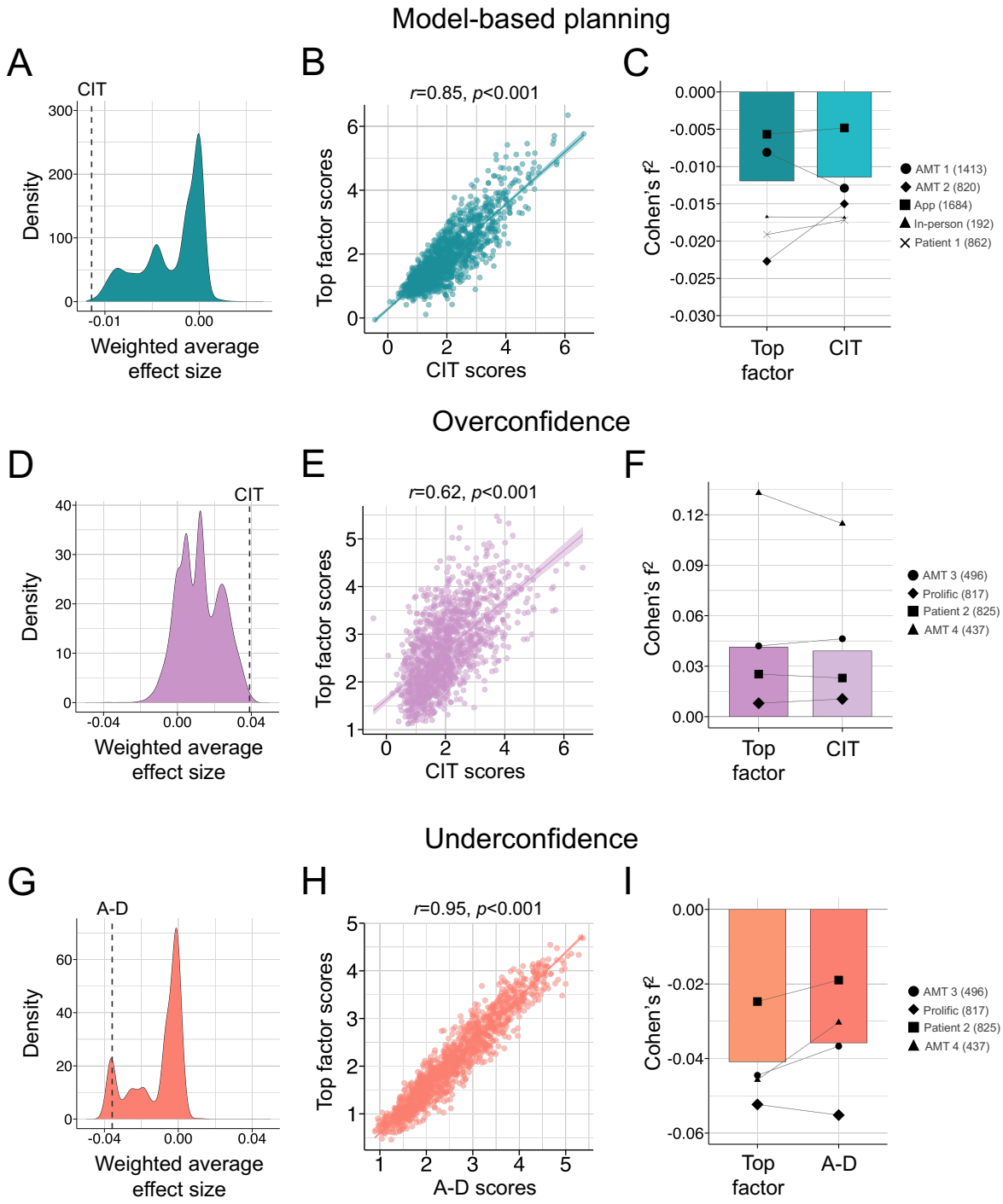


Figure 4. Variation in the selection of clinical symptoms. CIT = Compulsivity and Intrusive Thought, A-D = Anxious-depression, r= Pearson correlation coefficient, p = p-value. **Model-based planning. (A)** Density plot of effect sizes from 7665 candidate dimensions derived from all combinations of 9 questionnaires, retaining 1-5 factor solutions, predicting model-based planning. CIT was the 9th best performing of 7665. **(B)** Final scores for the top factor and CIT correlated at r=0.85, p<0.001 within the discovery dataset (N=1413). **(C)** Weighted average effect sizes across datasets (bars) and individual dataset effect sizes (dots) for the top factor and CIT associations with model-based planning. Overall effect sizes gains were negligible and inconsistent across datasets. **Overconfidence. (D)** Density plot of effect sizes predicting overconfidence. CIT was the 80th best performing of 7665. **(E)** Final scores between the top factor and CIT were correlated at r=.62, p<0.001 in the discovery dataset. **(F)** Weighted average effect sizes across datasets (bars) and individual dataset effect sizes (dots) for the top factor and CIT associations with overconfidence. Overall effect sizes gains were negligible and inconsistent across datasets. **Underconfidence. (G)** Density plot of effect sizes predicting underconfidence. A-D was the 731st best performing of 7665. **(H)** Final scores for the top factor and A-D were

correlated at r=0.95, p<0.001 in the discovery dataset. **(I)** Weighted average effect sizes (bars) and individual dataset effect size (dots) for the top factor and A-D associations with underconfidence. Overall effect sizes gains were negligible, and the top factor inconsistently outperformed A-D.

