npj mental health research

www.nature.com/npjmentalhealth

ARTICLE OPEN

Psychiatrically relevant signatures of domain-general decision-making and metacognition in the general population

Christopher S. Y. Benwell[1]✉, Greta Mohr[2], Jana Wallberg[1], Aya Kouadio[1] and Robin A. A. Ince[2]

Human behaviours are guided by how confident we feel in our abilities. When confidence does not reflect objective performance, this can impact critical adaptive functions and impair life quality. Distorted decision-making and confidence have been associated with mental health problems. Here, utilising advances in computational and transdiagnostic psychiatry, we sought to map relationships between psychopathology and both decision-making and confidence in the general population across two online studies (*N*'s = 344 and 473, respectively). The results revealed dissociable decision-making and confidence signatures related to distinct symptom dimensions. A dimension characterised by compulsivity and intrusive thoughts was found to be associated with reduced objective accuracy but, paradoxically, increased absolute confidence, whereas a dimension characterized by anxiety and depression was associated with systematically low confidence in the absence of impairments in objective accuracy. These relationships replicated across both studies and distinct cognitive domains (perception and general knowledge), suggesting that they are reliable and domain general. Additionally, whereas Big-5 personality traits also predicted objective task performance, only symptom dimensions related to subjective confidence. Domain-general signatures of decision-making and metacognition characterise distinct psychological dispositions and psychopathology in the general population and implicate confidence as a central component of mental health.

npj Mental Health Research (2022)1:10 ; https://doi.org/10.1038/s44184-022-00009-4

INTRODUCTION

When making decisions in everyday life, immediate external feedback is not always available to inform us of the utility of our choices. In the absence of external feedback, we often rely on an internally generated sense of confidence. This confidence informs metacognitive evaluations of our decisions, actions, and abilities. Though confidence and objective accuracy/utility are usually correlated, the ability to self-evaluate is often suboptimal[1] and this can impact diverse cognitive functions such as learning, decision-making and error-monitoring[2–5]. For example, if we know that we have performed poorly on a given task, we are likely to alter our behaviour to improve future performance[6,7]. Conversely, if we lack insight, we risk persevering with damaging choices/behaviours. Indeed, deficits in metacognitive insight have been shown to contribute to impaired life quality in various neurological and psychiatric disorders[8,9]. However, the psychological determinants of metacognitive ability remain poorly understood.

Consistent relationships have been identified between metacognition and clinically relevant psychiatric symptoms, particularly general under-confidence in major depression[10–12], under-confidence in memory in obsessive-compulsive disorder (OCD)[13–15], and impaired metacognitive insight in schizophrenia[16–19]. However, suboptimal self-evaluation is not only restricted to clinical samples[1,20,21]. Recent studies suggest that metacognitive distortions, such as under- and over-confidence, are associated with specific personality traits[22] and belief systems[23] in the general population, as well as subclinical psychopathology[24–26]. For instance, symptom dimensions cutting across traditional diagnostic categories have been found to correlate with metacognitive performance in general population samples: an 'anxious-depression' (AD) dimension and a 'compulsive behaviour and intrusive thought' (CIT) dimension. Those scoring highly for CIT displayed overconfidence in perceptual decisions, reduced sensitivity of confidence judgements to objective evidence and reduced metacognitive insight[26–29], whereas those scoring highly for AD showed low overall confidence but increased metacognitive insight[26]. These symptom-specific alterations of self-evaluation may represent enduring psychological phenotypes of psychopathology. However, metacognitive performance is governed by both domain-specific and domain-general mechanisms[30,31] and the degree to which metacognitive abnormalities are generalisable to cognitive domains outside of perception remains unknown.

Compared to psychopathology, fewer studies have investigated relationships between metacognition and personality traits. Due to the close links between personality traits and symptoms, it is possible that personality may play a key role in relationships between metacognition and psychopathology. Overall confidence has been positively associated with extraversion[22,32] and negatively associated with neuroticism[33]. Extraversion shows both positive and negative relationships with psychopathology[34], negatively predicting internalizing symptoms characterised by social/interpersonal dysfunction[35] and/or depression and anxiety[34,36], but positively predicting externalizing symptoms characterised by exhibitionism and mania[34]. Neuroticism positively predicts many forms of psychopathology, particularly anxiety and depression[37–40]. In the current study, we sought to quantify and dissociate the degree to which dimensions of both psychopathology and personality are predictive of metacognitive performance.

We adopted a computational modelling approach to measure 1st-order decision-making and metacognition across cognitive domains. This allowed for relationships with personality and psychopathology to be grounded in quantitative model-based measures[41,42]. This is important because confidence is influenced by multiple latent processes including metacognitive sensitivity

[1]Division of Psychology, School of Humanities, Social Sciences and Law, University of Dundee, Dundee, UK. [2]School of Psychology and Neuroscience, University of Glasgow, Glasgow, UK. ✉email: c.benwell@dundee.ac.uk

(the degree to which confidence dissociates between correct and incorrect decisions) and metacognitive bias (the absolute level of confidence experienced regardless of objective accuracy), as well as by 1st-order task performance itself[41,43]: any (or all) of which may be related to psychological dispositions. In addition to metacognitive abnormalities, some previous studies have found psychiatrically relevant 1st-order decision-making[27–29,44–46] and/or learning[47–51] deficits, whilst others have not[26,52,53]. Elucidation and dissociation of 1st- and 2nd-order (metacognitive) decision-making abnormalities represent key steps towards an accurate mapping of the deficits underlying core symptoms of psychopathology.

Here, across two separate online studies (N's = 344 and 473, respectively), we investigated relationships between both 1st-order and metacognitive decision-making parameters, self-reported psychopathology (utilising both classic categorical and transdiagnostic approaches) and personality traits. We replicated relationships between psychiatric symptomology and decision parameters from a perceptual task across both studies. In the 2nd study we also employed a knowledge-based task to test whether the relationships are domain-general, and hence likely to have a more pervasive influence in everyday life. Finally, we investigated the degree to which personality traits influenced 1st- and 2nd-order decision-making independently of symptoms of psychopathology.

METHODS

Participants

Participants were recruited online using the **Prolific** (www.prolific.co) and **Sona Systems** (https://www.sona-systems.com/) recruitment platforms (experiment 1: 393 participants, 16–73 years old (M = 25.32, SD = 10.83); experiment 2: 534 participants, 18–70 years old (M = 25.42, SD = 9.17)). Some participants (N = 374) were paid £7.50 for their time, whilst others received undergraduate course credits (N = 553). No *a priori* power analysis was performed for experiment 1, with the sample size being based on those employed in relevant previous studies[26,29]. However, to ensure adequate statistical power to replicate the effects observed in experiment 1, we conducted an *a priori* power analysis (using G*Power 3.1.9.7) to determine the appropriate sample size for experiment 2. We based the power analysis on the lowest significant effect size observed for a single symptom dimension across the symptom dimension-behaviour relationships in experiment 1 (Compulsive Behaviour and Intrusive Thought (CIT)-accuracy (***d'***) relationship: f^2 = 0.02). The power analysis indicated that 395 participants would be required to achieve 80% statistical power to detect such an effect. Hence, the total experiment 2 sample size (534) allowed for adequate statistical power to be maintained after data exclusion.

Due to predefined exclusion criteria (explained below), 49 participants were excluded from the experiment 1 analysis, leaving a total number of 344 participants (253 female/91 male, aged from 18 to 73 years (M = 25.35, SD = 10.5)), and 61 participants were excluded from experiment 2, leaving a total number of 473 participants (233 female/240 male aged from 18 to 65 years (M = 25.75, SD = 9.24)). A *post hoc* power analysis indicated that with the final sample (473) in experiment 2, we achieved 86% statistical power to detect an effect equal to the smallest significant effect size in experiment 1 (f^2 = 0.02). The only demographic information collected from participants was age and gender, thereby data anonymity was maintained. Both studies received ethical approval from the University of Dundee Research Ethics Committee and all participants provided informed consent.

Perceptual decision task

The perceptual decision task involved 2-alternative forced-choice (2-AFC) numerosity discrimination judgements with confidence ratings and was chosen to replicate Rouault et al., (2018). The perceptual decision task was employed in both experiments 1 and 2. Figure 1a shows a schematic of the trial procedure. On each trial, a black cross appeared at the centre of the screen for 1000 ms. This was followed by two black boxes, one on the left and the other on the right of the screen, which both contained numerous white dots. These were simultaneously presented for 400 ms. Participants were then asked to decide which box contained a larger number of dots by pressing the 'w' key for the box on the left or the 'e' key for the box on the right. One box (the reference box) constantly contained 272 dots (out of 544 possible dot locations), while the other box contained an increased or reduced number of dots ranging from either −72 to +72 dots (n = 79 in experiment 1) or −64 to +64 dots (n = 265 in experiment 1 and all participants in experiment 2) in increments of 8 dots in comparison to the reference box (including an identical condition). The location (left or right) of the reference box varied pseudo-randomly across trials and within each of the difficulty levels. The order of stimulus presentation was randomly generated for each participant. There was no time limit for the response and participants were not given feedback on whether their response was correct. After providing a response, participants were asked to rate how confident they were in their decision on a scale of 1 (not confident/guessing) to 6 (certain). There was no time limit for the confidence rating. Note that 82 participants in experiment 1 completed 152 trials over 2 blocks (8 trials per 19 conditions, 76 trials per block including ±72 stimuli), whereas the remaining 267 participants in experiment 1, and all participants in experiment 2, completed 136 trials over 2 blocks (8 trials per 17 conditions, 68 trials per block). Only the conditions that were shared by all participants were included in the analyses (−64 to +64 numerosity difference conditions). Participants could take a self-paced break between blocks. Before starting the task, participants completed ten practice trials in which only the easiest stimuli were presented (64 or 72 dot difference). The practice trials were identical to the experimental trials except that feedback (a green tick or red cross) was provided (indicating whether the response was correct or incorrect). Two further practice trials were used to familiarise participants with the confidence rating scale in which they were instructed how to respond if they were confident or not confident.

Knowledge decision task

To investigate whether the psychiatric symptom – decision-making relationships generalised to other cognitive domains, in experiment 2 we employed an additional 2-AFC task which tested prior knowledge of generally known quantities: national populations[54]. Figure 3a shows a schematic of the trial procedure. On each trial, a black cross appeared at the centre of the screen for 1000 ms. This was followed by the names of two countries and the participants were required to indicate which of the two has the largest human population by selecting the corresponding button on the screen. The country names remained on the screen until the response but if the participant did not respond within 10 s, then the trial was recorded as 'no response'. After each response, the participant was asked to rate how confident they were in their decision on a scale of 1 (not confident/guessing) to 6 (certain). No feedback about participants' decision-making was provided during the experimental trials. There was no time limit for the confidence rating.

The national populations for creating the stimuli were downloaded from The World Bank ('https://data.worldbank.org/indicator/SP.POP.TOTL') in December 2019. Eight different evidence discriminability 'bins' were created by grouping country pairs with similar population log ratios (bins created based on $\log_{10}$ (Country A Population/Country B Population)). The log ratio bins amounted to the following, ranging from least to most discriminable: bin 1 ($\log_{10}$ ratio = 0–0.225), bin 2 = (0.225–0.45), bin 3 (0.45–0.675), bin 4 = (0.675–0.9), bin 5 (0.9–1.125), bin 6 = (1.125–1.35), bin 7 (1.35–1.575), bin 8 = (1.575–1.8). Each bin included 18 different country pairs (full task stimuli available at

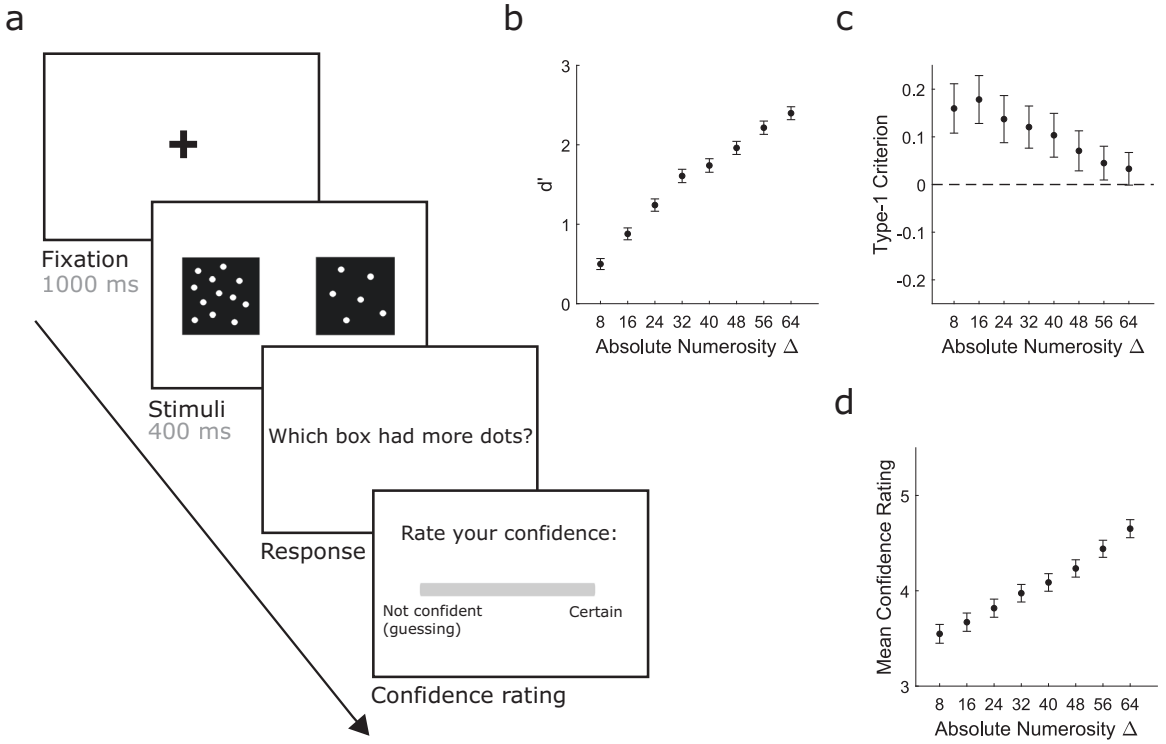

Fig. 1 Perceptual decision-making task and behaviour in experiment 1 ($n = 344$). a Perceptual task. On each trial, participants judged which box (left or right) contained the higher number of dots and provided a confidence rating in each decision (scale of 1–6, where 1 represented "not confident (guessing)" and 6 represented "certain"). **b** As expected, group-averaged ***d'*** increased as a function of absolute numerosity difference. **c** Group-averaged type-1 ***c'*** were biased towards 'left more numerous' responses across all evidence levels and were significantly different to 0 for all numerosity differences up to 56 dots (all p's < .014), but not for the easiest 64 dot difference condition ($p = .06$). This leftward bias may reflect either the pseudoneglect phenomenon, whereby neurotypical individuals tend to judge stimuli presented in the left visual field as more salient than comparable stimuli in the right visual field[75–77], and/or a motor-response bias. **d** Group-averaged overall mean confidence ratings increased as a function of evidence strength. All error bars reflect 95% confidence intervals for the mean.

https://osf.io/s3cth/). The location (left or right) of the most populous country varied pseudo-randomly across trials but was counterbalanced within each of the discriminability bins (i.e., same proportion of 'left' larger and 'right' larger stimuli). The order of stimulus presentation was randomly generated for each participant. Participants completed 144 trials over 2 blocks (9 trials per 16 log ratio conditions, 72 trials per block) and could take a self-paced break between blocks. Before starting the task, 10 practice trials were completed in which only examples of the most discriminable stimuli were presented (bin 8). The practice trials were identical to the experimental trials except that feedback (a green tick or red cross) was provided (indicating whether the response was correct or incorrect).

