

Metacognition in anxious-depression is state-dependent: an observational treatment study

Reviewed Preprint

Published from the original preprint after peer review and assessment by eLife.

About eLife's process

Reviewed preprint posted

July 13, 2023 (this version)

Posted to PsyArXiv

April 25, 2023

Sent for peer review

March 30, 2023

Celine A Fox , Chi Tak Lee, Anna K Hanlon, Tricia XF Seow, Kevin Lynch, Siobhán Harty, Derek Richards, Jorge Palacios, Veronica O'Keane, Klaas E Stephan, Claire M Gillan

Department of Psychology, Trinity College Dublin, Dublin, Ireland • Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland • Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom • SilverCloud Science, SilverCloud Health Ltd, Dublin, Ireland • Department of Psychiatry, Trinity College Dublin, Dublin, Ireland • Tallaght Hospital, Trinity Centre for Health Sciences, Tallaght University Hospital, Tallaght, Dublin, Ireland • Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Zurich, Switzerland • Max Planck Institute for Metabolism Research, Cologne, Germany • Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

 https://en.wikipedia.org/wiki/Open_access

 <https://creativecommons.org/licenses/by/4.0/>

Abstract

Objective: Prior studies have found metacognitive impairments are linked to a transdiagnostic dimension of anxious-depression, manifesting as reduced confidence in performance ('metacognitive bias'). However, previous work has been cross-sectional and so it is unclear if under-confidence is a trait-like marker of anxious-depression vulnerability, or if it resolves when anxious-depression improves.

Methods: Data were collected as part of the 'Precision in Psychiatry' study, a large-scale transdiagnostic, four-week observational study of individuals initiating internet-based cognitive behavioural therapy (iCBT) or antidepressant medication. Self-reported clinical questionnaires and perceptual task performance were gathered to assess anxious-depression and metacognitive bias at baseline and four-week follow-up. Primary analyses were conducted for individuals who received iCBT (n=649), with comparisons between smaller samples that received antidepressant medication (n=82) and a control group receiving no intervention (n=88).

Results: Prior to receiving treatment, anxious-depression severity was associated with under-confidence in performance in the iCBT arm, replicating previous work. From baseline to follow-up, levels of anxious-depression were significantly reduced, and this was accompanied by a significant increase in metacognitive confidence ($\beta=0.17$, $SE=0.02$, $p<0.001$). These changes were correlated ($r(647)=-0.12$, $p=0.002$); those with the greatest reductions in anxious-depression levels had the largest increase in confidence. In the antidepressant arm, anxious-depression reduced ($\beta=-0.61$, $SE=0.09$, $p<0.001$) and confidence increased ($\beta=0.31$, $SE=0.08$, $p<0.001$). Among controls, confidence remained stable from baseline to follow-up ($\beta=0.11$, $SE=0.07$, $p=0.103$).

Conclusions: Metacognitive biases in anxious-depression are state-dependent; when symptoms improve with treatment, so does confidence in performance. Our results suggest this is not specific to the type of intervention.

eLife assessment

This **valuable** study advances our knowledge of the effects of anxiety/depression treatment on metacognition, demonstrating that treatment increases metacognitive confidence alongside improving symptoms. The authors provide **convincing** evidence for the state-dependency of metacognitive confidence, based on a large longitudinal treatment dataset. However, it is unclear to what extent this effect is truly specific to treatment, as changes in metacognitive confidence in the group receiving online therapy were not statistically different from those in the control group.