Higher- versus first-order factor rotation

Next we examined how factor models might best be rotated, and tested for evidence of utility for a higher order solution, which includes a general hierarchical ('p') factor. The cross-solution correlations (Figure 5A) of scores were very high across oblique (oblimin) and orthogonal (varimax) first-order solutions, but the oblique rotation (2nd factor – 'Compulsivity and Intrusive Thought') produced the nominally largest association with deficits in model-based planning (Cohen's f2 = -0.014) across our 5 datasets (Figure 5B). The same was true for both under- and over-confidence in metacognitive biases. 'Compulsivity and Intrusive Thought' showed the largest positive association (Cohen's f2 = 0.038) and 'Anxious-depression' the largest negative association (Cohen's f2 = -0.030) (Figure 5B). Notably, the general ('p') factor from the bifactor model showed the relatively smaller associations with all three cognitive domains (Figure 5B). Zooming in on this, associations with the 'p' factor were not statistically significant within any of the datasets for model-based planning (all p>0.05). The 'p' factor was associated with metacognitive bias in two datasets, but these were in opposing directions (negative association in the Prolific, N=817 and positive association in the AMT 4, N=437 datasets) (Figure 5C).

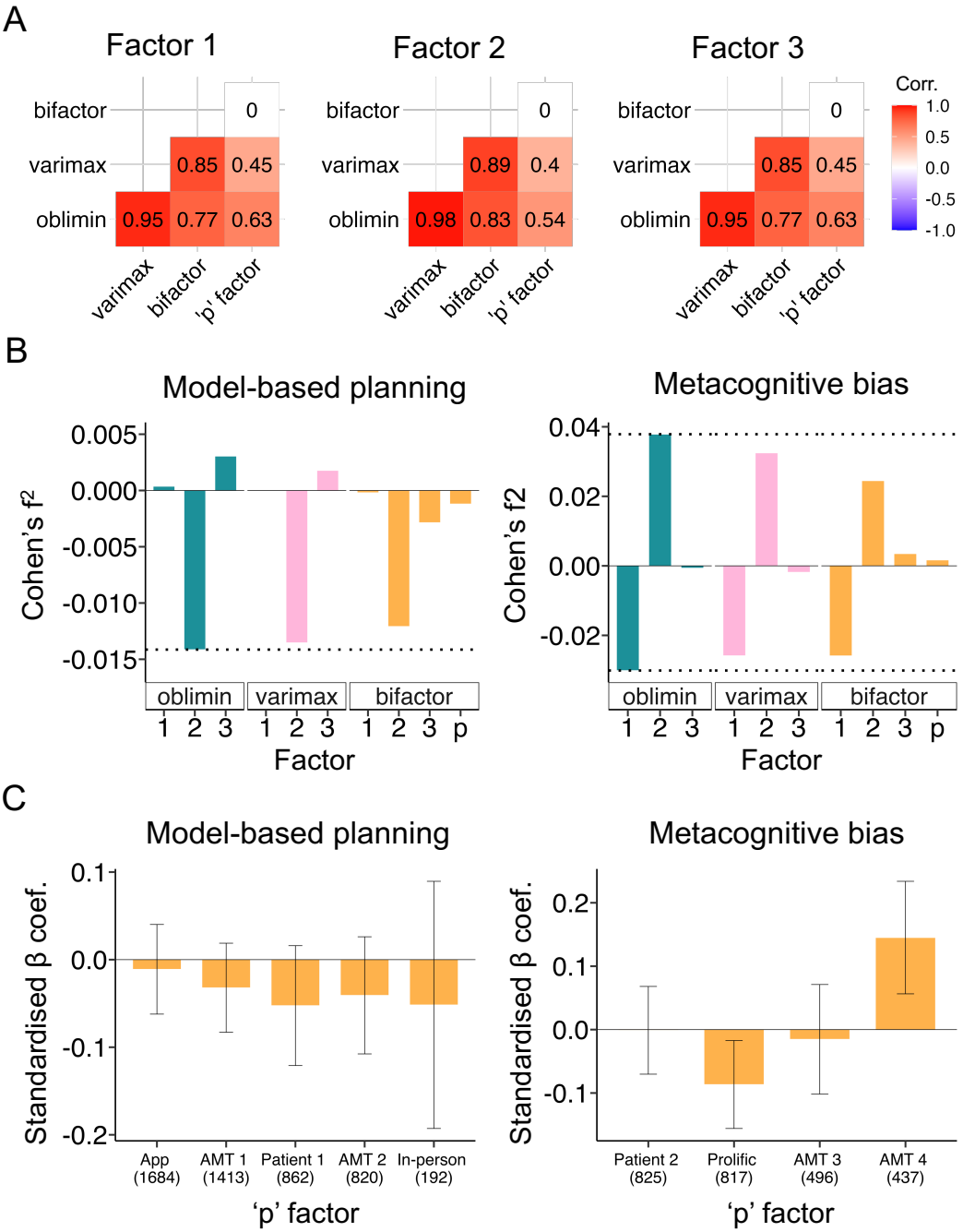


Figure 5. Higher- versus first-order factor rotation. AMT = Amazon's Mechanical Turk. p = general factor extracted from bifactor model. **(A)** Correlation of scores on factor 1, 2 and 3 from the 3 derivations – oblique, orthogonal and bifactor models in the discovery dataset (N=1413). Each correlation matrix also includes the correlation to the bifactor models general ('p') factor. **(B)** Weighted effect size for the association between derived factors and model-based planning and metacognitive bias. An oblique solution produces the winning model in all cases. **(C)** Individual results from component datasets for the 'p' factor reveals no consistent pattern of association with the cognitive constructs in question.

Comparison to a supervised method to derive factors

The preceding analyses focused on the extent to which natural patterns of symptom covariation possess mechanistic validity. In a final analysis, we compared the resulting

factors to those derived using a fully supervised approach, where cognitive performance (i.e., model-based planning and metacognitive bias) informs the selection of the factor itself. Specifically, we used partial least squares regression with 10-fold cross-validation within 75% of the discovery dataset[34] to first identify a transdiagnostic component linked to individual differences in model-based planning (Figure S7A). Across the test folds of the cross-validation procedure run on the discovery set (N=1061), higher component scores were associated with model-based planning (residualised for age, gender and IQ) with an average r^2=0.02 in the test folds of the cross-validation. Evaluating the model in the held out 25% of the data from the discovery sample (N=352), the PLS component score was significantly negatively associated with the residual of model-based planning abilities, r(350)=-0.15, p=0.004 (Figure S8A). Testing model performance out of sample, the PLS component performance was higher than 'Compulsivity and Intrusive Thought' in 3 of the 4 independent datasets (Figure S8B). While the weighted average effect size for the PLS component (Cohen's f^2 = -0.015) was nominally larger than the effect size for 'Compulsivity and Intrusive Thought' (Cohen's f^2 = -0.011), the average gain in effect size was not substantial (difference in Cohen's f^2=0.004) (Figure S8B). In the full discovery dataset (N=1413), scores on the PLS factor were strongly correlated at $r(1411)=0.86$, $p<0.001$ with 'Compulsivity and Intrusive Thought' (Figure S8C).