Modelling type-1 and type-2 sensitivity and bias

We modelled 1st-order decisions and confidence ratings on both tasks within an extended signal detection theory (SDT) framework. This model extends the classic SDT approach[55] to quantify latent parameters (i.e., sensitivity and bias) contributing to both type-1 and type-2 decisions. Type-1 sensitivity (***d'***) indexes how accurate the participant's 1st-order task decisions are. ***Meta-d'*** characterises type-2 (metacognitive) sensitivity as the value of ***d'*** that a metacognitively optimal observer, with the same type-1 ***criterion***, would have required to produce the observed type-2 (confidence) data. An individual with optimal metacognitive sensitivity will always be more confident when correct and less confident when incorrect. For a metacognitively ideal observer (a person who is rating confidence using the maximum possible metacognitive sensitivity), ***meta-d'*** should equal ***d'***. Importantly, we can therefore define the level of metacognitive insight/efficiency, controlling for 1st-order performance, as the value of ***meta-d'*** relative to ***d'*** (***meta-d'/d'***). A ***meta-d'/d'*** value of 1 indicates theoretically ideal metacognitive insight. A value below 1 indicates that evidence available for the type-1 decision is lost when making metacognitive judgements (type-2 decision), whereas a value above 1 indicates that more evidence is available for the type-2 decision than for the type-1 decision[41]. Note that we employed the ***meta-d'/d'*** measure of metacognitive efficiency, rather than the alternative ***meta-d'-d'*** measure, because it has been shown to better isolate metacognitive sensitivity from 1st-order accuracy[43].

The confidence criteria (***type-2 c'***) represent type-2 bias calculated within the meta-d' framework: the tendency to give high or low confidence ratings regardless of evidence strength. We calculated the absolute distance between ***type-2 c'*** and ***type-1 c'*** (|***type-2 c'*** - ***type-1 c'***|) to isolate confidence bias from perceptual response bias[56]. Lower confidence criteria (|***type-2 c'*** -***type-1 c'***|) values indicate an overall bias in favour of higher confidence ratings and higher values indicate a bias in favour of low confidence ratings (i.e., confidence criteria are inversely related to mean absolute confidence ratings). Confidence criteria values were calculated separately for each of the possible type-1 responses (i.e., 'left' or 'right' more numerous/higher population judgements in the perceptual and general knowledge tasks respectively) and for each of N-1 confidence ratings available to choose from (6 in the current experiment). To streamline the analysis, we averaged over the 5 |***type-2 c'*** - ***type-1 c'***| values for each response ('left' or 'right') separately and then averaged over the resulting 'left' and 'right' mean criteria to gain a single overall confidence criterion estimate.

All measures were calculated using individual participant fits (*fit_meta_d_mcmc* function) within the "Hmeta-d" toolbox[57] (https://github.com/metacoglab/HMeta-d) in Matlab (Mathworks, USA). The input parameters for the model fits were as follows:

```
mcmc_params.response_conditional = 0;
mcmc_params.estimate_dprime = 0;
mcmc_params.nchains = 3;
mcmc_params.nburnin = 1000;
mcmc_params.nsamples = 10000;
mcmc_params.nthin = 1;
mcmc_params.doparallel = 0;
mcmc_params.dic = 1;
```

The scripts for running the fits can be found at https://osf.io/s3cth/. It is important to note that the model-based measures were calculated collapsed across all discriminability levels from each participant independently (NOT within a hierarchical model) for regressions with self-reported psychiatric symptoms and personality traits (Figs. 2, 6 & 7). However, to test the reliability of the symptom-metacognitive efficiency relationships, we also employed alternative hierarchical analysis approaches which incorporated group-level prior densities when estimating metacognitive efficiency[57,58] (see 'Statistical Analyses' section below and Supplementary Figs. 7 and 8).

Self-report psychometric questionnaires

Each participant in both experiments completed a battery of nine mental health questionnaires which assessed symptomology across a range of disorders. Symptoms of depression were measured using the Zung Self-Rating Depression Scale[59]. Obsessive-Compulsive symptoms were measured using the Obsessive-Compulsive Inventory-Revised[60]. Trait anxiety was measured using the State-Trait Anxiety Inventory Form Y-2[61]. Alcohol addiction was measured using the Alcohol Use Disorder Identification Test (AUDIT)[62]. Apathy was measured using the Apathy Evaluation Scale[63]. Eating disorder symptomology was measured using the Eating Attitudes Test (EAT-26)[64]. Impulsivity was measured using the Barratt Impulsivity Scale (BIS-11)[65]. Schizotypy was measured using the Short Scales for Measuring Schizotypy[66]. Social anxiety was measured using the Liebowitz Social Anxiety Scale which contains 24-items[67]. These questionnaires were chosen to allow us to investigate the three underlying transdiagnostic symptom dimensions identified by[47] and replicated by[31]. In addition to the psychiatric symptom questionnaires, participants in experiment 2 also completed the Big Five Inventory[68].

Transdiagnostic symptom dimensions

Using the same psychiatric symptom questionnaires, Gillan et al., (2016)[47] conducted an exploratory factor analysis (FA) on data collected in a large sample ($n = 1413$). They found that the items from all 9 mental health questionnaires ($n = 209$ items) clustered around three latent 'factors' which they termed 'Anxious-Depression', 'Compulsive Behaviour and Intrusive Thoughts' and 'Social Withdrawal' based on the individual items loading most strongly on each respective factor. The 'Anxious-Depression' factor was most heavily weighted by items from the Generalised Anxiety, Depression, Apathy, and Impulsivity questionnaires (see Gillan et al., 2016). The 'Compulsive Behavior and Intrusive Thought' factor was most heavily weighted by items from the OCD, Eating Disorders, Alcoholism and Schizotypy questionnaires. Lastly, the 'Social Withdrawal' factor had the highest average loadings from the Social Anxiety questionnaire. These factors have subsequently been replicated in an independent sample[26]. We replicated the FA performed by Rouault et al., (2018)[26] to test whether the previously observed three transdiagnostic symptom dimensions[26,47] were replicated in our data ($N = 817$ participants from both experiments 1 and 2). The analysis was conducted on the 209 individual questionnaire items using the *fa()* function from the Psych package in R, with an oblique (oblimin) rotation and maximum likelihood estimation. For the Liebowitz Social Anxiety Scale (LSAS), the average of the avoidance and fear/anxiety answers of each item was taken. In line with previous studies[26,47], a 3-factor latent structure was found to provide the most parsimonious explanation for the item-level responses. Supplementary Fig. 3A plots correlations between the item weights from the FA performed on the current data and those of Gillan et al., (2016)[47] for each factor. Supplementary Fig. 3B plots correlations between the individual participant scores calculated using the item weights from our FA and those of Gillan et al., (2016) for each factor.

Due to the larger sample size used to conduct their factor analysis, we applied the weights from Gillan et al., (2016) to derive scores for the three symptom dimensions for the main analyses. First, the raw responses for each item were *z*-scored across participants, the individual item z-scores within each participant were then multiplied by their corresponding factor weights and the resulting products were summed across all items for each factor. Finally, the factor sums were *z*-scored across participants in preparation for statistical analyses. Note that the results were also reproduced using the item weights from the FA performed on the current data. The R script for running the FA can be found at https://osf.io/s3cth/.

Procedure

Both experiments were conducted online via the Gorilla experiment platform[69]. The experiments could only be completed on either a laptop, tablet, or personal computer (and not on a mobile phone) to facilitate a more optimal screen size for the visual perception task. After clicking an online link and providing informed consent, participants were first asked to provide demographic information of age and gender assigned at birth. The participants then completed the questionnaires and task(s) in a randomised order. The entire experimental session took between 40 min and 1 h for both experiments.

Exclusion criteria

Several predefined exclusion criteria were applied to the data from both experiments to ensure acceptable data quality. Across both studies, ~23% of participants were excluded based on the criteria, leaving 344 participants for experiment 1 and 473 participants for experiment 2.

Participants who met any one or more of the following criteria in experiment 1 were excluded from all analyses:

1. Did not provide gender information ($n = 5$, 1.28%).
2. Below- or near-chance perceptual decision task performance (overall accuracy < 55%) ($n = 9$, 2.29%).
3. Below the age of 18 ($n = 11$, 2.8%).
4. Incorrect response to a 'catch' item employed as an attention check ($n = 12$, 3.05%). The 'catch' item was embedded within the Zung Depression Scale and read as follows: "If you are paying attention, please select 'Good part of the time' for this answer".
5. Used the same single confidence rating across all trials of the perceptual decision task ($n = 1$, 0.25%).
6. A metacognitive efficiency (***meta-d'/d'***) ratio below 0 on the perceptual decision task ($n = 13$, 3.31%). A negative metacognitive efficiency score can occur when type-1 accuracy is around chance level and/or the participant is not using the confidence scale as expected (i.e. repeating a single confidence rating on the vast majority of trials or randomly selecting confidence ratings[70]).

Based on these criteria, a total of 49 participants (12.5%) were excluded from experiment 1.

The exclusion criteria for experiment 2 included all of those employed in experiment 1 plus additional criteria based on

knowledge task performance. Again, any participants who met any one or more of the following criteria were excluded from all analyses:

1. Did not provide gender information ($n = 0$, 0%).
2. Below- or near-chance perceptual decision task performance (overall accuracy < 55%) ($n = 19$, 3.56%).
3. Below the age of 18 ($n = 1$, 0.19%).
4. Incorrect response to the 'catch' item employed as an attention check ($n = 13$, 2.43%).
5. Used the same single confidence rating across all trials of the perceptual decision task ($n = 2$, 0.37%).
6. A metacognitive efficiency (***meta-d'/d'***) ratio below 0 on the perceptual decision task ($n = 27$, 5.06%).
7. Below- or near-chance knowledge decision task performance (overall accuracy < 55%) ($n = 11$, 2.06%).
8. Used the same single confidence rating across all trials of the knowledge decision task ($n = 0$, 0%).
9. A metacognitive efficiency (***meta-d'/d'***) ratio below 0 on the knowledge decision task ($n = 7$, 1.31%).
10. Failed to respond on >4 trials (out of 144) on the knowledge task ($n = 12$, 2.25%).

Based on these criteria, a total of 61 participants (11.42%) were excluded from experiment 2.

Statistical analyses

To examine the relationships between task measures and both self-reported symptoms and personality traits, we conducted a series of multiple linear regressions (always controlling for age and gender). All regressions were conducted using the *fitlm* function in MATLAB R2021a (Mathworks, USA). All variables were z-scored to ensure comparability of the regression coefficients.

The dependent measures derived from the perceptual decision-making task and the general knowledge task were type-1 accuracy (***d'***), metacognitive sensitivity (***meta-d'***), metacognitive efficiency (log(***meta-d'/d'***)) and confidence criteria (|***type-2 c'*** − ***type-1 c'***|). Due to high correlations between some of the different psychiatric symptom questionnaires, we assessed relationships between individual questionnaire scores (log-transformed) and the task measures, and between individual questionnaire scores (log-transformed) and personality dimensions, in separate regression models. In the syntax of the *fitlm* function, the regressions were as follows:

$$\text{Dependent variable} \sim \log(\text{Questionnaire Score}) + \text{age} + \text{gender}.$$

For the regressions assessing relationships between the psychiatric symptom dimensions and the task measures, and between symptom dimensions and personality dimensions, all symptom dimensions were entered in the same regression model:

$$\text{Dependent variable} \sim \text{Factor1 'anxious-depression'} + \text{Factor2 'compulsive behaviour and intrusive thought'} + \text{Factor3 'social withdrawal'} + \text{age} + \text{gender}.$$

This was also the case for the regressions assessing relationships between personality dimensions and task measures, whilst controlling for symptom dimensions:

$$\text{Dependent variable} \sim \text{extraversion} + \text{agreeableness} + \text{conscientiousness} + \text{openness} + \text{neuroticism} + \text{Factor1 'anxious-depression'} + \text{Factor2 'compulsive behaviour and intrusive thought'} + \text{Factor3 'social withdrawal'} + \text{age} + \text{gender}.$$

To correct for multiple comparisons, Bonferroni correction was applied over the number of dependent variables tested in each different analysis. For the individual questionnaire-behaviour relationships presented in Figs. 2a and 6a, c and Supplementary Fig. 6A, the corrected alpha level was 0.0014. For the symptom dimension-behaviour relationships presented in Figs. 2b and 6b, d and Supplementary Fig. 6B, the corrected alpha level was 0.0125. For the personality dimension-behaviour relationships presented in Fig. 7a, b, the corrected alpha level was 0.0167. For the individual questionnaire-personality relationships presented in Fig. 8a, the corrected alpha level was 0.0011. For the symptom dimension-personality relationships presented in Fig. 8b, the corrected alpha level was 0.01.