Introduction

Metacognition refers to the ability to accurately monitor and appraise one's own cognitive experience (1). Metacognition is crucial for adaptive behaviour: it allows for flexible adjustment of behavioural strategies in order to improve performance, signals when to engage or withdraw from an activity, and guides the engagement in social interactions (2,3). Metacognitive abilities vary across individuals; one can be under- or over-confident, and these biases can be associated with maladaptive thoughts, feelings and behaviours (4). There is a growing interest in these confidence abnormalities in psychiatry, with studies implicating alterations of metacognition in depression, obsessive-compulsive disorder, and psychosis (5). While case-control studies have mainly found patterns of reduced confidence across several disorders, newer methods that can separate transdiagnostic dimensions of mental health using large online samples have revealed specific and bi-directional effects of confidence (6). Using these methods, studies have shown that the transdiagnostic dimension 'anxious-depression' is linked to under-confidence in one's own performance, while a separate dimension 'compulsivity and intrusive thought' is related to elevated confidence (7–11).

A major gap in this area is that studies to-date only measure metacognition and transdiagnostic psychopathology at a single time point. Therefore, it is unclear if metacognitive biases are stable, fixed traits, or if they might change with treatment response. Preliminary evidence suggests metacognition may indeed be malleable; metacognitive abilities can be improved with metacognitive interventions, such as training, in unselected online samples (12,13) and in clinical populations (14,15). However, it remains unknown if metacognitive changes generalise beyond its specific training context and are associated with any real-world improvement in psychiatric symptoms. In clinical studies, research has identified confidence abnormalities in at-risk populations (16,17), suggestive of a trait-dependence. In contrast, stimulant use disorders remitters have better metacognition than active users, suggesting state-dependence (18). Within-subject designs are needed to extend this work and understand if metacognition can improve in parallel to symptom alleviation, or if those with greater metacognitive deficits are simply the most vulnerable to illness onset and persistence.

The present study aimed to address this by examining metacognition in a large cohort of individuals before and after internet-based cognitive behavioural therapy (iCBT). iCBT has emerged as an important intervention for reducing the treatment-gap in mental healthcare provision globally; it is low-cost, scalable, geographically unconstrained and flexible (19,20). iCBT offers patients standardised content and records objective metrics of treatment engagement, making it particularly well-suited to treatment-oriented research in psychiatry (21). However, most studies on iCBT have examined clinical effectiveness (22–24), but little research on how it might impact on metacognition (25). In this study, we used an objective task measure of

metacognition (26 [↗](#)), which allowed us to test if successful treatment is linked to within-person improvements in metacognition. We also tested if any changes in metacognition were iCBT-specific, by comparing data gathered from smaller samples of individuals receiving antidepressant medication and a control group receiving no intervention. Similar to iCBT, antidepressants have established transdiagnostic efficacy (27 [↗](#)–29 [↗](#)), but studies typically lacked power to detect effects of antidepressants on cognitive abilities (30 [↗](#)–33 [↗](#)). Accordingly, a secondary aim of this study was to compare metacognitive changes across the different intervention arms, which may shed light on differential therapeutic mechanisms and potentially augment therapeutic decision-making in the future.

Results

Cross-sectional Findings at Baseline: iCBT

At baseline, participants with higher levels of anxious-depression had lower levels of mean confidence ($\beta=-0.09$, $SE=0.03$, $p=0.008$; **Figure 2A & 2B** [↗](#)), while those with higher levels of compulsivity and intrusive thought had elevated mean confidence ($\beta=0.11$, $SE=0.03$, $p=0.002$; **Figure 2A & 2C** [↗](#)), controlling for age, gender, and education. Levels of social withdrawal were not associated with mean confidence ($\beta=-0.05$, $SE=0.03$, $p=0.168$; **Figure 2A** [↗](#)).