A PLS component was generated to predict individual differences in overconfidence within the AMT 3 sample (N=496)[40] (Figure S7B). The PLS component was associated with the residual values of confidence (after controlling for age, gender, IQ and levels of 'Anxious-depression') at r^2=0.07, averaged across the cross-validation test folds in the training 75% of the data (N=372). The PLS component was significant positively associated with mean confidence in the held out 25% of the data (N=124), $r(122)=0.21$, $p=0.017$ (Figure S8D). With out-of-sample testing, the weighted average effect size of the PLS component (Cohen's f^2=0.037) was not nominally larger than the effect of 'Compulsivity and Intrusive Thought' (Cohen's f^2=0.037) when predicting individual differences in overconfidence across all 3 independent datasets (Figure S8E). Within the full discovery dataset (N=496), scores on the PLS component were highly correlated ($r(494)=0.91$, $p<0.001$) with 'Compulsivity and Intrusive Thought' scores (Figure S8F). We repeated this analysis residualising 'Compulsivity and Intrusive Thought', instead of 'Anxious-depression', from the confidence measure to identify a PLS component significantly negatively associated with confidence (Figure S7C). This component had an average r^2=0.06 in the training set (N=372) and was negatively

associated with the residual of confidence in the held out 25% of the data (N=124), $r(122)=-0.17$, p=0.055 (Figure S8G). The PLS component did not outperform 'Anxious-depression' in 2 of the 3 out-of-sample tests, and had a lower weighted average effect size (Cohen's $f^2=-0.035$) compared to 'Anxious-depression' (Cohen's $f^2=0.036$) (Figure S8H). Scores for the PLS component factor were strongly correlated ($r(494)=0.95$, $p<0.001$) with 'Anxious-depression' scores (Figure S8I).

DISCUSSION

There is broad agreement that current diagnostic categories present challenges for mechanistic research in psychiatry[53,54]. There is less agreement, however, about how alternative classification frameworks should be generated. Some have suggested that correlation at the level of symptoms can reveal an important latent structure of mental health, which is statistically robust and reproducible[25,26,47]. Others have argued that this approach may be fraught, because there is no reason to think that a descriptive construct, albeit one that is statistically valid, has any specific mechanistic influence[27,55]. This issue is compounded by the analytic degrees of freedom inherent in such correlation-based approaches, wherein researcher choices at the time of data collection and analysis can produce many alternative factor solutions.

In this paper, we conducted an exhaustive multiverse analysis of potential factor solutions and tested their mechanistic validity, operationalised as their association with three aspects of cognition. In nine datasets, comprising N=7565 individuals, we provide evidence that those factors with the greatest mechanistic validity converge on the factors derived from a first-order factor analysis with oblique rotation, applied to a transdiagnostic sets of the maximal number of theoretically-informed self-report symptoms available to the researcher. These canonical factors ('Compulsivity and Intrusive Thought' and 'Anxious-depression') performed better than 99% and 90% (respectively) of alternatives in their association with cognition. These factors consistently out-performed total scores on constituent questionnaires and, in some cases, dimension effects were twice as large as classic clinical questionnaire effects. In cases where they performed nominally worse than the top-performing factor from multiverse analyses, the differences in effect size were trivial, inconsistent across datasets and

correlations between competing factors approached unity. Perhaps most surprisingly, we showed that dimensions derived from factor analysis at the symptom-level had as strong an association to cognition as dimensions derived from a supervised approach that expressly maximised that same association to cognitive performance. This suggests that novel dimensions of psychopathology can be identified using symptoms alone, opening the door to leveraging other large existing datasets, without cognitive test data, to identify novel transdiagnostic dimensions. Our results generalised across nine independent samples with markedly different characteristics, including patients, the general population, paid and unpaid participants, of different ages, collected in-person, online or in-app, and using several variants of the cognitive tasks in question. This remarkable stability across heterogeneous datasets suggests that factor solutions generated based on the correlation of experienced symptoms are a valid and powerful approach for defining novel transdiagnostic dimensions for research in psychiatry.

As part of this multiverse analysis, we showed that a first-order correlated factors model outperformed a bifactor model. The 'p' factor has garnered significant interest in psychiatric research since its initial identification, and has been suggested by some to reflect the liability, comorbidity, persistence, and severity of psychopathology[31]. It features prominently in the HiTOP framework as a 'superspectrum', and although it is without doubt a statistical feature of symptom features sets, concerns have been raised about the interpretation of the 'p' factor[56,57], and more broadly the evaluation and interpretation of bifactor models, including model overfitting[58,59], and poor replication across samples[60]. Across the nine datasets and their aggregates, we found no association between the 'p' factor and any of the three cognitive phenotypes under study. These results suggest two things. First, it suggests that the 'p' factor, if it constitutes an important level of clinical description, may reflect nonspecific psychopathology[31]. Second, the 'p' factor is near indistinguishable from the sum of the feature matrix that generates it[57], and so the finding that it is unrelated to cognitive test performance further underscores that the associations we report between cognition, 'Anxious-depression' and 'Compulsivity and Intrusive Thought' are not generalised effects.