Pearson correlation coefficients were calculated for each of the between-subject correlations of interest. Paired- and independent-samples *t*-tests were employed to test for differences in decision parameters both within- and between-tasks.

For analysis of the relationships between psychiatric symptom dimensions and metacognitive efficiency, in addition to the linear regression approach outlined above, we also adopted two approaches which have recently been employed to test for group differences and to link external qualities to metacognitive efficiency[57,58]. These approaches incorporate Bayesian priors to constrain estimates of both group-average and individual participant metacognitive efficiency using hierarchical modelling. Two separate analyses were performed using the hierarchical fitting option in the "HMeta-d" toolbox[57]. These analyses were conducted to test the reliability of the null relationships between psychiatric symptom dimensions and metacognitive efficiency observed in the multiple linear regression analyses performed using non-hierarchical individual participant Meta-d' fits (presented in Figs. 2, 6 and 7). For both hierarchical analyses, we used the perception task data from both experiments combined ($N = 817$) to maximize statistical power.

In the first hierarchical analysis, we used median splits to create 'high' and 'low' symptom dimension score groups for each of the three dimensions (AD, CIT, and SW). The hierarchical Bayesian estimation implemented in HMeta-d' specifies group-level prior densities over each of the participant-level parameters and provides a group-level estimate of metacognitive efficiency (***meta-d'/d'***). We estimated group-level metacognitive efficiency separately for the high and low symptom groups across all three symptom dimensions. The group-level fits were performed using the *fit_meta_d_mcmc_group* function[57] with the following input parameters:

```
mcmc_params.response_conditional = 0;
mcmc_params.estimate_dprime = 0;
mcmc_params.nchains = 3;
mcmc_params.nburnin = 1000;
mcmc_params.nsamples = 10000;
mcmc_params.nthin = 1;
mcmc_params.doparallel = 0;
mcmc_params.dic = 1;
```

Group difference in metacognitive efficiency were assessed by first calculating the distribution of differences in posterior parameter samples from each group (high > low), and then determining the 95% highest-density interval (HDI) for this distribution. The group-level posterior densities were then used to test the statistical significance of differences in metacognitive efficiency. Specifically, if the 95% highest-density interval (HDI) of the difference between groups did not include 0 then the difference was judged to be statistically significant, whereas if the HDI did include 0 then the difference was judged not statistically significant[57].

In the second hierarchical analysis, we adopted a recently developed approach which allows for relationships between potential covariates and metacognitive efficiency (***meta-d'/d'***) to be estimated within the hierarchical meta-d' model[57,58]. This approach embeds the estimation of symptom-metacognitive efficiency relationships into the parameter inference routine, such

that the group-level estimate of regression coefficients reflects the influence of individual differences in symptom severity on metacognitive efficiency[58]. The regressors included in the hierarchical model were Age, Gender, AD scores, CIT scores and SW scores, with the outcome variable being metacognitive efficiency (***meta-d'/d'***) scores. The fitting was performed using the *fit_meta_d_mcmc_regression* function[57] with the following input parameters:

```
mcmc_params.response_conditional = 0;
mcmc_params.estimate_dprime = 0;
mcmc_params.nchains = 3;
mcmc_params.nburnin = 1000;
mcmc_params.nsamples = 10000;
mcmc_params.nthin = 1;
mcmc_params.doparallel = 0;
mcmc_params.dic = 1;
```

Again, posterior densities were used to test the statistical significance of the regression coefficients. Specifically, if the 95% highest-density interval (HDI) of a regression coefficient did not include 0 then the relationship was judged to be statistically significant, whereas if the HDI did include 0 then the relationship was judged not statistically significant.

RESULTS

In study 1 ($N = 344$ after data exclusion), participants performed a visual two-alternative forced-choice (2-AFC) numerosity discrimination task (Fig. 1a) and completed a battery of nine self-report psychiatric symptom questionnaires. The task involved deciding which of two simultaneously presented black boxes contained a greater number of white dots on each trial, and then rating confidence in the decision (on a scale of 1—'not confident (guessing)' to 6—'certain'). The true numerosity difference between the boxes (and hence task difficulty) was manipulated from trial-to-trial. Figure 1 provides a schematic of the trial procedure and an overview of task performance.

Psychiatric symptom dimensions are associated with dissociable 1st-order and metacognitive decision-making signatures

To quantify latent parameters contributing to both 1st- and 2nd-order decisions, we modelled the task data within an extended signal detection theory (SDT) framework[41,42,57,71]. This provided measures of both 1st-order accuracy (***d'***) and the degree to which confidence ratings dissociated correct from incorrect decisions (metacognitive sensitivity (***meta-d'***)) (see Methods for full details). Because ***d'*** and ***meta-d'*** are measured in the same units (signal-to-noise ratio), their ratio can be used to index the level of metacognitive efficiency of the observer[43,57]. This measure quantifies how much of the information available for 1st-order decisions is retained when rating confidence. The *meta-d'* model also separates sensitivity measures from measures of both 1st-order (***criterion (c')***) and 2nd-order (***confidence criteria***) bias.

We investigated relationships between self-reported psychiatric symptoms and the task measures of interest (perceptual accuracy (***d'***), metacognitive sensitivity (***meta-d'***), metacognitive efficiency (***meta-d'/d'***), confidence criteria), whilst controlling for age and gender (see Methods). Note that to calculate the task measures, individual meta-d' fits were applied to the data (collapsed across all levels of absolute numerosity difference) from each participant independently (NOT within a hierarchical model), thereby providing overall metrics of both perceptual and metacognitive performance for each participant which were independent of the data from other participants. This importantly satisfies the assumption that observations should be independent of each other for regression analysis. Full sample distributions of all measures are shown in Supplementary Fig. 1, and relationships with demographic variables (age and gender) are reported in Supplementary Fig. 2 and Supplementary Results.

Figure 2a plots standardised regression coefficients indexing the strength and direction of the relationships between questionnaire scores and each task measure. Self-reported apathy ($\beta = 0.18$, $p = .033$, corrected) and generalised anxiety ($\beta = 0.19$, $p = .027$, corrected) were positively associated with confidence criteria (indicating negative relationships with absolute confidence). No other relationships survived Bonferroni correction.

As well as relating scores on each questionnaire separately, we performed a transdiagnostic analysis[26,47] to relate underlying dimensions of psychopathology to both perceptual and metacognitive performance. The transdiagnostic approach accounts for high comorbidity between diagnostic categories (indicated by strong correlations between individual questionnaire scores (Supplementary Fig. 2A)) and potentially heterogenous symptom clusters within categories[72,73]. The questionnaires were chosen to match those of previous studies[26,47] that used factor analysis to identify three symptom dimensions underlying the 209 items across all nine questionnaires: an 'anxious-depression' (AD) dimension, a 'compulsive behaviour and intrusive thought' (CIT) dimension and a 'social withdrawal' (SW) dimension. We conducted the same factor analysis in our entire sample (across both studies: $N = 817$) and replicated the three dimensions (Supplementary Fig. 3).

We tested relationships between the symptom dimensions and task measures, again controlling for age and gender (Fig. 2b). The CIT dimension showed a dissociation between 1st- and 2nd-order effects: despite being associated with lower objective accuracy ($\beta = -0.16$, $p = .047$, corrected), CIT was also associated with reduced confidence criteria (indicating high levels of absolute confidence) ($\beta = -0.17$, $p = .022$, corrected). Conversely, whilst the AD dimension showed no relationship with objective accuracy ($\beta = 0.01$, $p = .85$), it was associated with increased confidence criteria (indicating low levels of absolute confidence) ($\beta = 0.33$, $p < .001$, corrected). The confidence criteria effects replicate Rouault, Seow, et al. (2018)[26] who found AD/CIT to be associated with low/high levels of absolute confidence, respectively. It is noteworthy that the CIT-confidence effect was not captured in the standard questionnaire analyses (Fig. 2a), and therefore the transdiagnostic approach revealed relationships masked by classic diagnostic categories.

Overall, the results of study 1 show that dissociable psychiatric symptom dimensions are associated with distinct 1st-order and metacognitive decision-making signatures, with CIT predicting reduced perceptual accuracy but high absolute confidence levels and AD predicting low absolute confidence levels despite intact perceptual accuracy.

Both domain-specific and domain-general factors contributed to performance, and confidence was the most strongly correlated measure across cognitive domains

In a 2nd study ($N = 473$ after data exclusion), we sought to extend the results in an independent sample by testing (1) whether the relationships generalise across cognitive domains and (2) whether big-5 personality dimensions explain additional variance in either 1st- and/or 2nd-order decision measures, over and above that explained by symptom dimensions. Participants performed the same perceptual task but also performed an additional 2-AFC task which tested prior knowledge of generally known quantities: national populations (Fig. 3a)[54,74]. The knowledge task was chosen to maintain a similar trial and response structure to the perceptual task while indexing performance in a different cognitive domain. The task involved judging which of two countries had the highest human population, and then rating decision confidence on the

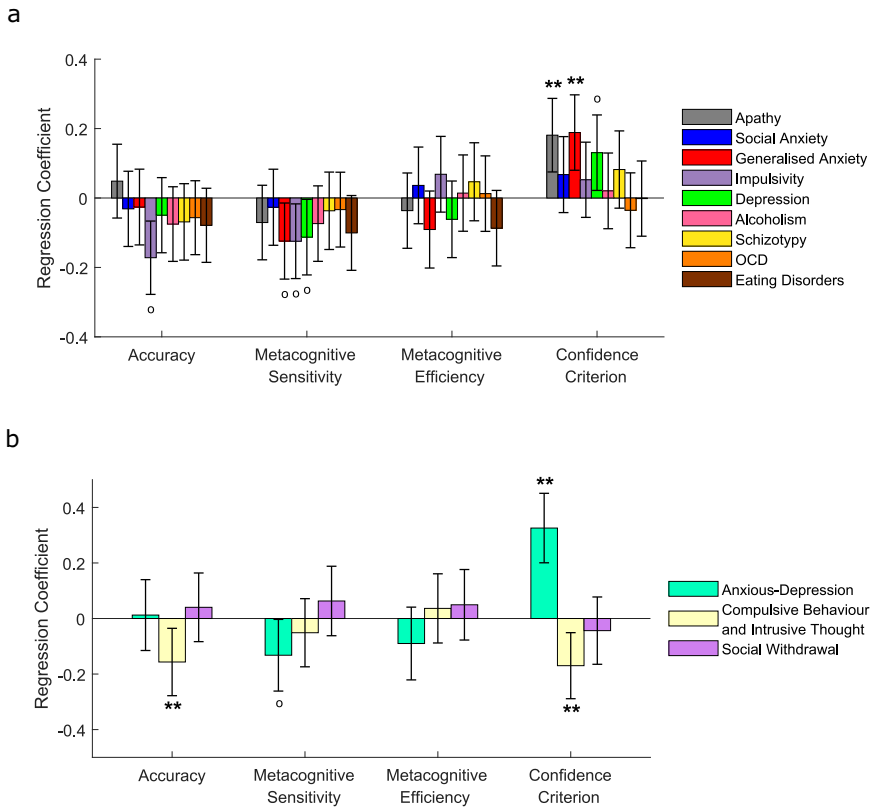

Fig. 2 Associations between 1st- and 2nd-order decision parameters and self-reported psychopathology, additionally controlling for the influence of age and gender, in experiment 1. a Associations between psychiatric symptom questionnaire scores and Meta-d′ parameters from separate regression models. Given that all variables were *z*-scored prior to entry into the regression models, the *y*-axis indicates the change in each decision parameter (in standard deviations) for each change of 1 standard deviation of questionnaire scores. Accuracy = ***d'***, Metacognitive sensitivity = ***meta-d'***, Metacognitive efficiency = log(***meta-d'***/***d'***). **b** In line with previous studies[26,47], factor analysis on the correlation matrix of all 209 questionnaire items revealed a three-factor solution comprising anxious-depression (AD), compulsive behaviour and intrusive thought (CIT) and social withdrawal (SW). The relationships between these transdiagnostic symptom dimension scores and Meta-d′ parameters were investigated using multiple regression models. CIT showed negative relationships with both 1st-order accuracy and confidence criteria, whereas AD showed a positive relationship with confidence criteria. All error bars denote 95% Confidence Intervals for the regression coefficients. °$P < 0.05$ uncorrected; **$P < 0.05$ Bonferroni corrected for multiple comparisons over the number of dependent variables tested.

same 6-point scale (1—'not confident (guessing)' to 6—'certain'). The true population difference between the two countries (and hence task difficulty) was manipulated from trial-to-trial (Methods). Figure 3 provides a schematic of the trial procedure and an overview of performance on both tasks. In addition to the nine psychiatric symptom questionnaires, participants also completed the Big Five Inventory (BFI)[68] to assess personality dimensions of 'extraversion', 'agreeableness', 'conscientiousness', 'openness to experience' and 'neuroticism'. Full sample distributions of outcome measures for study 2 are shown in Supplementary Fig. 4 and relationships with age and gender are reported in Supplementary Fig. 5 and Supplementary Results.