Treatment Findings: iCBT

The transdiagnostic dimensions and psychiatric scale scores all significantly improved from baseline to four-week follow-up, except for impulsivity (**Figure 3A** [↗](#), **Table S2** [↗](#)). In tandem with these clinical changes, there was a small but significant increase in mean confidence from baseline ($M=3.78$, $SD=0.85$) to follow-up ($M=3.95$, $SD=0.89$), ($\beta=0.17$, $SE=0.02$, $p<0.001$, $r^2=0.01$) (**Figure 3B** [↗](#)). Performance accuracy improved, which due to the staircasing calibration was reflected as increased task difficulty from baseline (dot difference: $M=41.82$, $SD=11.61$) to follow-up (dot difference: $M=39.80$, $SD=12.62$), ($\beta=-2.02$, $SE=0.44$, $p<0.001$, $r^2=0.01$) (**Figure 3C** [↗](#)). Additionally, the effect of time on confidence was not dependent on how much participants engaged in iCBT, as indexed by time spent in the program ($\beta<0.01$, $SE<0.01$, $p=0.756$) and percentage of the iCBT program viewed ($\beta=0.09$, $SE=0.21$, $p=0.650$). Change in confidence was also not dependent on receiving concurrent treatment ($\beta=-0.03$, $SE=0.06$, $p=0.566$), with 175 participants (27.0%) in the iCBT group receiving another treatment during the study (further detailed in the **Supplement** [↗](#)).

Importantly, change in anxious-depression was significantly associated with change in confidence ($r(647)=-0.12$, $p=0.002$), such that those with the largest decrease in anxious-depression had the greatest increase in confidence (**Figure 3D** [↗](#)). Change in confidence was not significantly associated with change in compulsivity and intrusive thought ($r(647)=-0.05$, $p=0.211$). The relationship between change in anxious-depression and change in task difficulty was not significant ($r(647)=0.01$, $p=0.835$). A significant interaction effect of time and anxious-depression on mean confidence held when including change in compulsivity and intrusive thought and change in social withdrawal as covariates in the model ($\beta=-0.07$, $SE=0.03$, $p=0.005$). There was no significant interaction effect of time and baseline confidence ($\beta=0.06$, $SE=0.04$, $p=0.144$) or an interaction effect of time and baseline anxious-depression ($\beta=-0.03$, $SE=0.05$, $p=0.611$) on mean confidence. Exploratory analyses determined the specificity of these effects to anxious-depression by examining the interaction effect of time and change in each psychiatric score on mean confidence. Changes in trait anxiety ($\beta=-0.08$, $SE=0.02$, $p=0.002$), depression ($\beta=-0.06$, $SE=0.02$, $p=0.011$) and alcohol misuse ($\beta=-0.05$, $SE=0.02$, $p=0.037$) also showed an association with changes in confidence (**Figure 3E** [↗](#) and **Table S3** [↗](#)).