Mechanistic validity is just one way in which we might evaluate a candidate classification system in psychiatry. Alternative objective functions, for example maximising discriminative treatment-prediction or future risk, could produce different dimensions that are of more immediate clinical value[55], even if they provide little-to-no insights with respect to aetiology

or brain-behaviour correspondence. We do not see these as competing goals, *per se,* but rather different levels of analysis. The current focus on improved mapping of symptom dimensions to neurocomputational mechanisms is designed to provide a new, theory-based path to developing precision therapeutics that target cognitive processes, or using these symptom dimensions to guide selection of existing treatments[28]. Future research should test if treatment-oriented vs. mechanistic approaches to symptom validation are distinct and complementary or converge on the same or similar factors. The present study cannot speak directly to this potential. However, there is some preliminary evidence that transdiagnostic dimensions derived from computational factor modelling may help us understand how treatments may work. For example, those with the largest increases in metacognitive confidence following internet-based cognitive behaviour therapy had the greatest reductions in 'Anxious-depression'[43]. To further elucidate the causal computational mechanisms of this effect, more intensive repeated testing would be a promising approach to account for temporal and contextual variation[61].

While the top performing factors had considerably stronger cognitive correlates than individual questionnaire total scores, the average effect sizes remained small in magnitude overall. This underscores how individual cognitive tests can at best provide a highly simplistic mechanistic account of psychopathology[21]. For a comprehensive explanatory model, multivariable models are required, incorporating micro (e.g., neural circuits) and macro (e.g., societal factors) levels of information[62]. An additional limitation of the current study is that we only considered nine specific clinical questionnaires, which yielded three factors, only two of which had robust cognitive correlates. Future work aiming to capture the full spectrum of psychopathology must consider additional facets of mental health and illness (e.g., autism spectrum disorder or attention deficit hyperactive disorder[63]), which are theoretically relevant to the cognitive facet on interest. Finally, cognitive test performance is but one layer of mechanistic precision – future work aiming to link brain-based measures to these dimensions may reveal stronger associations than the existing literature has been able to[36,64], providing a more compelling path towards novel pharmacological or stimulation-based interventions.

In conclusion, we used data-driven multiverse analysis of large-scale datasets to demonstrate the factor structure of psychopathology with optimal mechanistic validity. The natural covariance patterns of symptoms corresponded to specific cognitive mechanisms, regardless

of manipulations to factor number retained, combinations of questionnaires or rotation type. Specifically, dimensions that maximally corresponded to distinct cognitive capacities were generated through first-order factor analysis, with oblique rotation, a factor number indexed by the scree plot and a broad set of theoretically-informed psychiatric questionnaires. This demonstrated that computational factor modelling can derive dimensions of psychopathology that are clinically sensible and cognitively meaningful.

METHODS

Participants

In total, nine datasets were included in this study (Figure 1A). Five of these datasets included measures of model-based planning[33–36,50], and four included measures of metacognitive bias [39,40,42,43]. N=1413 individuals were included in the discovery sample, previously described in Gillan et al. (2016) (labelled AMT 1, to distinguish it from other AMT samples). We chose this dataset as the discovery sample, as this study was the first to derive the established transdiagnostic dimensions and is the sample from which the other studies have applied the factor weights to generate independent dimension scores. Individuals in this samples were recruited remotely, online through AMT[34]. The sample had a mean age of 32.97 (SD=10.81), was mostly female (n=823, 58.20%), with a mean IQ of 98.00 (SD=9.55) (Table 1).

Table 1. Sociodemographic Characteristics

	Model-based planning					Metacognitive bias			
Socio-demographic characteristic	AMT 1 (N = 1413)	Smartphone App (N = 1684)[a]	AMT 2 (N = 820)	Student (N = 192)	Patient 1 (N = 862)	AMT 3 (N = 496)	AMT 4 (N = 437)	Patient 2 (N=825)	Prolific (N = 817)
Gender, *N (%)*									
Male	590 (41.80)	483 (28.68)	421 (51.34)	80 (41.67)	178 (20.65)	256 (51.61)	241 (55.15)	173 (20.97)	331 (40.51)
Female	823 (58.20)	1163 (69.06)	399 (48.66)	112 (58.33)	676 (78.42)	240 (48.39)	196 (44.85)	644 (78.06)	486 (59.49)
Other		36 (2.14)			8 (0.93)			8 (0.97)	
Age, *M* (*SD*)	32.97 (10.81)	46.14 (14.70)	34.89 (10.05)	31.89 (12.10)	32.07 (11.06)	35.59 (10.57)	37.54 (10.39)	32.00 (11.02)	25.58 (9.79)
IQ, *M* (*SD*)	98.00 (9.55)		99.00 (9.74)	7.95 (3.29)		7.98 (3.47)	-0.27 (0.79)		
Education, *N (%)*									
Below undergraduate		609 (36.16)			203 (23.55)			196 (23.76)	
Completed undergraduate		678 (40.26)			453 (52.55)			433 (52.48)	
Above undergraduate		395 (23.46)			206 (23.90)			196 (23.76)	

[a] = missing data for 2 participants. AMT = Amazon's Mechanical Turk, IQ = Intelligence quotient, M = mean, SD = standard deviation, N = count.

The other samples with measures of model-based planning included three previously published datasets[33,35,36] and unpublished data from the Precision in Psychiatry (PIP) study[50]. Participants in the Patzelt et al. (2019) dataset were recruited online using (N=820) (labelled AMT 2). We had a slightly lower sample than the sample published in the Patzelt et al. (2019) paper, as we only included those with full model-based planning, questionnaire and sociodemographic information. Participants in the Seow et al. (2021) dataset were from the general population and recruited in-person (N=192). The Donegan et al. (2023) dataset included participants from the Neureka Project, which enrols members from the general public who voluntarily download a smartphone application to contribute to scientific research. We had a slightly larger sample (N=1785) than the sample published in Donegan et al. (2023) paper, as we included additional Neureka users who have completed the model-based planning task since the study publication. The PIP study recruited participants from clinical sites that made referrals for internet-based cognitive behavioural therapy[50]. Among PIP study completers, N=862 completed a model-based planning task[50] (labelled Patient 1), and N=825 completed the metacognitive task[43] (labelled Patient 2). The sample size in this paper (N=825) differs from the size published in Fox, Lee, et al. (2023), as we included the updated sample of individuals that completed the metacognitive task, relevant self-report questionnaires and sociodemographic information. Three of the other datasets with measures of metacognitive bias were pre-existing, published datasets, that recruited participants online through crowdsourcing AMT (Rouault et al., 2018, N=496 (labelled AMT 3) and Seow & Gillan, 2020, N=437 (labelled AMT 4)), or Prolific (Benwell et al., 2022, N=817). The sociodemographic descriptives for all datasets are included in Table 1.