Comparing performance between the tasks (Fig. 4), participants performed better on the perceptual (mean ***d'*** = 1.70; SD = 0.56) compared to the knowledge (mean ***d'*** = 1.32; SD = 0.47) task ($t(472) = 13.25$, $p < .001$). However, ***meta-d'*** did not significantly differ (mean perceptual ***meta-d'*** = 1.32; SD = 0.63, mean knowledge ***meta-d'*** = 1.37; SD = 0.66: $t(472) = -1.32$, $p = .186$). ***Meta-d'*** values were more closely aligned with ***d'*** values in the knowledge task, as can be seen by comparing Fig. 4a, b. Accordingly, overall metacognitive efficiency was higher for knowledge (mean ***meta-d'/d'*** = 1.07; SD = 0.44) relative to perception (mean ***meta-d'/d'*** = 0.8; SD = 0.36) ($t(472) = 10.9$, $p < .001$) (Fig. 4c). Leftward group-level response biases (indexed by ***c'***) were significantly stronger for perception (mean perceptual ***c'*** = 0.12; SD = 0.34, mean knowledge ***c'*** = 0.04; SD = 0.13: $t(472) = 5.01$, $p < .001$) (Fig. 4d). Given that the leftward bias was present for both tasks, but stronger for perception, suggests that both motor and perceptual biases likely contributed[75–77]. Finally, despite the knowledge task being objectively more difficult than the perceptual task (as reflected by the ***d'*** differences), knowledge ***confidence criteria*** were lower (indicating higher mean confidence ratings) (mean perceptual ***confidence c'*** = 0.73; SD = 0.3, mean knowledge ***confidence c'*** = 0.65; SD = 0.24: $t(472) = -5.93$, $p < .001$) (Fig. 4e).

To estimate the contribution of domain-general mechanisms, we tested the correlation of each measure (collapsed across evidence levels) between tasks (Fig. 5). We reasoned that significant correlation of a given measure between tasks suggests that a shared latent mechanism must contribute across cognitive domains[78,79]. The only non-significant correlation was for ***c'*** ($r(471) = 0.06$, $p = .18$). All other correlations indicated influence of domain-general mechanisms on performance, though with marked differences in correlation strength across measures. Both ***d'*** ($r(471) = 0.26$, $p < .001$) and ***meta-d'*** ($r(471) = 0.16$, $p < .001$) showed moderate correlations across tasks, whilst ***meta-d'/d'*** ($r(471) = 0.09$, $p = .043$) showed the weakest correlation of the significant measures. In line with previous studies[78,79], the most strongly correlated measure across tasks was ***confidence c'*** ($r(471) = 0.52$, $p < .001$), suggesting that overall confidence calibration represents a stable, 'trait-like' measure which strongly influences metacognitive judgements across cognitive domains.

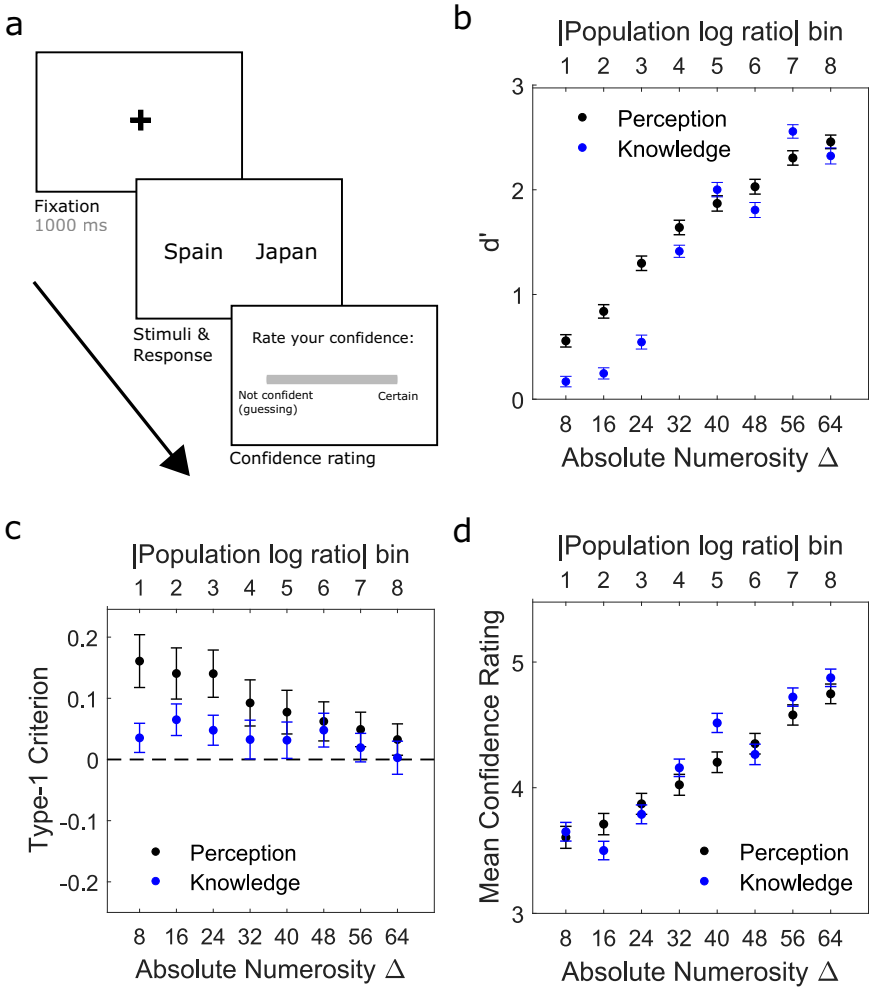

Fig. 3 Knowledge decision-making task and behaviour in study 2 ($n = 473$). a In addition to the perception task, participants also completed a task which tested knowledge of national populations. On each trial, participants judged which of two countries had the higher human population and provided a confidence rating (scale of 1–6, where 1 represented "not confident (guessing)" and 6 represented "certain"). Eight evidence discriminability bins were created by grouping pairs of countries with similar population log ratios. The log ratio bins amounted to the following, ranging from least to most discriminable: bin 1 (log10 ratio = 0–0.225), bin 2 = (0.225–0.45), bin 3 (0.45–0.675), bin 4 = (0.675–0.9), bin 5 (0.9–1.125), bin 6 = (1.125–1.35), bin 7 (1.35–1.575), bin 8 = (1.575–1.8). **b** In both tasks, group-averaged ***d'*** increased as a function of evidence strength**. c** The systematic type-1 leftward biases (here indexed by the mean type-1 ***c'***) decreased as a function of evidence level for both tasks but were systematically stronger for the perceptual task. **d** Group-averaged overall mean confidence ratings increased as a function of evidence strength. All error bars reflect 95% confidence intervals for the mean.

It is important to note that estimates of ***confidence c'*** may be inherently less noisy than estimates of ***meta-d'*** and ***meta-d'/d'***, and that this may contribute to the differences in correlation strength of these measures across tasks. Further work is needed to ascertain whether absolute confidence levels are indeed an inherently more stable trait across cognitive domains than metacognitive sensitivity/efficiency.

Psychiatrically relevant 1st- and 2nd-order decision-making signatures are domain-general

Next, we investigated whether the relationships between psychiatric symptoms and task measures are themselves domain-specific or domain-general. For perception, similar relationships between task measures and both individual questionnaires and symptom dimensions were observed to those in experiment 1, though in experiment 2 additional significant relationships were found between CIT and metacognitive sensitivity ($\beta = -0.15$, $p = .013$, corrected) and between SW and 1st-order accuracy ($\beta = 0.13$, $p = .048$, corrected) (see Fig. 6a, b compared to Fig. 2a, b). The perception-symptom relationships across both studies combined ($N = 817$) are presented in Supplementary Fig. 6. To test whether the relationships generalised across cognitive domains, we turned to the knowledge task (Fig. 6c, d). In line with perception, knowledge confidence criteria were positively associated with apathy ($\beta = 0.18$, $p < .001$, corrected) and generalised anxiety ($\beta = 0.14$, $p = .047$, corrected).

The knowledge task-symptom dimension results closely replicated those of the perceptual task (Fig. 6d). CIT was associated with reduced 1st-order accuracy ($\beta = -0.19$, $p < .001$, corrected) and metacognitive sensitivity ($\beta = -0.13$, $p = .039$, corrected) as well as reduced confidence criteria ($\beta = -0.18$, $p < .001$, corrected), whereas AD was positively associated with confidence criteria ($\beta = 0.24$, $p < .001$, corrected). However, SW and 1st-order accuracy were not correlated for the knowledge task ($\beta = 0.05$, $p = .351$). As in experiment 1, no significant relationships with metacognitive efficiency were found for any of the symptom dimensions in either task. Importantly, these null results held when we further tested them (on the combined perceptual data from both experiments) using alternative hierarchical analysis approaches which incorporated

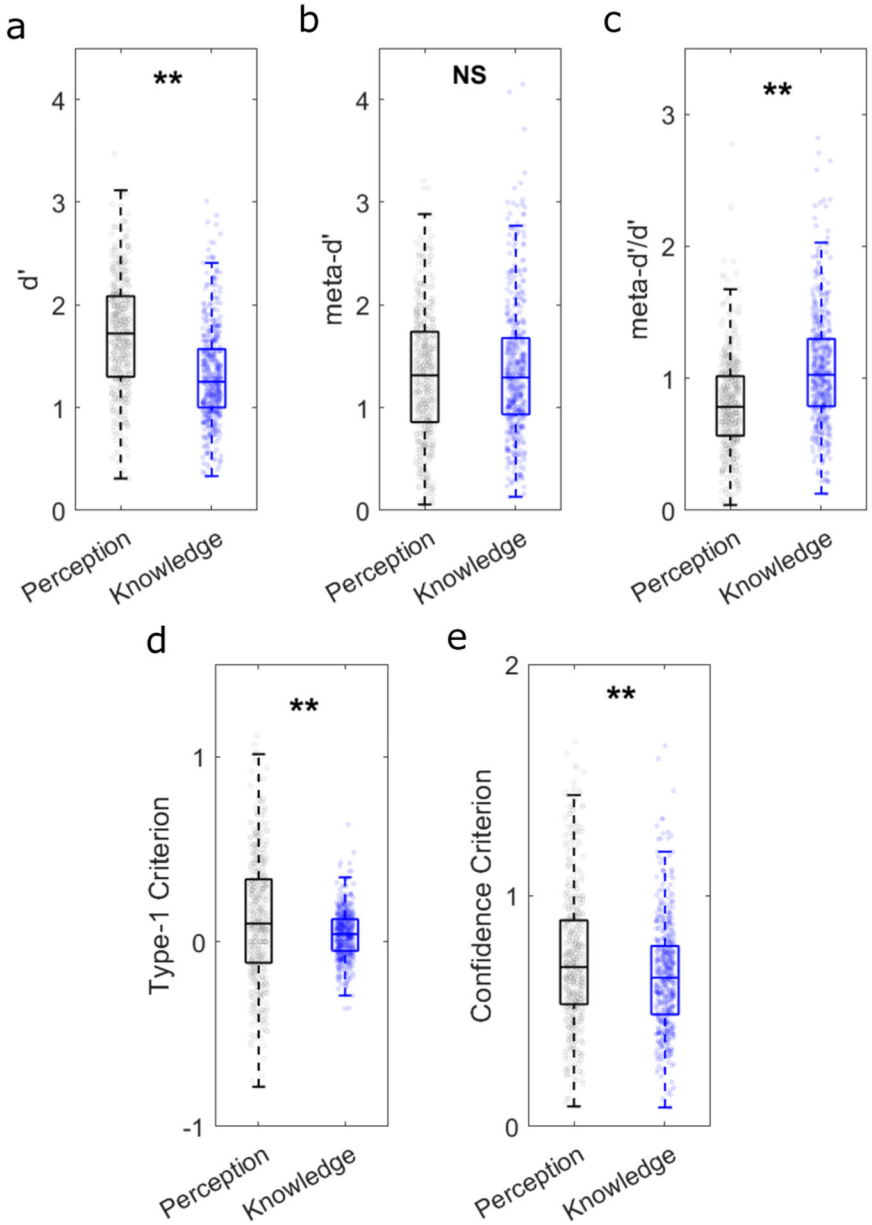

Fig. 4 Between-task comparisons of overall performance. The data are shown for (**a**) type-1 accuracy (***d'***), (**b**) metacognitive sensitivity (***meta-d'***), (**c**) metacognitive efficiency (***meta-d'/d'***), (**d**) criterion (***type-1 c'***) and (**e**) type-2 criterion (***confidence c'***). On each box, the central line is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend ±2.7 standard deviations from the median. $^{*}P < 0.05$, $^{**}P < 0.01$.

group-level prior densities when estimating metacognitive efficiency[57,58] (see Supplementary Results and Supplementary Figs. 7 and 8). Hence, we found no evidence for any relationship between symptom dimensions and metacognitive efficiency. Note that the negative relationships between CIT and metacognitive sensitivity in both tasks may be accounted for by the relationships between CIT and 1st-order accuracy, given that ***meta-d'*** positively correlates with ***d'***. Indeed, the lack of any relationship between CIT and metacognitive efficiency (***meta-d'/d'***) indicates that CIT is primarily associated with 1st-order accuracy rather than metacognitive sensitivity. Overall, overlap in relationships with psychiatric symptoms between the perceptual and knowledge tasks suggests that domain-general mechanisms largely underlie the associations between distinct dimensions of psychopathology and 1st-order and metacognitive decision signatures.

Personality explains additional variance in 1st-order decisions, but not confidence

Having established domain-general associations with dimensions of psychopathology, we next investigated whether Big-5 personality traits account for additional variance in 1st- and/or 2nd-order performance across both tasks.