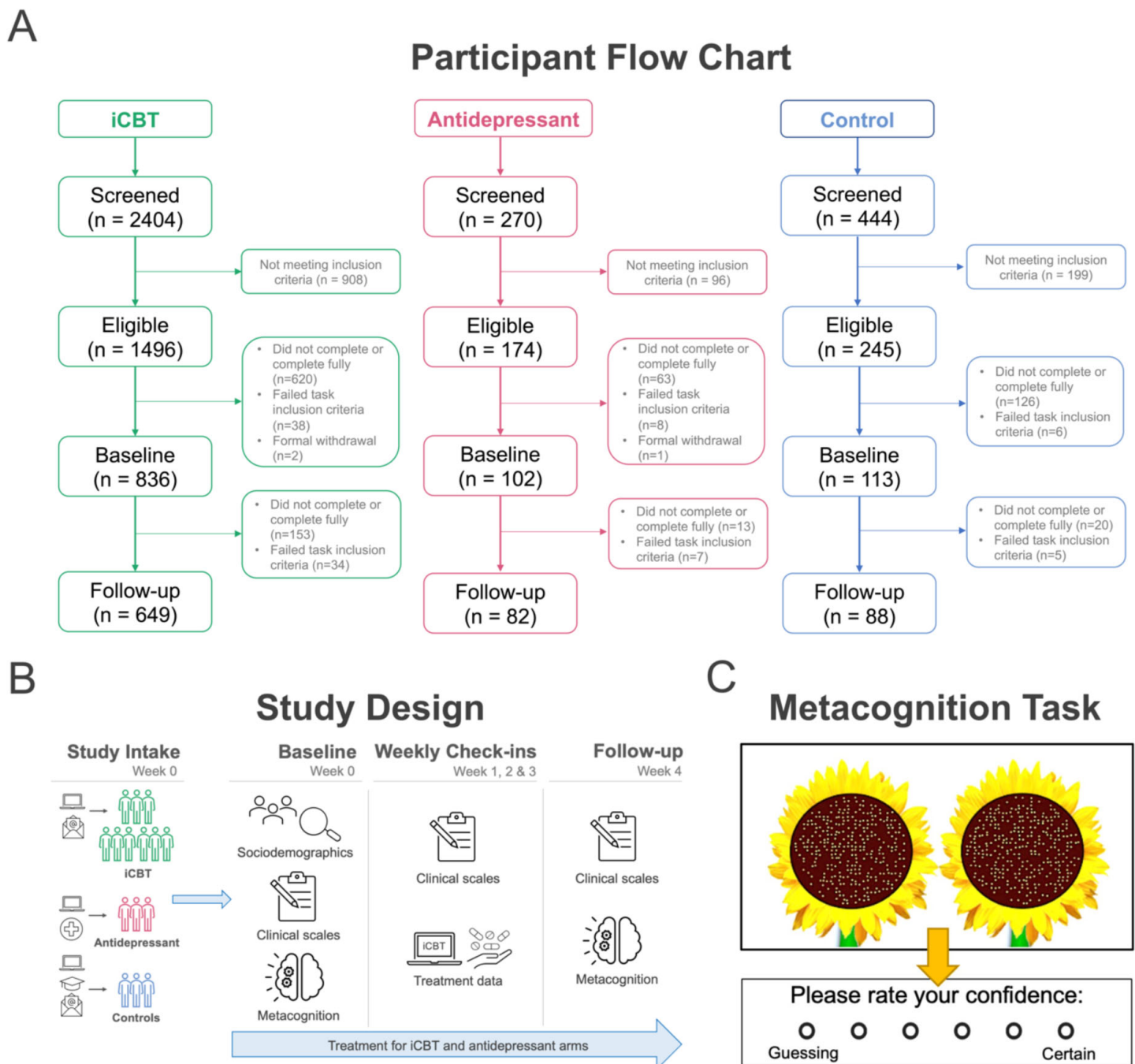


Figure 1

(A) Participant flow chart (CONSORT chart). Participants were considered 'completers' if they had metacognitive and transdiagnostic psychiatric dimension data at baseline and follow-up and met task inclusion criteria. (B) Overview of study design from study intake (week 0) to follow-up (week 4) assessments across groups. (C) Metacognitive (visuo-perceptual decision-making) task design (N = 210 trials). On each trial, participants were asked to judge and choose the sunflower that contained more seeds (i.e., higher number of dots) and then provide a confidence rating on their decision.