Procedures

Self-report clinical questionnaires. In each study, participants completed 209 items from nine self-report clinical questionnaires that assess a variety of psychiatric symptoms, including depression (Zung Self-Rating Depression Scale)[65], trait anxiety (State Trait Anxiety Inventory)[66], schizotypy (Short Scales for Measuring Schizotypy)[67], impulsivity (Barratt Impulsiveness Scale 11)[68], OCD (Obsessive-Compulsive Inventory-Revised, OCI-R)[69], social anxiety (Liebowitz Social Anxiety Scale)[70], eating disorders (Eating Attitudes Test)[71], apathy (Apathy Evaluation Scale)[72], and alcohol misuse (Alcohol Use Disorders Identification Test)[73]. These questionnaires were chosen based on prior factor analysis in the discovery sample study[34], which demonstrated that these items could be used to generate the established

'Anxious-depression' and 'Compulsivity and Intrusive Thought' dimensions previously associated with model-based planning and metacognitive bias.

Model-based planning tasks. Alternative versions of the reinforcement-learning 'two-step' task[32,74] were used in the five samples that measured model-based planning. In the two-step task used to quantify model-based planning, participants are presented with a series of choices between two stimuli (often represented as abstract symbols or images). Each choice leads to another set of options, creating a two-step decision-making process (Figure 1A). The key feature of the two-step task is that the outcomes of the initial choices are probabilistically associated with different outcomes in the subsequent steps. Participants must learn these associations through trial and error. Model-based learning refers to the ability to learn and utilize an explicit model of the task structure to make decisions. In the context of the two-step task, this involves understanding the probabilistic relationships between choices and outcomes and using this knowledge to plan and select actions that are expected to yield favourable outcomes in the long run. More detailed descriptions of the adapted two-step task versions are included in the original publications for each study[33–36,50].

Metacognitive tasks. Metacognitive bias was measured through adapted versions of a visuo-perceptual decision-making task[39,40,43] or a predictive inference task[42]. In the visuo-perceptual decision-making tasks, participants make a choice as to which of two stimuli contains more dots, and then subsequently rate their confidence in the accuracy of their choice, across multiple trials (Figure 1A). In contrast, the predictive inference task involves participants aiming a particle from the centre of a large circle to hit a target. Participants then rate their confidence that the particle would hit the target (Figure 1A). Detailed descriptions of the tasks can be found in the prior publications from which the datasets were taken[39,40,42,43].

Data Preparation and Analysis

Quantifying model-based planning. The 'two-step' task can be used to assess an individual's model-based planning by using logistic regression analyses with mixed-effect models to predict their choice on the subsequent trial. Specifically, model-based planning was indexed as the interaction effect of reward and transition on their choice stochasticity[33,34,36,50]. Alternatively, model-based planning was calculated as the fit of a computational learning model, in which choices were reflected as the weighted combination of model-free and

model-based planning[35]. For all datasets, deficits in model-based planning were indicated by lower model-based learning values.

Quantifying metacognitive bias. Explicit post-decisional confidence judgements were used to measure metacognition across the four datasets[39,40,42,43]. Metacognitive bias was calculated as the mean confidence reported by participants across experimental trials. Higher mean confidence, relative to within-sample estimates, was used to index 'overconfidence', while 'underconfidence' was indicated as relatively lower mean confidence.

Extracting factor weights from discovery dataset. Using the self-reported clinical questionnaire data from the Gillan et al. (2016) dataset only, we aimed to determine which transdiagnostic dimension had the strongest association with individual differences in model-based planning and metacognitive bias, following manipulations to the number of factors retained, the sets of questionnaire items analysed, and the rotations implemented (Figure 1B).

> ***Variation of the number of factors selected for retention.*** First, we generated a heterogenous correlation matrix of the 209 items from the nine questionnaires using the hector function in the polycor package in R. We conducted maximum likelihood estimation (MLE) factor analysis using the fa function from the psych package in R. We specified that each iteration would run from a single to 10-factor structure, using regression with an oblique rotation. The oblique rotational 'oblimin' was used, consistent with prior factor analysis of this questionnaire set[34]. Factor analysing each factor number from a single to 10-factor structure generated factor weights and loadings for 55 candidate dimensions in total (Figure 1B(i)).
>
> ***Variation in the selection of clinical symptoms.*** Candidate dimensions for multiverse analysis were identified in the Gillan et al. (2016) sample by factor analysing every possible combination of the nine clinical questionnaires. Combinations ranged from each questionnaire on its own, to all nine questionnaires considered together, giving a total of 511 possible combinations (Figure 1B(ii)). We then took each of the standalone combinations of questionnaires and generated a heterogenous correlation matrix for the items included in that combination. The correlation matrices were then factor analysed (MLE in the fa package in R), with the oblique rotational 'oblimin'. Rather than using the scree plot to determine factor number, we set each of the 511

dimensions as having a single factor structure up to a 5-factor structure. Factor analysing each combination with up to 5 potential factor structures generated factor weights and loadings for 7665 candidate dimensions in total Figure 1B(ii).

Higher- versus first-order factor rotation. The heterogenous correlation matrix of all 209 item responses was factor analysed (MLE in the fa package in R), specifying a three-factor structure for analyses. A three-factor structure was chosen following the results from the factor number manipulation analyses. To manipulate factor rotation type, we considered 3 commonly employed rotation types: the oblique rotation, 'oblimin' (used in the prior sections), and orthogonal rotations, 'varimax', and a 'bifactor' solution (Figure 1B(iii)). While the no rotation, oblimin and varimax rotations were generated using the fa package in R, the bifactor solution was calculated using the 'omega' function from the Psych package in R, as per prior research on bifactor modelling in psychiatry[75]. The within-solution correlations between factor scores verified the oblique and orthogonal nature across rotation types (Figure S6). Factor analysing each rotation type with a three-factor structure (with an additional hierarchical general 'p' factor for the bifactor solution) produced weights and loadings for 10 candidate dimensions in total (Figure 1B(iii)).