We entered Big-5 factor scores into regression models as predictors along with the symptom dimensions (and age and gender) (Fig. 7). Note that variance inflation factors (VIFs) were ≤2.83 for all predictors, indicating a negligible influence of multicollinearity on the estimated coefficients[80]. The analysis was only performed for ***d'***, ***meta-d'*** and ***confidence c'*** as no relationships were found with metacognitive efficiency (***meta-d'/d'***) for any of the symptom (Fig. 6b, d) or personality (Supplementary Fig. 9) dimensions when tested independently. For the personality dimensions, extraversion was negatively

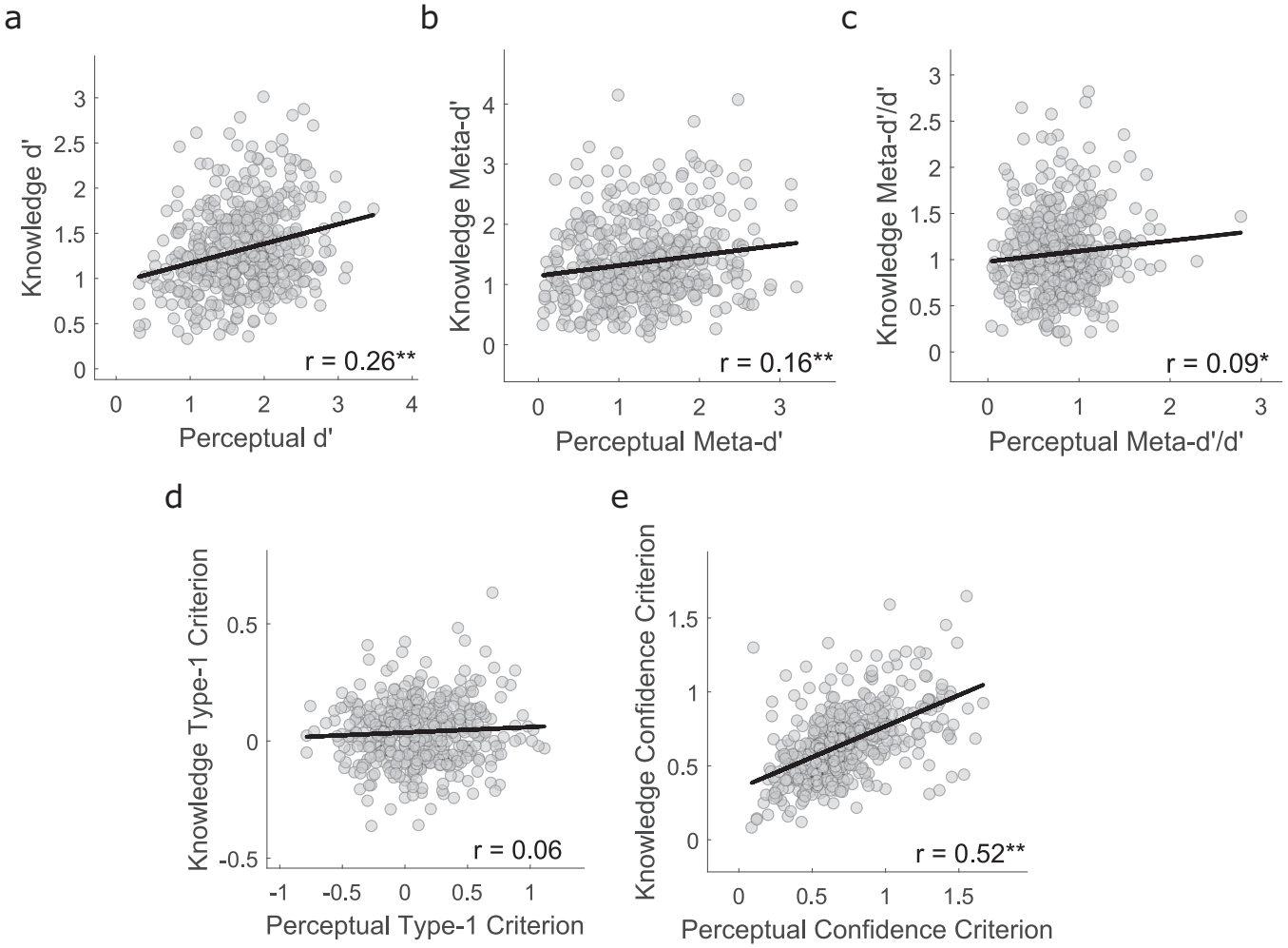

Fig. 5 Between-participant Pearson correlations across the two tasks. Data are plotted for overall (**a**) type-1 accuracy (***d'***), (**b**) metacognitive sensitivity (***meta-d'***), (**c**) metacognitive efficiency (***meta-d'/d'***), (**d**) criterion (***type-1 c'***) and (**e**) type-2 criterion (***confidence c'***). *$P < 0.05$, **$P < 0.01$.

correlated with 1st-order accuracy on the knowledge task ($\beta = -0.18$, $p = .014$, corrected) and a similar but weaker negative relationship was observed on the perceptual task ($\beta = -0.15$, $p = .023$, uncorrected). Additionally, openness to experience was positively correlated with 1st-order accuracy on the perception task ($\beta = 0.12$, $p = .034$, corrected). With personality dimensions included in the regression models, CIT scores remained significant independent predictors of 1st-order accuracy for both tasks (perception: $\beta = -0.17$, $p = .006$, corrected; knowledge: $\beta = -0.15$, $p = .019$, corrected).

No personality dimensions significantly predicted metacognitive performance (***meta-d'*** or ***confidence c'***) in either task after multiple comparison correction, whereas confidence criteria were positively related to AD (perception: $\beta = 0.19$, $p = .026$, corrected; knowledge: $\beta = 0.23$, $p = .003$, corrected), and negatively related to CIT (perception: $\beta = -0.14$, $p = .036$, corrected; knowledge: $\beta = -0.17$, $p = .003$, corrected) for both tasks. Hence, whilst both personality and psychiatric symptom dimensions were independently associated with 1st-order accuracy, symptom dimensions were the only significant predictors of domain-general confidence.

Transdiagnostic symptom dimensions elucidate relationships between personality traits and psychopathology

Finally, we investigated relationships between Big-5 personality dimensions and symptoms of psychopathology (Fig. 8). Controlling for age and gender, extraversion was negatively associated with apathy ($\beta = -0.39$, $p < .001$, corrected), social anxiety ($\beta = -0.53$, $p < .001$, corrected), generalised anxiety ($\beta = -0.47$, $p < .001$, corrected), depression ($\beta = -0.36$, $p < .001$, corrected) and schizotypy ($\beta = -0.28$, $p < .001$, corrected), but positively associated with alcoholism ($\beta = 0.16$, $p = .017$, corrected). Agreeableness was significantly negatively associated with scores on 6 out of 9 questionnaires (all β's ≤ -0.1, all p's $\leq .001$, corrected). Conscientiousness was significantly negatively associated with scores on 7 questionnaires (all β's ≤ -0.11, all p's $\leq .001$, corrected). Openness to experience was negatively associated with apathy ($\beta = -0.36$, $p < .001$, corrected). Neuroticism was significantly positively associated with scores on 8 of the questionnaires (all β's ≥ 0.09, all p's $\leq .001$, corrected).

For symptom dimensions (Fig. 8b), extraversion was negatively associated with both AD ($\beta = -0.24$, $p < .001$, corrected) and SW ($\beta = -0.62$, $p < .001$, corrected), but positively associated with CIT ($\beta = 0.28$, $p < .001$, corrected). Only AD showed a significant negative association with agreeableness ($\beta = -0.31$, $p < .001$, corrected), suggesting that this transdiagnostic dimension may account for the ubiquitous negative relationships observed across the individual questionnaires (Fig. 8a). For conscientiousness, AD was negatively associated ($\beta = -0.66$, $p < .001$, corrected) whilst SW was positively associated ($\beta = 0.17$, $p < .001$, corrected). This suggests that AD may also account for the negative relationships between multiple questionnaires and conscientiousness (Fig. 8a). Openness was negatively correlated with AD ($\beta = -0.15$, $p = .016$, corrected), but positively correlated with CIT ($\beta = 0.17$, $p = .002$,

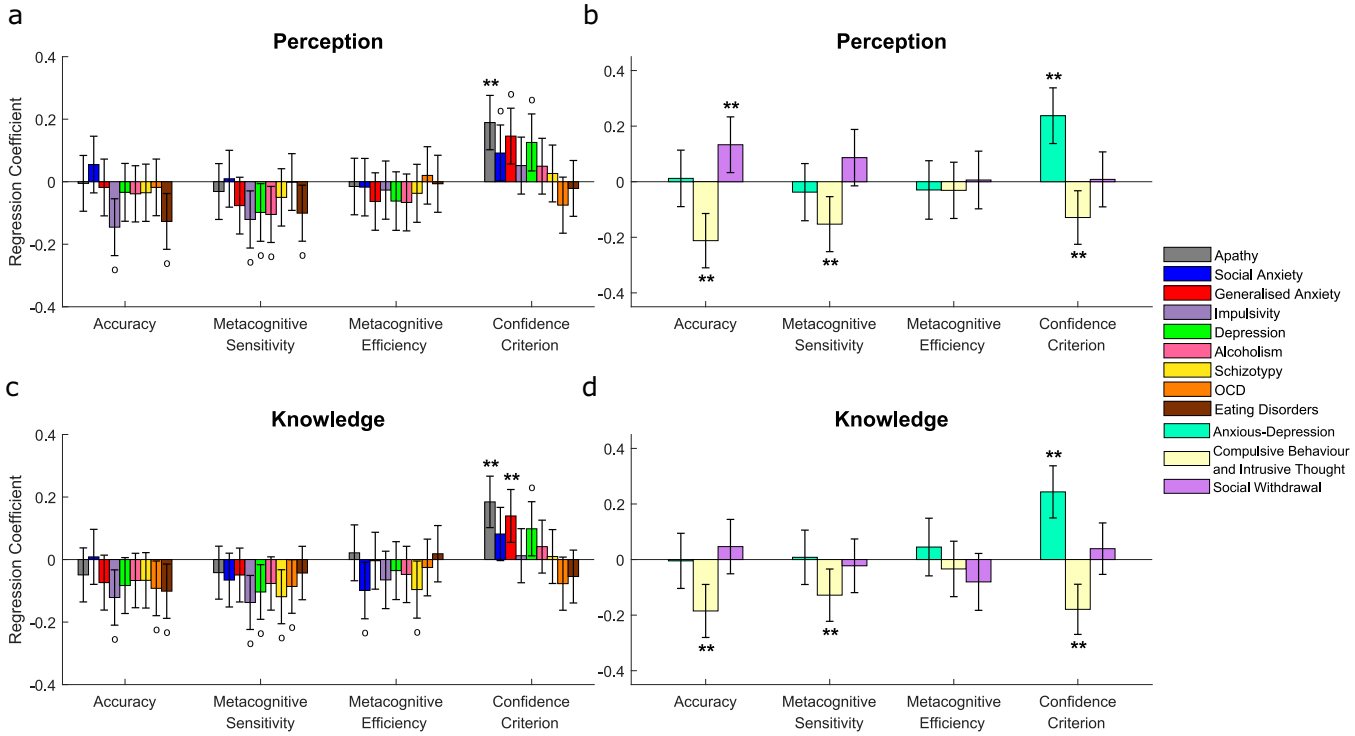

Fig. 6 Associations between 1st- and 2nd-order decision parameters and self-reported psychopathology, additionally controlling for age and gender, in experiment 2. a Associations between psychiatric symptom questionnaire scores and *perceptual* Meta-d' parameters. Given that all variables were z-scored prior to entry into the regression models, the *y*-axis indicates the change in each decision parameter (in standard deviations) for each change of 1 standard deviation of questionnaire scores. Accuracy = ***d'***, Metacognitive sensitivity = ***meta-d'***, Metacognitive efficiency = log(***meta-d'/d'***). **b** Associations between transdiagnostic symptom dimension scores and *perceptual* Meta-d' parameters. **c** Associations between psychiatric symptom questionnaire scores and *knowledge* Meta-d' parameters. **d** Associations between transdiagnostic symptom dimension scores and *knowledge* Meta-d' parameters. All error bars denote 95% Confidence Intervals for the regression coefficients. °$P < 0.05$ uncorrected; **$P < 0.05$ corrected for multiple comparisons over the number of dependent variables tested.

corrected). It is notable that no positive relationships were observed between either conscientiousness or openness and any of the individual questionnaire scores (Fig. 8a), whereas the transdiagnostic analysis revealed positive relationships with SW (conscientiousness) and CIT (openness), respectively (Fig. 8b). Hence, the transdiagnostic approach revealed relationships which were masked by classic diagnostic categories. Finally, neuroticism was positively associated with all three symptom dimensions (all β's ≥ 0.24, all p's $< .001$, corrected). The results confirm strong relationships between dimensions of personality and psychopathology and highlight that the transdiagnostic approach provides information about the nature of these relationships which is not apparent using classical diagnostic categories.

DISCUSSION

Distortions of both 1st-order perceptual decision-making and metacognitive evaluation have been suggested to characterise various forms of psychopathology. To date it has remained unclear exactly which latent processes are involved and whether the distortions generalise across cognitive domains. Here, employing a battery of self-report psychiatric symptom questionnaires and computational modelling of psychophysical performance across two studies, we found a symptom dimension characterised by 'compulsive behaviour and intrusive thought' (CIT) to be associated with reduced 1st-order objective accuracy but, paradoxically, increased confidence. Conversely, an 'anxious-depression' (AD) dimension was associated with systematically low absolute confidence in the absence of any relationship with 1st-order accuracy. These relationships replicated across perception and general knowledge tasks and occurred independently of age and gender. Alongside dimensions of psychopathology, we also investigated whether Big-5 personality traits explained additional variance in either 1st-order and/or metacognitive decision-making. Whilst dimensions of both personality (extraversion, openness) and symptoms (CIT) were independently associated with 1st-order accuracy, only symptom dimensions (AD, CIT) predicted metacognitive performance. Overall, the results reveal robust, domain-general signatures of decision-making and metacognition related to distinct psychological dispositions and psychopathology in the general population, and further elucidate the nature of relationships between personality and psychopathology.