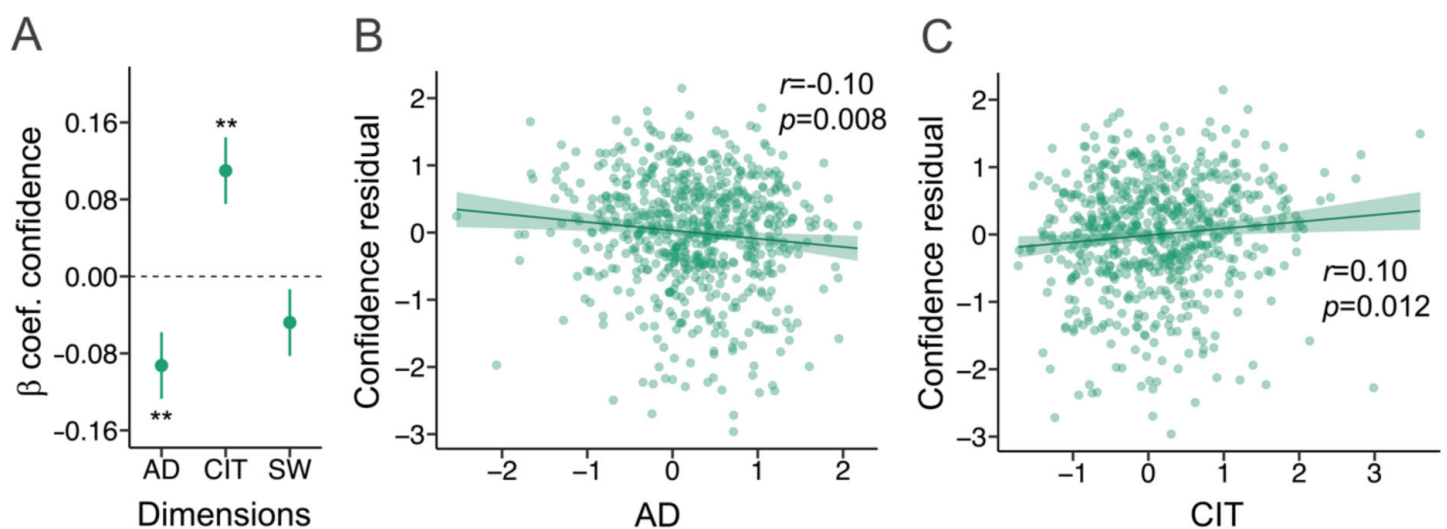


Figure 2

β = standardised beta coefficient, r = correlation coefficient, p = p-value, AD = Anxious-Depression, CIT = Compulsivity and Intrusive Thought, SW = Social Withdrawal. (A) AD and CIT were associated with metacognitive bias, while SW was not. The residual values for confidence (controlling for age, gender and education) were negatively associated with AD. (C) The residual values for confidence (controlling for age, gender and education) were positively associated with CIT.