Testing candidate dimensions across datasets. Considering each manipulation type separately (factor number, combinations of scales and rotation type), we calculated the factor scores for each participant within each of the eight datasets separately. Participants' factor scores were calculated as the sum of the item weight by participants' response to that item, for each factor (i.e., dimension score = sum(item response*weight)) (Figure 1C(i)). This only differed slightly for our bifactor model, as the 'omega' output does not provide factor weights. To account for this, factor scores for our bifactor model were calculated from the factor loadings using the Anderson-Rubin method, in which a least-squares formula is applied to maintain the orthogonality of the general and specific factor scores[76].

We then ran linear regression analysis to determine the association between neurocognitive abilities (model-based planning/metacognitive bias) and factors scores, controlling for age, gender and IQ/education. Age, gender and IQ/education were included as covariates to account for their potential effects on model-based planning and confidence. For each regression analysis, we calculated the effect size of each factor as Cohen's f^2 value[52] (Figure

1C(ii)). For the factor number and rotation type manipulations, our regression models included all the dimensions generated within that factor structure. For example, the regression model with a three-factor structure would be: model-based planning/metacognition ~ factor 1 scores + factor 2 scores + factor 3 scores + age + gender + IQ/education (Figure 1C(ii)). We included all factors within the structure in the same model, as the dimensions across the factor structure are obliquely rotated, meaning that factor scores are related across the structure. Including all factors means the model accounts for the effects of the related factors. This is more important for metacognition, as it has two poles of clinical association. For the interested reader, see supplementary Figure S2 for sensitivity analyses showing individual model results without these controls, where overall effect sizes are smaller. When we tested the impact of the inclusion/exclusion of questionnaires, we only included individual factor scores in each of the 7665 models predicting model-based planning. For metacognitive bias, we included 'Anxious-depression' as an additional covariate to determine which factors explained the most variance in overconfidence. When we were interested in factors that predicted underconfidence, we reran the models and included 'Compulsivity and Intrusive Thought' instead of 'Anxious-depression' as a covariate.

Following this, we took each dataset with a measure of model-based planning and calculated the weighted (by sample size) average Cohen's f^2 for each factor across the samples, consistent with a meta-analytic approach. We then repeated this separately for each dataset that had a measure of metacognitive bias to get the weighted average Cohen's f^2 when mean confidence was the dependent variable. The weighted average Cohen's f^2 were used to determine which factor performed best when predicting individual differences in model-based planning and metacognitive bias (i.e., which factors were the winning solutions). For model-based planning, we were specifically interested in deficits (lowest Cohen's f^2 values). For metacognitive bias, we were interested in both directions (highest and lowest Cohen's f^2 values, measuring over- and under-confidence respectively) (Figure 1C(iii)).

PLS regression. We used partial least squared regression to identify transdiagnostic latent dimensions, comprised of the 209 questionnaire items, which are linked to individual differences in cognitive outcomes. The AMT 1 sample (N=1413)[34] was used as discovery datasets to generate the PLS component for model-based planning, as this sample was also used to discover the factor structures from unsupervised factor analysis. For metacognitive bias, the AMT 3 sample (N=496)[40] were used as discovery datasets to generate the PLS

component. The Rouault et al. (2018) sample was chosen because it is the largest dataset available with a general population sample and data on metacognition and all covariates (age, gender and IQ). To avoid model overfitting, we split each discovery dataset into training and test sets, comprising of 75% and 25% of the data, respectively[77]. To identify the optimal number of components and equivalent item loadings for components within the training set, we used a 10-fold cross-validation procedure, fitting the model on 90% and testing performance on the left-out 10% of the data. The mean squared error of the model's predictions was then averaged across test folds to provide an index of the model's predictive accuracy with different numbers of components. Within the training set, the optimal models for model-based planning (RMSE=0.99), overconfidence (RMSE=0.97) and underconfidence (RMSE=0.98) all contained a single component. The item loadings for components were used to generate participants' dimension scores (dimension score = sum(item response x loading)). We then evaluated if component scores were significantly associated with cognitive outcomes within the training and test sets separately, before testing model performance out of sample.

Acknowledgements

This work was funded by a fellowship awarded to Claire M Gillan from Science Foundation Ireland's Frontiers for the Future Scheme (19/FFP/6418). Claire M Gillan holds additional funding from a European Research Council (ERC) Starting Grant (ERC-H2020-HABIT). The PhD studentship of Celine A Fox is funded by the Government of Ireland Postgraduate Scholarship Programme (GOIPG/2020/662). The authors thank all the participants for their involvement in this study. We thank the researchers that kindly shared their data with us, including Steve Fleming, Marion Rouault, Sam Gershman, Robin Ince, and Edward Patzelt.

Competing interests

The authors declare no competing interests.

REFERENCES

1. Kotov, R., Krueger, R. F. & Watson, D. A paradigm shift in psychiatric classification: the Hierarchical Taxonomy Of Psychopathology (HiTOP). *World Psychiatry* **17**, 24–25 (2018).
2. Dalgleish, T., Black, M., Johnston, D. & Bevan, A. Transdiagnostic approaches to mental health problems: Current status and future directions. *J. Consult. Clin. Psychol.* **88**, 179–195 (2020).
3. American Psychiatric Association. (2022). Diagnostic and statistical manual of mental disorders (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787
4. Fried, E. I., Flake, J. K. & Robinaugh, D. J. Revisiting the theoretical and methodological foundations of depression measurement. *Nat. Rev. Psychol.* **1**, 358–368 (2022).
5. Haslam, N. Categorical versus dimensional models of mental disorder: the taxometric evidence. *Australian & New Zealand Journal of Psychiatry* **37**, 696–704 (2003).
6. Fried, E. I. & Nesse, R. M. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. *J. Affect. Disord.* **172**, 96–102 (2015).
7. Zimmerman, M., Ellison, W., Young, D., Chelminski, I. & Dalrymple, K. How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Compr. Psychiatry* **56**, 29–34 (2015).
8. Forbes, M. K. *et al.* Elemental psychopathology: distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychol. Med.* **54**, 886–894 (2024).
9. Kessler, R. C., Chiu, W. T., Demler, O. & Walters, E. E. prevalence, severity, and comorbidity of 12-Month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch. Gen. Psychiatry* **62**, 617 (2005).