The CIT dimension most prominently links features of impulsivity, OCD, schizotypy, addiction and eating disorders. Our results suggest domain-general alterations across multiple levels of the decision hierarchy in CIT, in line with previous studies which have found compulsivity to be associated with alterations in 1st-order perceptual decision-making[27,28,81], goal-directed control[9,47,51,82] and confidence judgements[26,28,29]. The CIT dimension was associated with a positive confidence bias (across both experiments and tasks) and reduced metacognitive sensitivity (across both tasks but only in study 2) but showed no relationship with metacognitive efficiency. Previous studies have found a reduction in metacognitive efficiency associated with compulsivity[26,27] but we did not find evidence for this here. The lack of an association with metacognitive efficiency suggests that the relationship between CIT and metacognitive sensitivity (***meta-d'***) may have been driven by the negative relationship between CIT and first order accuracy (***d'***). Our results suggest that confidence ratings still dissociate between correct and incorrect trials to the degree expected given the 1st-order performance in CIT, but overall confidence calibration is high. The apparent contradiction of

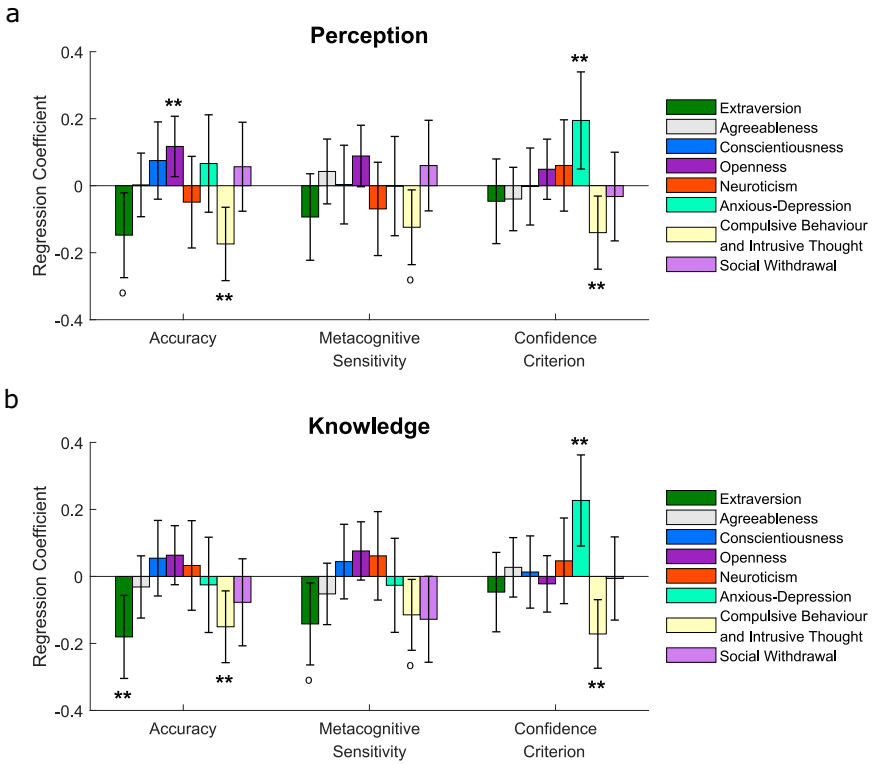

Fig. 7 Associations between 1st- and 2nd-order decision parameters and both self-reported personality traits and symptom dimensions, controlling for age and gender, in experiment 2. Data are plotted separately for the (**a**) perception and (**b**) knowledge tasks. Note that these analyses were only performed for ***d'***, ***meta-d'*** and ***confidence c'*** as no relationships were found with metacognitive efficiency (***meta-d'/d'***) for any of the symptom or personality dimensions when tested alone. All error bars denote 95% Confidence Intervals for the regression coefficients. °$P < 0.05$ uncorrected; **$P < 0.05$ corrected for multiple comparisons over the number of dependent variables tested.

reduced objective performance but inflated confidence is in line with an altered connection between confidence and behaviour[29].

The 1st-order decision deficits associated with CIT, and related disorders, have been attributed to alterations in decision formation processes such as evidence accumulation[27,44,83]. Here we show that the deficits extend beyond decisions about external sensory stimuli to include semantic memory/knowledge decisions based on internal evidence. Hence, they cannot be explained by low level sensory dysfunction. Higher order deficits in the internal modelling of task structures have also been shown to characterise compulsivity[9,82,84]. However, as optimal performance on our tasks did not require participants to learn underlying state transition probabilities, but rather depended in a straightforward manner on their decision accuracy on each individual trial, it seems unlikely that impaired internal task models can explain the 1st-order effects observed here. The effects may be explained by a recently proposed 'decision acuity' (*d*) trait found to underlie decision-making performance, independently of IQ, across a large range of decision tasks[46]. Interestingly, both *d* and IQ scores were found to be negatively related to a psychiatric dimension characterised by compulsivity/obsessionality/schizotypy (labelled 'aberrant thinking')[46].

The AD dimension, which most prominently linked features of apathy, anxiety, and depression, was associated with low confidence across cognitive domains in the absence of any relationships with objective performance. These findings confirm negative confidence bias as a feature of anxious-depressive symptomology, even in sub-clinical samples[25,26,53,85], and have implications for prominent theories of the role of metacognition in depression. Whereas the negativity hypothesis[86] posits that depressed individuals evaluate themselves in an overly negative way, the depressive realism hypothesis[87] posits that depressed individuals are more accurate in their evaluations of themselves and that it is non-depressed individuals whose evaluations are distorted by a positivity bias. Under these theories, we would expect depressive symptoms to be associated with either an increase in confidence criteria (negativity hypothesis) or an increase in metacognitive sensitivity/insight (depressive realism). Our results were more in line with the former, as we found no evidence for a relationship between AD symptoms and metacognitive efficiency. Hence, while individuals reporting high levels of AD were more negative in their confidence ratings overall (in line with the negativity hypothesis), this was not associated with a reliable alteration in their ability to dissociate correct from incorrect responses. Indeed, other recent studies have also found no relationship between metacognitive efficiency and anxious-depressive symptoms[28,53].

The computations underlying metacognitive sensitivity and bias have been suggested to arise from dissociable neural networks. For instance, in the prefrontal cortex (PFC), metacognitive sensitivity is associated primarily with anterior (aPFC) structure and activity[20,88,89], whereas absolute confidence is associated with ventromedial (vmPFC), posterior medial (mPFC) and dorsolateral (dlPFC) regions[30,90,91]. Our results suggest that anxious-depressive symptoms may be associated with changes in networks subserving absolute confidence, but not metacognitive sensitivity. Intriguingly, recent evidence suggests that interactions between confidence and reward valuation/motivation are reflected in activity in the vmPFC and dorsal anterior cingulate cortex (dACC)[25]. These regions have also been associated with symptoms of apathy[92], anxiety[93] and depression[94] and hence represent promising candidates for the neural locus of the AD effects.

The functional consequences of confidence biases in both AD and CIT should be investigated further. Negative confidence bias may have a pernicious long-term influence on motivation[95,96], learning[97,98], information seeking[6] and self-esteem[53,99] which in turn may cause and/or exacerbate anxious-depressive symptoms.

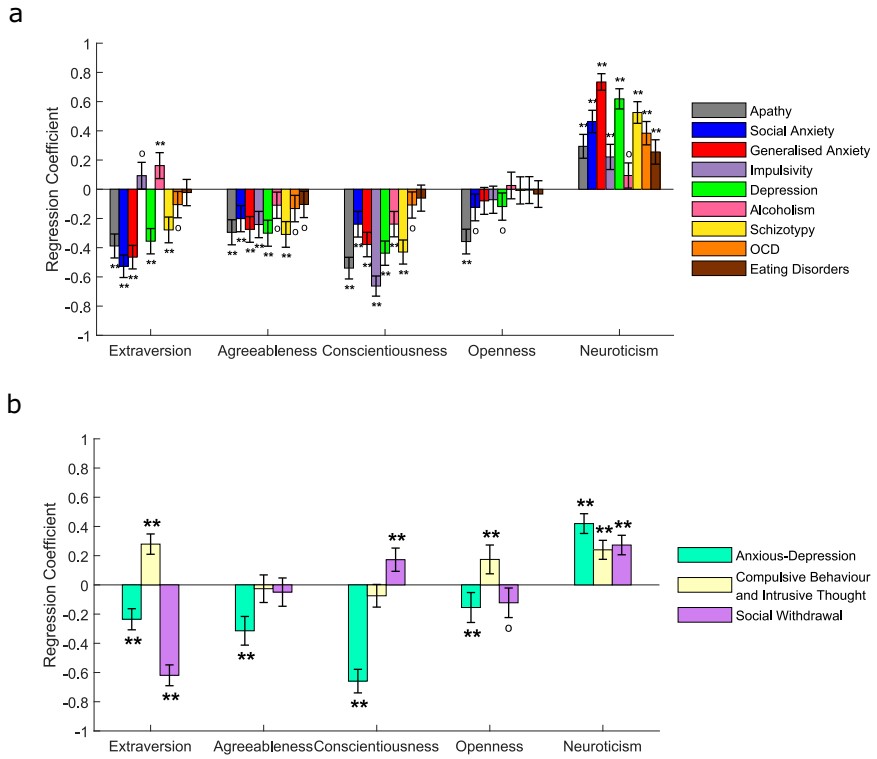

Fig. 8 Widespread associations between self-reported personality traits and psychopathology, controlling for the influence of age and gender. a Associations between psychiatric symptom questionnaire scores and personality dimension scores from separate regression models. The *y*-axis indicates the change in each personality dimension score for each change of 1 standard deviation of questionnaire scores. **b** Associations between transdiagnostic symptom dimension scores and personality dimension scores. All error bars denote 95% Confidence Intervals for the regression coefficients. °$P < 0.05$ uncorrected; **$P < 0.05$ corrected for multiple comparisons over the number of dependent variables tested.

Conversely, inflated confidence may result in rigid beliefs and cognitive inflexibility, symptoms often observed in OCD[100], addiction[101] and schizophrenia[102,103]. Changes in confidence calibration may be linked to maladaptive beliefs about self-efficacy and the level of control one has over their thoughts and/or behaviours. It would be of interest to assess whether successfully challenging these maladaptive beliefs, through techniques such as cognitive behavioural[86] or metacognitive[104] therapies, would result in corresponding changes in confidence criteria. As well as providing a useful neuro-computational outcome measure for clinical research[105], this would help to elucidate a key open question of the causal direction of the relationship between symptoms and metacognitive bias: Do the biases arise prior to, and potentially confer risk for, the onset of symptomology; or are they rather concomitant symptoms themselves? Incorporating quantitative measurement of metacognitive bias into studies employing longitudinal and/or interventional designs could shed light on this question.

We found no evidence that personality traits play a role in the relationships between metacognition and psychopathology. Metacognitive bias related to dimensions of psychopathology directly rather than through a shared link with general psychological dispositions. Indeed, Big-5 dimensions did not predict confidence in either cognitive domain. Interestingly, 1st-order accuracy was negatively associated with extraversion for both tasks. These relationships occurred independently of the accuracy-CIT relationships and, though they were not hypothesized, are in line with previous studies[32,106,107]. Hence, both personality and symptom dimensions were related to 1st-order performance. To elucidate the source of these relationships, future studies may investigate whether factors known to influence decision-making performance, such as choice history bias[108,109], attention deficits[110], confirmation bias[111], and/or alteration in reward/loss sensitivity[81,112], contribute to the observed 1st-order CIT and/or personality effects. We did not measure IQ here and so it is possible that variation in general intelligence may contribute to the effects, though evidence for relationships between IQ and both extraversion[113,114] and compulsivity[26,46] is mixed. Future studies may also investigate whether IQ and/or the recently proposed *d* factor[46] play a role in the observed 1st-order effects.

Although they were not significantly related to metacognition, personality dimensions were strongly correlated with psychopathology. Numerous relationships with classic diagnostic categories were observed for each Big-5 dimension[34–40]. However, relationships between personality and transdiagnostic symptom dimensions were also found which were masked by the classical categories: positive relationships between SW and conscientiousness, and between CIT and openness. These findings suggest links between personality traits and symptoms which do not neatly fit established diagnostic boundaries, thereby further validating interest in the identification of transdiagnostic symptom predictors[72,73]. Given that the Big-5 represent one level within a hierarchy of traits[115,116], it would be interesting to investigate exactly which subordinate facets of each dimension are most strongly linked to transdiagnostic symptoms.

Our results have implications for current models of metacognition. A normative model posits that confidence computations reflect the probability of being correct in a statistically optimal manner[117–119]. However, the relationships between symptoms and confidence ratings, and the dissociations between ***d'*** and ***meta-d'*** observed across both tasks, show that the normative model alone cannot fully account for subjective confidence. Rather, our results align with models positing that confidence

judgements arise from processes which are dissociable from the decision itself[74,120].

Both domain-specific and domain-general factors influenced metacognitive performance. At the group-level, objective accuracy was lower for knowledge than perception, but overall metacognitive efficiency and absolute confidence levels were higher. The differences in metacognitive efficiency and confidence criteria between the tasks support an influence of domain-specific factors[30,121], though it is difficult to identify exactly which as these measures are not only influenced by differences in metacognitive mechanisms between cognitive domains, but also by differences in task characteristics such as 1st-order difficulty[41] and variability in difficulty across stimulus levels[122] which were not equalised between tasks. However, a possible explanation for increased metacognitive efficiency in the knowledge task is that, whereas self-evaluation of perceptual task performance required assessment of evidence presented very briefly and then fading in iconic memory, the internally generated knowledge evidence was presumably available to the same degree throughout the trial, including during confidence judgements. Alternatively, given that confidence levels were also higher for the knowledge task here, the increased metacognitive efficiency scores may be explained by a recently discovered positive correlation between efficiency and confidence[1,123].