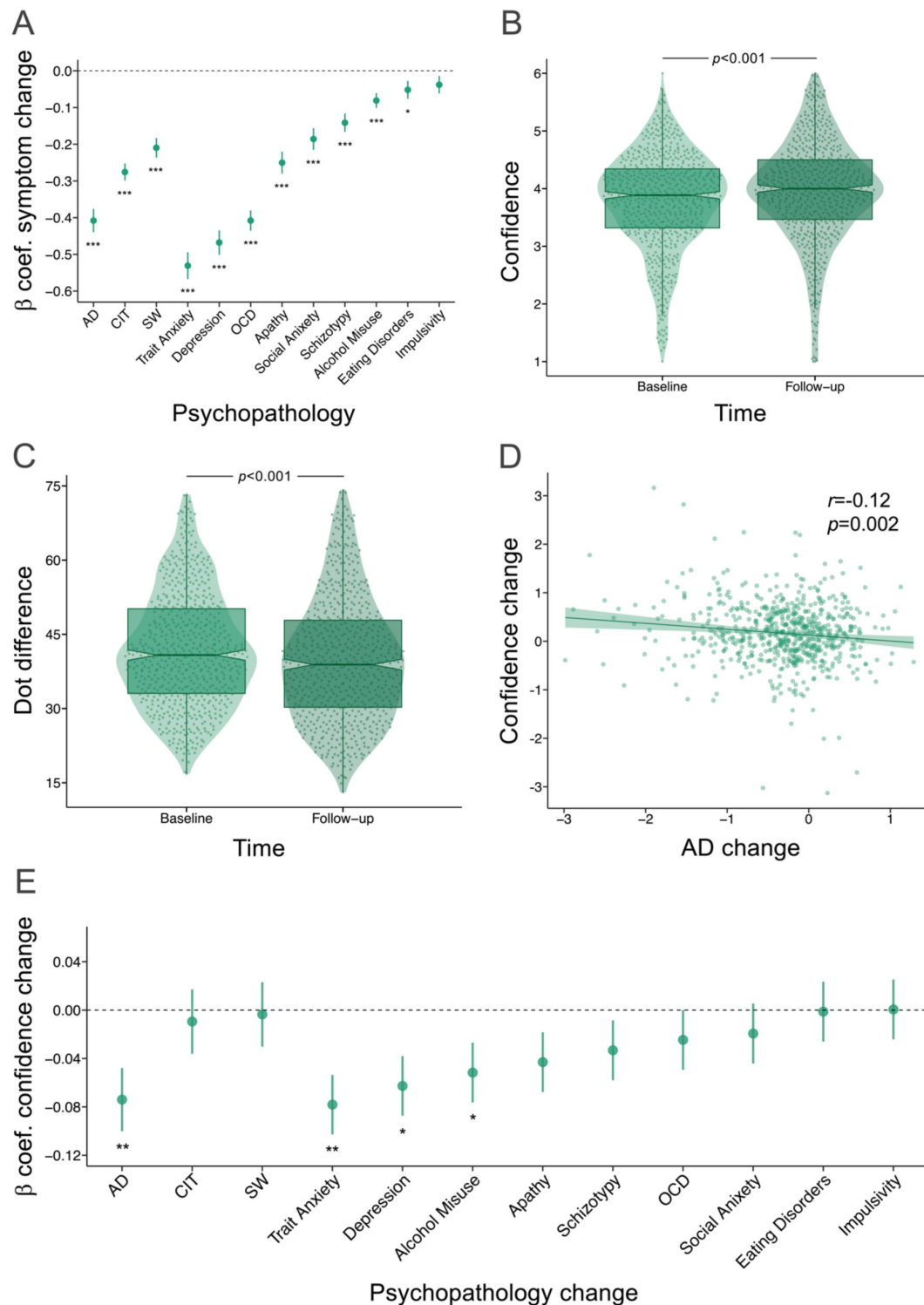


Figure 3

β = standardised beta coefficient, AD = Anxious-Depression, CIT = Compulsivity and Intrusive Thought, SW = Social Withdrawal, OCD = Obsessive compulsive disorder, r = correlation coefficient, p = p-value (unadjusted), *** = $p < 0.001$, ** = $p < 0.01$, * = $p < 0.05$. (A) Psychopathology symptoms improved with four weeks of iCBT. (B) Confidence was significantly higher and, (C) the task was more difficult at four-week follow-up. (D) Those with the largest improvements in AD had the greater increases in confidence. (E) Change in confidence also scaled with improvements in trait anxiety, depression and alcohol misuse.

Comparing iCBT, Antidepressant and Control Groups

Examining simple effects in the antidepressant arm, there was a significant reduction in anxious-depression from baseline to follow-up ($\beta=-0.61$, $SE=0.09$, $p<0.001$). Among controls, levels of anxious-depression did not significantly change ($\beta=0.10$, $SE=0.06$, $p=0.096$). Further details of transdiagnostic clinical changes for the antidepressant and controls groups are presented in **Figure 4A** and **Table S4**. When comparing the three groups directly, ANOVA analysis predicting anxious-depression scores with group and time as independent variables revealed a main effect of time ($F(1, 1632)=62.99$, $p<0.001$), a main effect of group ($F(2, 1632)=249.74$, $p<0.001$), and an interaction effect of group and time ($F(2, 1632)=9.23$, $p<0.001$).

With respect to confidence, tests of simple effects revealed that mean confidence significantly increased from baseline ($M=3.77$, $SD=0.88$) to follow-up ($M=4.07$, $SD=0.79$) in the antidepressant arm ($\beta=0.31$, $SE=0.08$, $p<0.001$) (**Figure 4B**). Among controls, there was no significant change in confidence from baseline ($M=3.68$, $SD=0.86$) to follow-up ($M=3.79$, $SD=0.92$) ($\beta=0.11$, $SE=0.07$, $p=0.103$), though it went in the same direction (**Figure 4B**). Predicting confidence scores using ANOVA analysis with group and time as independent variables revealed a main effect of time ($F(1, 1632)=16.26$, $p<0.001$), and no significant main effect of group ($F(2, 1632)=2.35$, $p=0.096$). The interaction effect of group and time on mean confidence was not significant ($F(2, 1632)=0.60$, $p=0.550$), suggesting that change in confidence did not differ across the three groups.