10. McGrath, J. J. *et al.* Comorbidity within mental disorders: a comprehensive analysis based on 145 990 survey respondents from 27 countries. *Epidemiol. Psychiatr. Sci.* **29**, e153 (2020).
11. Abi-Dargham, A. *et al.* Candidate biomarkers in psychiatric disorders: state of the field. *World Psychiatry* **22**, 236–262 (2023).
12. Strawbridge, R., Young, A. H. & Cleare, A. J. Biomarkers for depression: recent insights, current challenges and future prospects. *Neuropsychiatr. Dis. Treat.* **Volume 13**, 1245–1262 (2017).
13. Tiego, J. *et al.* Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. *Nat. Ment. Health* **1**, 304–315 (2023).
14. Abramovitch, A., Short, T. & Schweiger, A. The C Factor: Cognitive dysfunction as a transdiagnostic dimension in psychopathology. *Clin. Psychol. Rev.* **86**, 102007 (2021).
15. East-Richard, C., R. -Mercier, A., Nadeau, D. & Cellard, C. Transdiagnostic neurocognitive deficits in psychiatry: A review of meta-analyses. *Can. Psychol. Psychol. Can.* **61**, 190–214 (2020).
16. Smoller, J. W. *et al.* Psychiatric genetics and the structure of psychopathology. *Mol. Psychiatry* **24**, 409–420 (2019).
17. Marek, S. *et al.* Reproducible brain-wide association studies require thousands of individuals. *Nature* **603**, 654–660 (2022).
18. Sha, Z., Wager, T. D., Mechelli, A. & He, Y. Common dysfunction of large-scale neurocognitive networks across psychiatric disorders. *Biol. Psychiatry* **85**, 379–388 (2019).
19. Winter, N. R. *et al.* Quantifying deviations of brain structure and function in major depressive disorder across neuroimaging modalities. *JAMA Psychiatry* **79**, 879 (2022).

20. Etkin, A. A reckoning and research agenda for neuroimaging in psychiatry. *Am. J. Psychiatry* **176**, 507–511 (2019).
21. Huys, Q. J. M., Browning, M., Paulus, M. P. & Frank, M. J. Advances in the computational understanding of mental illness. *Neuropsychopharmacology* **46**, 3–19 (2021).
22. Bzdok, D. & Meyer-Lindenberg, A. machine learning for precision psychiatry: opportunities and challenges. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 223–230 (2018).
23. Eaton, N. R. *et al.* A review of approaches and models in psychopathology conceptualization research. *Nat. Rev. Psychol.* **2**, 622–636 (2023).
24. Insel, T. *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* **167**, 748–751 (2010).
25. Kotov, R. *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J. Abnorm. Psychol.* **126**, 454–477 (2017).
26. Wise, T., Robinson, O. & Gillan, C. Identifying transdiagnostic mechanisms in mental health using computational factor modeling. *Biol. Psychiatry* **0**, (2023).
27. Wittchen, H. & Beesdo-Baum, K. "Throwing out the baby with the bathwater"? Conceptual and methodological limitations of the HiTOP approach. *World Psychiatry* **17**, 298–299 (2018).
28. Strauss, G. P. & Cohen, A. S. A transdiagnostic review of negative symptom phenomenology and etiology. *Schizophr. Bull.* **43**, 712–719 (2017).
29. Haeffel, G. J. *et al.* Folk classification and factor rotations: whales, sharks, and the problems with the Hierarchical Taxonomy of Psychopathology (HiTOP). *Clin. Psychol. Sci.* **10**, 259–278 (2022).
30. Borsboom, D. A network theory of mental disorders. *World Psychiatry* **16**, 5–13 (2017).

31. Caspi, A. & Moffitt, T. E. All for one and one for all: mental disorders in one dimension. *Am. J. Psychiatry* **175**, 831–844 (2018).
32. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-based influences on humans' choices and striatal prediction errors. *Neuron* **69**, 1204–1215 (2011).
33. Donegan, K. R. *et al.* Using smartphones to optimise and scale-up the assessment of model-based planning. *Commun Psychol* **1**, 31 (2023).
34. Gillan, C. M., Kosinski, M., Whelan, R., Phelps, E. A. & Daw, N. D. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* **5**, e11305 (2016).
35. Patzelt, E. H., Kool, W., Millner, A. J. & Gershman, S. J. Incentives boost model-based control across a range of severity on several psychiatric constructs. *Biol. Psychiatry* **85**, 425–433 (2019).
36. Seow, T. X. F. *et al.* Model-based planning deficits in compulsivity are linked to faulty neural representations of task structure. *J. Neurosci.* **41**, 6539–6550 (2021).
37. Fleming, S. M. & Lau, H. C. How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443 (2014).
38. Hoven, M. *et al.* Abnormalities of confidence in psychiatry: an overview and future perspectives. *Transl. Psychiatry* **9**, 268 (2019).
39. Benwell, C. S. Y., Mohr, G., Wallberg, J., Kouadio, A. & Ince, R. A. A. Psychiatrically relevant signatures of domain-general decision-making and metacognition in the general population. *Npj Ment. Health Res.* **1**, 1–17 (2022).
40. Rouault, M., Seow, T., Gillan, C. M. & Fleming, S. M. Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biol. Psychiatry* **84**, 443–451 (2018).