In support of domain-general processes also influencing performance, we found significant between-task correlations. In line with previous studies, type-1 accuracy[46], metacognitive sensitivity[124] and metacognitive efficiency[125,126] were all somewhat correlated across tasks. However, overall confidence bias was the most strongly correlated measure[78,79] and most strongly linked to symptoms. This suggests that a trait-like, global metacognitive process[9,28] links to psychopathology, as opposed to more 'local', domain-specific processes such as uncertainty about sensory evidence or model-based task representations. Global metacognitive evaluations may be intimately linked to beliefs about overall self-efficacy and are likely to have a more pervasive influence on everyday functioning[9,28]. One important consideration is that the task measures of interest here may be affected by different levels of noise[121,127] and this may have influenced both estimates of their reliability across tasks and the strength of their relationships with other variables (such as symptom scores). For instance, it is possible that estimates of confidence bias may be inherently less noisy than estimates of metacognitive sensitivity and efficiency. Although Meta-d' measures of metacognitive performance are widely adopted and currently represent the state-of-the-art in the field[41,42,57], alternative approaches to modelling/quantifying metacognitive abilities[128–130] are emerging which may be applied in future research to further characterise relationships between metacognition and psychopathology.

Testing symptom variation in the general population affords the advantage of efficient collection of large samples and overcomes the arbitrary boundaries between psychopathology and normality imposed by diagnostic manuals including the DSM[131] and ICD[132]. However, it remains to be seen whether these results can be extended to clinical samples with the highest levels of symptom severity. The transdiagnostic approach revealed relationships between psychopathology and both metacognition and personality traits which were not apparent in analyses using classic diagnostic categories (see also Rouault et al., 2018[26]), and this may be due to relationships being masked by overlap of symptom dimensions within single categorical disorders, such as overlap of AD and CIT within OCD[9,24,25,47]. This creates challenges both in terms of relating results to previous research and for translation to clinical practice[72]. Future research should investigate whether diagnostic categories (such as OCD) or transdiagnostic dimensions (such as compulsivity) are stronger predictors of cognitive and/or metacognitive deficits in clinical samples. Along these lines, Gillan et al., (2020)[133] showed that the CIT dimension was a significant predictor of deficits in goal-directed planning whereas having a diagnosis of OCD was not. Furthermore, identification and quantification of relationships between symptoms and cognition at the level of the individual, rather than at the population level[134], could remain agnostic to over-arching diagnostic labels and provide direct targets for therapeutic intervention, in line with a move towards precision psychiatry[135,136].

We employed the same battery of questionnaires as previous studies[26,47] and were able to replicate three previously reported symptom dimensions (AD, CIT, and SW). However, the questionnaire items contributing to these dimensions do not exhaustively cover all forms of psychopathology and other transdiagnostic symptom structures have been proposed[36,51,137,138] which may capture a broader range of cognitive/metacognitive alterations. It is also important to note that the age ranges of both samples here were heavily skewed towards young adults (Supplementary Figs. S1 and S4), likely due to the online recruitment strategy. Future studies should investigate decision-making and metacognition over extended symptom and age ranges and across different transdiagnostic structures. Additionally, it will be important to ascertain whether relationships between psychopathology and both 1st and 2nd-order decision-making are relatively invariant, or whether they depend on time and context[139]. For instance, the relationships may fluctuate as a function of disorder trajectory or symptom provocation. Understanding temporal dynamics and contextual triggers will help to refine models of the neurocomputational signatures associated with psychopathology and potentially facilitate the identification of novel treatment techniques.

DATA AVAILABILITY

All data are openly available on the Open Science Framework (OSF) under the URL: https://osf.io/s3cth/.

CODE AVAILABILITY

All code to reproduce the analyses are available on the OSF under the URL: https://osf.io/s3cth/.

Received: 21 April 2022; Accepted: 4 August 2022;
Published online: 30 August 2022

REFERENCES

1. Shekhar, M. & Rahnev, D. Sources of metacognitive inefficiency. *Trends Cogn. Sci.* **25**, 12–23 (2021).
2. Bahrami, B. et al. What failure in collective decision-making tells us about metacognition. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1350–1365 (2012).
3. Correa, C. M. C. et al. How the level of reward awareness changes the computational and electrophysiological signatures of reinforcement learning. *J. Neurosci.* **38**, 10338–10348 (2018).
4. van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N. & Wolpert, D. M. Confidence is the bridge between multi-stage decisions. *Curr. Biol.* **26**, 3157–3168 (2016).
5. Yeung, N. & Summerfield, C. Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1310–1321 (2012).
6. Desender, K., Boldt, A. & Yeung, N. Subjective confidence predicts information seeking in decision making. *Psychol. Sci.* **29**, 761–778 (2018).
7. Folke, T., Jacobsen, C., Fleming, S. M., & De Martino, B. Explicit representation of confidence informs future value-based decisions. *Nat. Hum. Behav.* **1**, 0002 (2017).
8. David, A. S., Bedford, N., Wiffen, B. & Gilleen, J. Failures of metacognition and lack of insight in neuropsychiatric disorders. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 1379–1390 (2012).
9. Seow, T. X. F., Rouault, M., Gillan, C. M. & Fleming, S. M. How local and global metacognition shape mental health. *Biol. Psychiatry* **90**, 436–446 (2021).

10. Fieker, M., Moritz, S., Köther, U. & Jelinek, L. Emotion recognition in depression: an investigation of performance and response confidence in adult female patients with depression. *Psychiatry Res* **242**, 226–232 (2016).
11. Fu, T., Koutstaal, W., Fu, C. H. Y., Poon, L. & Cleare, A. J. Depression, confidence, and decision: evidence against depressive realism. *J. Psychopathol. Behav. Assess.* **27**, 243–252 (2005).
12. Hancock, J. A. "Depressive realism" assessed via confidence in decision-making. *Cognit. Neuropsychiatry* **1**, 213–220 (1996).
13. Macdonald, P. A., Antony, M. M., Macleod, C. M. & Richter, M. A. Memory and confidence in memory judgments among individuals with obsessive compulsive disorder and non-clinical controls. *Behav. Res. Ther.* **35**, 497–505 (1997).
14. McNally, R. J. & Kohlbeck, P. A. Reality monitoring in obsessive-compulsive disorder. *Behav. Res. Ther.* **31**, 249–253 (1993).
15. Moritz, S. & Jaeger, A. Decreased memory confidence in obsessive–compulsive disorder for scenarios high and low on responsibility: is low still too high? *Eur. Arch. Psychiatry Clin. Neurosci.* **268**, 291–299 (2018).
16. Eifler, S. et al. Metamemory in schizophrenia: retrospective confidence ratings interact with neurocognitive deficits. *Psychiatry Res* **225**, 596–603 (2015).
17. Gawęda, Ł. et al. Impaired action self-monitoring and cognitive confidence among ultra-high risk for psychosis and first-episode psychosis patients. *Eur. Psychiatry* **47**, 67–75 (2018).
18. Lysaker, P. H. et al. Metacognitive function and fragmentation in schizophrenia: relationship to cognition, self-experience and developing treatments. *Schizophr. Res. Cogn.* **19**, 100142 (2020).
19. Moritz, S., Woodward, T. S., Whitman, J. C. & Cuttler, C. Confidence in errors as a possible basis for delusions in schizophrenia. *J. Nerv. Ment. Dis* **193**, 9–16 (2005).
20. Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J. & Rees, G. Relating introspective accuracy to individual differences in brain structure. *Science* **329**, 1541–1543 (2010).
21. Maniscalco, B., McCurdy, L. Y., Odegaard, B. & Lau, H. Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. *J. Neurosci.* **37**, 1213–1224 (2017).
22. Buratti, S., Allwood, C. M. & Kleitman, S. First- and second-order metacognitive judgments of semantic memory reports: The influence of personality traits and cognitive styles. *Metacognition Learn* **8**, 79–102 (2013).
23. Rollwage, M., Dolan, R. J. & Fleming, S. M. Metacognitive failure as a feature of those holding radical beliefs. *Curr. Biol.* **28**, 4014–4021.e8 (2018).
24. Gillan, C. M. & Seow, T. X. F. Carving out new transdiagnostic dimensions for research in mental health. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **5**, 932–934 (2020).
25. Hoven, M. et al. Abnormalities of confidence in psychiatry: an overview and future perspectives. *Transl. Psychiatry* **9**, 268 (2019).
26. 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).
27. Hauser, T. U. et al. Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Sci. Rep.* **7**, 6614–6614 (2017).
28. Hoven, M., Denys, D., Rouault, M., Luigjes, J. & Holst, R. van. How do confidence and self-beliefs relate in psychopathology: a transdiagnostic approach. *PsyArXiv* https://doi.org/10.31234/osf.io/d45gn (2022).
29. Seow, T. X. F. & Gillan, C. M. Transdiagnostic phenotyping reveals a host of metacognitive deficits implicated in compulsivity. *Sci. Rep.* **10**, 2883–2883 (2020).
30. Morales, J., Lau, H. & Fleming, S. M. Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* **38**, 3534–3546 (2018).
31. Rouault, M., McWilliams, A., Allen, M. G. & Fleming, S. M. Human metacognition across domains: insights from individual differences and neuroimaging. *Personal. Neurosci.* **1**, e17 (2018).
32. Schaefer, P. S., Williams, C. C., Goodie, A. S. & Campbell, W. K. Overconfidence and the big five. *J. Res. Personal.* **38**, 473–480 (2004).
33. Burns, K. M., Burns, N. R. & Ward, L. Confidence—more a personality or ability trait? It depends on how it is measured: a comparison of young and older adults. *Front. Psychol.* **7**, 518 (2016).
34. Watson, D., Stanton, K., Khoo, S., Ellickson-Larew, S. & Stasik-O'Brien, S. M. Extraversion and psychopathology: a multilevel hierarchical review. *J. Res. Personal.* **81**, 1–10 (2019).
35. Watson, D. et al. Aspects of extraversion and their associations with psychopathology. *J. Abnorm. Psychol.* **128**, 777–794 (2019).
36. Kotov, R., Gamez, W., Schmidt, F. & Watson, D. Linking "big" personality traits to anxiety, depressive, and substance use disorders: ameta-analysis. *Psychol. Bull.* **136**, 768–821 (2010).
37. Barlow, D. H., Sauer-Zavala, S., Carl, J. R., Bullis, J. R. & Ellard, K. K. The nature, diagnosis, and treatment of neuroticism: back to the future. *Clin. Psychol. Sci.* **2**, 344–365 (2014).
38. Brandes, C. M., Herzhoff, K., Smack, A. J. & Tackett, J. L. The p factor and the n factor: associations between the general factors of psychopathology and neuroticism in children. *Clin. Psychol. Sci.* **7**, 1266–1284 (2019).
39. Griffith, J. W. et al. Neuroticism as a common dimension in the internalizing disorders. *Psychol. Med.* **40**, 1125–1136 (2010).
40. Zinbarg, R. E. et al. Testing a Hierarchical model of neuroticism and its cognitive facets: latent structure and prospective prediction of first onsets of anxiety and unipolar mood disorders during 3 years in late adolescence. *Clin. Psychol. Sci.* **4**, 805–824 (2016).
41. Fleming, S. M. & Lau, H. C. How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443 (2014).
42. Maniscalco, B. & Lau, H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* **21**, 422–430 (2012).
43. Guggenmos, M. Measuring metacognitive performance: type 1 performance dependence and test-retest reliability. *Neurosci. Conscious.* **2021**, niab040 (2021).
44. Banca, P. et al. Evidence accumulation in obsessive-compulsive disorder: the role of uncertainty and monetary reward on perceptual decision-making thresholds. *Neuropsychopharmacology* **40**, 1192–1202 (2015).
45. Kim, J. et al. Selective impairment in visual perception of biological motion in obsessive-compulsive disorder. *Depress. Anxiety* **25**, E15–E25 (2008).
46. Moutoussis, M. et al. Decision-making ability, psychopathology, and brain connectivity. *Neuron* **109**, 2025–2040 (2021).
47. 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).
48. Huang, H., Thompson, W. & Paulus, M. P. Computational dysfunctions in anxiety: failure to differentiate signal from noise. *Biol. Psychiatry* **82**, 440–446 (2017).
49. Hunter, L. E., Meer, E. A., Gillan, C. M., Hsu, M. & Daw, N. D. Increased and biased deliberation in social anxiety. *Nat. Hum. Behav.* **6**, 146–154 (2022).
50. Patzelt, E. H., Kool, W., Millner, A. J. & Gershman, S. J. The transdiagnostic structure of mental effort avoidance. *Sci. Rep.* **9**, 1689–1689 (2019).
51. Suzuki, S., Yamashita, Y. & Katahira, K. Psychiatric symptoms influence reward-seeking and loss-avoidance decision-making through common and distinct computational processes. *Psychiatry Clin. Neurosci.* **75**, 277–285 (2021).
52. Koizumi, A. et al. Atypical spatial frequency dependence of visual metacognition among schizophrenia patients. *NeuroImage Clin* **27**, 102296 (2020).
53. Moses-Payne, M. E., Rollwage, M., Fleming, S. M. & Roiser, J. P. Postdecision evidence integration and depressive symptoms. *Front. Psychiatry* **10**, 639 (2019).
54. Sanders, J. I., Hangya, B. & Kepecs, A. Signatures of a statistical computation in the human sense of confidence. *Neuron* **90**, 499–506 (2016).
55. Green, D. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Peninsula Publishing, 1966).
56. Sherman, M. T., Seth, A. K. & Barrett, A. B. Quantifying metacognitive thresholds using signal-detection theory. *bioRxiv* https://doi.org/10.1101/361543 (2018).
57. Fleming, S. M. HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neurosci. Conscious.* **2017**, 1–14 (2017).
58. Harrison, O. K. et al. The Filter Detection Task for measurement of breathing-related interoception and metacognition. *Biol. Psychol.* **165**, 108185 (2021).
59. Zung, W. W. A self-rating depression scale. *Arch. Gen. Psychiatry* **12**, 63–70 (1965).
60. Foa, E. B. et al. The Obsessive-Compulsive Inventory: development and validation of a short version. *Psychol. Assess.* **14**, 485–496 (2002).
61. Spielberger, C. D. State-trait anxiety inventory for adults. *PA PsycTests* https://doi.org/10.1037/t06496-000 (2012).
62. 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).
63. Marin, R. S., Biedrzycki, R. C. & Firinciogullari, S. Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Res* **38**, 143–162 (1991).
64. 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).
65. Patton, J. H., Stanford, M. S. & Barratt, E. S. Factor structure of the Barratt impulsiveness scale. *J. Clin. Psychol.* **51**, 768–774 (1995).
66. Mason, O., Linney, Y. & Claridge, G. Short scales for measuring schizotypy. *Schizophr. Res.* **78**, 293–296 (2005).
67. Liebowitz, M. R. Social Phobia. in *Modern Trends in Pharmacopsychiatry* Vol 22 (ed Klein, D. F.) 141–173 (S. Karger AG, 1987).
68. John, O. P., Donahue, E. M. & Kentle, R. L. Big five inventory. *APA PsycTests* https://doi.org/10.1037/t07550-000 (2012).
69. Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N. & Evershed, J. K. Gorilla in our midst: an online behavioral experiment builder. *Behav. Res. Methods* **52**, 388–407 (2020).