In the antidepressant arm, mean dot difference decreased from baseline ($M=41.2$, $SD=13.3$) to follow-up ($M=35.3$, $SD=13.1$) ($\beta=-5.91$, $SE=1.25$, $p<0.001$), indicating increased task difficulty. There was no significant change in task difficulty among controls from baseline ($M=43.0$, $SD=11.8$) to follow-up ($M=41.4$, $SD=13.6$) ($\beta=-1.64$, $SE=1.30$, $p=0.210$) (**Figure 4C**). With the three groups in the model, there was a significant main effect of time ($F(1, 1632)=15.17$, $p=0.001$) and group ($F(2, 1632)=4.56$, $p=0.011$) on mean dot difference. The interaction effect of time and group on mean dot difference was not significant ($F(2, 1632)=1.91$, $p=0.148$), suggesting no differences across the groups in task difficulty changes.

While our sample was underpowered to examine individual differences, we conducted an exploratory analysis examining the connection between changes in anxious-depression symptoms and changes in confidence in the antidepressant and controls groups. Although not significant, the association between change in confidence and change in anxious-depression was in the expected negative direction in the antidepressant arm ($r(80)=-0.10$, $p=0.381$), and among controls ($r(86)=-0.17$, $p=0.111$) (**Figure 4D**). When examining the effects of time, group and anxious-depression change on mean confidence, there was a significant interaction effect of time and anxious-depression change on mean confidence ($F(1, 1626)=4.04$, $p=0.045$), suggesting change in confidence was dependent on change in anxious-depression. There was no significant three-way interaction of anxious-depression change, time and group on mean confidence when comparing the three groups ($F(2, 1626)=0.08$, $p=0.928$), indicating that the significant association between confidence change and anxious-depression change was not specific to any group.

Discussion

Metacognitive biases are linked to transdiagnostic dimensions of mental health, but it is presently unclear if these biases are stable traits, or if they fluctuate alongside symptoms and change during the course of treatment (10). To answer these questions, we administered a previously validated adaptive task of metacognitive ability that controls for objective performance differences (26) in a large sample of individuals before and after four weeks of iCBT or antidepressant medications (21). As expected, a four-week course of iCBT or antidepressant medication led to transdiagnostic improvements in mental health (27–29). Alongside this, there was a significant increase in metacognitive confidence following four weeks of iCBT or antidepressant