41. Seow, T. X. F., Fleming, S. M. & Hauser, T. U. Metacognitive biases in anxious-depression and compulsivity extend across perception and memory. Preprint at https://doi.org/10.31234/osf.io/avyph (2024).
42. Seow, T. X. F. & Gillan, C. M. Transdiagnostic Phenotyping Reveals a Host of Metacognitive Deficits Implicated in Compulsivity. *Sci. Rep.* **10**, 2883 (2020).
43. Fox, C. A. *et al.* An observational treatment study of metacognition in anxious-depression. *eLife* **12**, (2023).
44. Fox, C. A. *et al.* Reliable, rapid, and remote measurement of metacognitive bias. Preprint at https://osf.io/c5abx (2024).
45. Wittchen, H. *et al.* The structure of mental disorders re-examined: Is it developmentally stable and robust against additions? *Int. J. Methods Psychiatr. Res.* **18**, 189–203 (2009).
46. Krueger, R. F. The structure of common mental disorders. *Arch. Gen. Psychiatry* **56**, 921 (1999).
47. Caspi, A. *et al.* The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
48. Caspi, A., Houts, R. M., Fisher, H. L., Danese, A. & Moffitt, T. E. The general factor of psychopathology (p): choosing among competing models and interpreting p. *Clin. Psychol. Sci.* **12**, 53–82 (2024).
49. Simonsohn, U., Simmons, J. P. & Nelson, L. D. Specification curve analysis. *Nat. Hum. Behav.* **4**, 1208–1214 (2020).
50. Lee, C. T. *et al.* The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation. *BMC Psychiatry* **23**, 25 (2023).
51. Gorsuch, R. & Nelson, J. CNG screen test: an objective procedure for determining the number of factors. *Present. Annu. Meet. Soc. Multivar. Exp. Psychol.* (1981).

52. Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D. & Mermelstein, R. J. A practical guide to calculating cohen's f2, a measure of local effect size, from PROC MIXED. *Front. Psychol.* **3**, (2012).
53. Kapur, S., Phillips, A. G. & Insel, T. R. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol. Psychiatry* **17**, 1174–1179 (2012).
54. Stephan, K. E. *et al.* Charting the landscape of priority problems in psychiatry, part 1: classification and diagnosis. *Lancet Psychiatry* **3**, 77–83 (2016).
55. Haeffel, G. J. *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP) Is not an improvement over the *DSM*. *Clin. Psychol. Sci.* **10**, 285–290 (2022).
56. Aristodemou, M. E. & Fried, E. I. Common factors and interpretation of the p factor of psychopathology. *J. Am. Acad. Child Adolesc. Psychiatry* **59**, 465–466 (2020).
57. Fried, E. I., Greene, A. L. & Eaton, N. R. The p factor is the sum of its parts, for now. *World Psychiatry* **20**, 69–70 (2021).
58. Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J. & Wiernik, B. M. Appropriate use of bifactor analysis in psychopathology research: appreciating benefits and limitations. *Biol. Psychiatry* **88**, 18–27 (2020).
59. Luciano, J. V., Sanabria-Mazo, J. P., Feliu-Soler, A. & Forero, C. G. The pros and cons of bifactor models for testing dimensionality and psychopathological models: A commentary on the manuscript "A systematic review and meta-analytic factor analysis of the depression anxiety stress scales". *Clin. Psychol. Sci. Pract.* **27**, (2020).
60. Watts, A.L., Greene, A.L., Bonifay, W., Fried. E. I. A critical evaluation of the p-factor literature. *Nat Rev Psychol* **3**, (2024).
61. Karvelis, P., Paulus, M. P. & Diaconescu, A. O. Individual differences in computational psychiatry: A review of current challenges. *Neurosci. Biobehav. Rev.* **148**, 105137 (2023).

62. Kendler, K. S. Explanatory models for psychiatric illness. *Am. J. Psychiatry* **165**, 695–702 (2008).
63. Keidel, K., Lu, X., Suzuki, S., Murawski, C. & Ettinger, U. Association of temporal discounting with transdiagnostic symptom dimensions. *Npj Ment. Health Res.* **3**, 13 (2024).
64. Xia, C. H. *et al.* Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat. Commun.* **9**, 3003 (2018).
65. Zung, W. W. A self-rating depression scale. *Arch. Gen. Psychiatry* **12**, 63–70 (1965).
66. Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P. & Jacobs, G. *Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2). Palo Alto, CA: Consulting Psychologists Press;* vol. IV (1983).
67. Mason, O., Linney, Y. & Claridge, G. Short scales for measuring schizotypy. *Schizophr. Res.* **78**, 293–296 (2005).
68. Patton, J. H., Stanford, M. S. & Barratt, E. S. Factor structure of the Barratt impulsiveness scale. *J. Clin. Psychol.* **51**, 768–774 (1995).
69. Foa, E. B. *et al.* The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychol. Assess.* **14**, 485–496 (2002).
70. Liebowitz, M. R. Social phobia. *Mod. Probl. Pharmacopsychiatry* **22**, 141–173 (1987).
71. Garner, D. M., Olmsted, M. P., Bohr, Y. & Garfinkel, P. E. The eating attitudes test: psychometric features and clinical correlates. *Psychol. Med.* **12**, 871–878 (1982).
72. Marin, R. S., Biedrzycki, R. C. & Firinciogullari, S. Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Res.* **38**, 143–162 (1991).
73. Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R. & Grant, M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO

Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addict. Abingdon Engl.* **88**, 791–804 (1993).

74. Kool, W., Gershman, S. J. & Cushman, F. A. Cost-benefit arbitration between multiple reinforcement-learning systems. *Psychol. Sci.* **28**, 1321–1333 (2017).
75. Gagne, C., Zika, O., Dayan, P. & Bishop, S. J. Impaired adaptation of learning to contingency volatility in internalizing psychopathology. *eLife* **9**, e61387 (2020).
76. Anderson, T. & Rubin, H. Statistical inference in factor analysis. *Proc. Third Berkeley Symp. Math. Stat. Probab.* 111–150 (1956).
77. Wise, T. & Dolan, R. J. Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nat. Commun.* **11**, 4179 (2020).