70. Rahnev, D. et al. The confidence database. *Nat. Hum. Behav.* **4**, 317–325 (2020).
71. Galvin, S. J., Podd, J. V., Drga, V. & Whitmore, J. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon. Bull. Rev.* **10**, 843–876 (2003).
72. 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).
73. 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).
74. Pleskac, T. J. & Busemeyer, J. R. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* **117**, 864–901 (2010).
75. Benwell, C. S. Y., Harvey, M. & Thut, G. On the neural origin of pseudoneglect: EEG-correlates of shifts in line bisection performance with manipulation of line length. *NeuroImage* **86**, 370–380 (2014).
76. Jewell, G. & McCourt, M. E. Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. *Neuropsychologia* **38**, 93–110 (2000).
77. Veniero, D., Benwell, C. S. Y., Ahrens, M. M. & Thut, G. Inconsistent effects of parietal α-tACS on Pseudoneglect across two experiments: a failed internal replication. *Front. Psychol.* **8**, 952 (2017).
78. Ais, J., Zylberberg, A., Barttfeld, P. & Sigman, M. Individual consistency in the accuracy and distribution of confidence judgments. *Cognition* **146**, 377–386 (2016).
79. Mazancieux, A., Dinze, C. & Souchay, C. & Moulin, C. J. A. Metacognitive domain specificity in feeling-of-knowing but not retrospective confidence. *Neurosci. Conscious.* **2020**, niaa001 (2020).
80. James, G., Witten, D., Hastie, T. & Tibshirani, R. *An Introduction to Statistical Learning* (Springer New York, 2013).
81. Franken, I. H. A., van Strien, J. W., Nijs, I. & Muris, P. Impulsivity is associated with behavioral decision-making deficits. *Psychiatry Res* **158**, 155–163 (2008).
82. Vaghi, M. M. et al. Compulsivity is linked to reduced adolescent development of goal-directed control and frontostriatal functional connectivity. *Proc. Natl. Acad. Sci.* **117**, 25911–25922 (2020).
83. Solway, A., Lin, Z. & Vinaik, E. Transfer of information across repeated decisions in general and in obsessive–compulsive disorder. *Proc. Natl. Acad. Sci.* **118**, e2014271118 (2021).
84. Voon, V. et al. Disorders of compulsivity: a common bias towards learning habits. *Mol. Psychiatry* **20**, 345–352 (2015).
85. Orth, U. & Robins, R. W. Understanding the link between low self-esteem and depression. *Curr. Dir. Psychol. Sci.* **22**, 455–460 (2013).
86. Beck. A. *Depression: Clinical, Experimental and Theoretical Aspects* (Harper and Row, 1967).
87. Alloy, L. B. & Abramson, L. Y. Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *J. Exp. Psychol. Gen.* **108**, 441–485 (1979).
88. Allen, M. et al. Metacognitive ability correlates with hippocampal and prefrontal microstructure. *NeuroImage* **149**, 415–423 (2017).
89. Baird, B., Smallwood, J., Gorgolewski, K. J. & Margulies, D. S. Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *J. Neurosci.* **33**, 16657–16665 (2013).
90. Bang, D. & Fleming, S. M. Distinct encoding of decision confidence in human medial prefrontal cortex. *Proc. Natl Acad. Sci. USA* **115**, 6082–6087 (2018).
91. Vaccaro, A. G. & Fleming, S. M. Thinking about thinking: a coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain Neurosci. Adv.* **2**, 239821281881059 (2018).
92. Hogeveen, J., Hauner, K. K., Chau, A., Krueger, F. & Grafman, J. Impaired valuation leads to increased apathy following ventromedial prefrontal cortex damage. *Cereb. Cortex* **27**, 1401–1408 (2016).
93. Wang, H.-Y. et al. Prefrontoparietal dysfunction during emotion regulation in anxiety disorder: a meta-analysis of functional magnetic resonance imaging studies. *Neuropsychiatr. Dis. Treat.* **14**, 1183–1198 (2018).
94. Dillon, D. G. et al. Peril and pleasure: an rdoc-inspired examination of threat responses and reward processing in anxiety and depression: neighborhood characteristics and mental health. *Depress. Anxiety* **31**, 233–249 (2014).
95. Jiang, Y. & Kleitman, S. Metacognition and motivation: links between confidence, self-protection and self-enhancement. *Learn. Individ. Differ.* **37**, 222–230 (2015).
96. Schunk, D. H. & DiBenedetto, M. K. Motivation and social cognitive theory. *Contemp. Educ. Psychol.* **60**, 101832 (2020).
97. Greven, C. U., Harlaar, N., Kovas, Y., Chamorro-Premuzic, T. & Plomin, R. More than just IQ: school achievement is predicted by self-perceived abilities—but for genetic rather than environmental reasons. *Psychol. Sci.* **20**, 753–762 (2009).
98. Lebreton, M., Bacily, K., Palminteri, S. & Engelmann, J. B. Contextual influence on confidence judgments in human reinforcement learning. *PLoS Comput. Biol.* **15**, e1006973 (2019).
99. Rouault, M., Will, G.-J., Fleming, S. M. & Dolan, R. J. Low self-esteem and the formation of global self-performance estimates in emerging adulthood. *Transl. Psychiatry* **12**, 1–10 (2021).
100. Chamberlain, S. R., Solly, J. E., Hook, R. W., Vaghi, M. M. & Robbins, T. W. In *The Neurobiology and Treatment of OCD:* Accelerating *Progress* Vol 49 (eds Fineberg, N. A. & Robbins, T. W.) 125–145 (Springer International Publishing, 2021).
101. Perandrés-Gómez, A., Navas, J. F., van Timmeren, T. & Perales, J. C. Decision-making (in)flexibility in gambling disorder. *Addict. Behav.* **112**, 106534 (2021).
102. Joyce, D. W., Averbeck, B. B., Frith, C. D. & Shergill, S. S. Examining belief and confidence in schizophrenia. *Psychol. Med.* **43**, 2327–2338 (2013).
103. Serrano-Guerrero, E., Ruiz-Veguilla, M., Martín-Rodríguez, A. & Rodríguez-Testal, J. F. Inflexibility of beliefs and jumping to conclusions in active schizophrenia. *Psychiatry Res* **284**, 112776 (2020).
104. Wells, A. et al. Metacognitive therapy in recurrent and persistent depression: a multiple-baseline study of a new treatment. *Cogn. Ther. Res.* **33**, 291–300 (2009).
105. Reiter, A. M., Atiya, N. A., Berwian, I. M. & Huys, Q. J. Neuro-cognitive processes as mediators of psychological treatment effects. *Curr. Opin. Behav. Sci.* **38**, 103–109 (2021).
106. Chamorro-Premuzic, T., Furnham, A. & Ackerman, P. L. Ability and personality correlates of general knowledge. *Personal. Individ. Differ.* **41**, 419–429 (2006).
107. Graham, E. K. & Lachman, M. E. Personality stability is associated with better cognitive performance in adulthood: are the stable more able? *J. Gerontol. B. Psychol. Sci. Soc. Sci.* **67**, 545–554 (2012).
108. Benwell, C. S. Y., Beyer, R., Wallington, F. & Ince, R. A. A. History biases reveal novel dissociations between perceptual and metacognitive decision-making. *bioRxiv* https://doi.org/10.1101/737999 (2019).
109. Urai, A. E., de Gee, J. W., Tsetsos, K. & Donner, T. H. Choice history biases subsequent evidence accumulation. *eLife* **8**, e46331 (2019).
110. Dekkers, T. J. et al. Decision-making deficits in ADHD are not related to risk seeking but to suboptimal decision-making: meta-analytical and novel experimental evidence. *J. Atten. Disord.* **25**, 486–501 (2021).
111. Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J. & Fleming, S. M. Confidence drives a neural confirmation bias. *Nat. Commun.* **11**, 1–11 (2020).
112. Gullo, M. J. & Stieger, A. A. Anticipatory stress restores decision-making deficits in heavy drinkers by increasing sensitivity to losses. *Drug Alcohol Depend* **117**, 204–210 (2011).
113. Pincombe, J. L., Luciano, M., Martin, N. G. & Wright, M. J. Heritability of NEO PI-R extraversion facets and their relationship with IQ. *Twin Res. Hum. Genet.* **10**, 462–469 (2007).
114. Roberts, M. J. The relationship between extraversion and ability. *Personal. Individ. Differ.* **32**, 517–522 (2002).
115. Costa, P. T. Jr & McCrae, R. R. Domains and facets: hierarchical personality assessment using the revised NEO personality inventory. *J. Pers. Assess.* **64**, 21–50 (1995).
116. DeYoung, C. G., Quilty, L. C. & Peterson, J. B. Between facets and domains: 10 aspects of the big five. *J. Pers. Soc. Psychol.* **93**, 880–896 (2007).
117. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231 (2008).
118. Meyniel, F., Schlunegger, D. & Dehaene, S. The sense of confidence during probabilistic learning: a normative account. *PLOS Comput. Biol.* **11**, e1004305 (2015).
119. Pouget, A., Drugowitsch, J. & Kepecs, A. Confidence and certainty: distinct probabilistic quantities for different goals. *Nat. Neurosci.* **19**, 366–374 (2016).
120. Fleming, S. M. & Daw, N. D. Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation. *Psychol. Rev.* **124**, 91–114 (2017).
121. Shekhar, M. & Rahnev, D. The nature of metacognitive inefficiency in perceptual decision making. *Psychol. Rev.* **128**, 45–70 (2021).
122. Rahnev, D. & Fleming, S. M. How experimental procedures influence estimates of metacognitive ability. *Neurosci. Conscious.* **2019**, niz010 (2019).
123. Xue, K., Shekhar, M. & Rahnev, D. Examining the robustness of the relationship between metacognitive efficiency and metacognitive bias. *Conscious. Cogn.* **95**, 103196 (2021).
124. Samaha, J. & Postle, B. R. Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. *Proc. R. Soc. B Biol. Sci.* **284**, 20172035 (2017).
125. Faivre, N., Filevich, E., Solovey, G., Kühn, S. & Blanke, O. Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. *J. Neurosci.* **38**, 263–277 (2018).
126. McCurdy, L. Y. et al. Anatomical coupling between distinct metacognitive systems for memory and visual perception. *J. Neurosci.* **33**, 1897–1906 (2013).
127. Bang, J. W., Shekhar, M. & Rahnev, D. Sensory noise increases metacognitive efficiency. *J. Exp. Psychol. Gen.* **148**, 437–452 (2019).

128. Guggenmos, M. Reverse engineering of metacognition. *bioRxiv* https://doi.org/10.1101/2021.10.10.463812 (2021).
129. Paulewicz, B., Siedlecka, M. & Koculak, M. Confounding in studies on metacognition: a preliminary causal analysis framework. *Front. Psychol.* **11**, 1933 (2020).
130. Shekhar, M. & Rahnev, D. How do humans give confidence? A comprehensive comparison of process models of metacognition. *PsyArXiv* https://doi.org/10.31234/osf.io/cwrnt (2022).
131. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (CBS, 2013).
132. World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. (World Health Organization., 1992).
133. Gillan, C. M. et al. Comparison of the association between goal-directed planning and self-reported compulsivity vs obsessive-compulsive disorder diagnosis. *JAMA Psychiatry* **77**, 77 (2020).
134. Ince, R. A., Paton, A. T., Kay, J. W. & Schyns, P. G. Bayesian inference of population prevalence. *eLife* **10**, e62461 (2021).
135. Friston, K. J., Redish, A. D. & Gordon, J. A. Computational nosology and precision psychiatry. *Comput. Psychiatry* **1**, 2 (2017).
136. Y Niv, P Hitchcock, I M Berwian, & G Schoen. in *Precision Psychiatry: Using Neuroscience Insights to Inform Personally Tailored, Measurement-Based Care* (Williams, C. et al) Ch. 12 (American Psychiatric Association Publishing, 2021).
137. Caspi, A. et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).
138. Keyes, K. M. et al. Thought disorder in the meta-structure of psychopathology. *Psychol. Med.* **43**, 1673–1683 (2013).
139. Hitchcock, P. F., Fried, E. I. & Frank, M. J. Computational psychiatry needs time and context. *Annu. Rev. Psychol.* **73**, 243–270 (2022).

ACKNOWLEDGEMENTS

C.S.Y.B. was supported by the British Academy/Leverhulme Trust and the United Kingdom Department for Business, Energy and Industrial Strategy [SRG19/191169]. C.S.Y.B. and A.K. were supported by an Undergraduate Research Assistantship Award from the British Psychological Society (BPS). R.A.A.I. was supported by the Wellcome Trust [214120/Z/18/Z]. G.M. was supported by the United Kingdom Engineering and Physical Sciences Research Council (EPSRC) [EP/R513222/1].

AUTHOR CONTRIBUTIONS

Conceptualization, C.S.Y.B.; data collection, C.S.Y.B., J.W. and A.K.; formal analysis, C.S.Y.B., G.M. and R.A.A.I.; writing, C.S.Y.B., J.W., A.K., G.M. and R.A.A.I. Supervision, C.S.Y.B., R.A.A.I.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44184-022-00009-4.

Correspondence and requests for materials should be addressed to Christopher S. Y. Benwell.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.