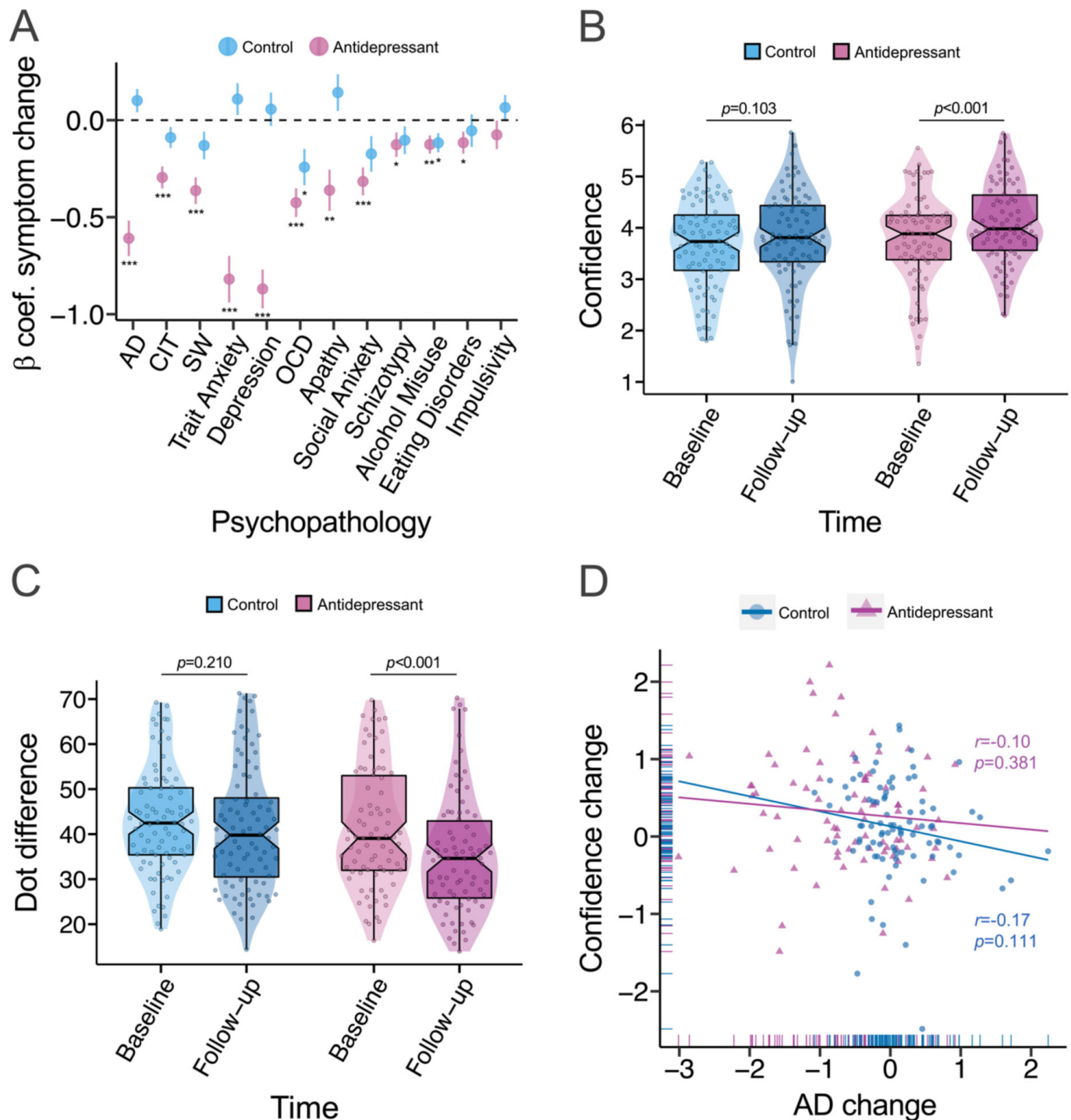


Figure 4

β = standardised beta coefficient, AD = Anxious-Depression, CIT = Compulsivity and Intrusive Thought, SW = Social Withdrawal, OCD = Obsessive compulsive disorder, r = correlation coefficient, p = p-value, *** = $p<0.001$, ** = $p<0.01$, * = $p<0.05$. (A) The majority of psychiatric scales improved in the antidepressant arm after 4 weeks of treatment, while the controls only had significant reductions in OCD symptoms and alcohol misuse at follow-up. (B) The larger increase in confidence in the antidepressant arm compared to controls was trended towards significant. (C) The antidepressant arm had a greater increase in task difficulty (a reduction in dot difference across stimuli) from baseline to follow-up, relative to controls. (D) Although not significant, the association between change in confidence and change in anxious-depression was in the expected negative direction in the antidepressant arm and among controls.

medication. Not simply a practice effect, we found that individuals in the iCBT arm with the greatest improvements in anxious-depression had the largest increase in confidence at follow-up. These findings suggest that metacognitive biases in anxious-depression are state-dependent. This builds on previous findings in small samples that have shown iCBT improves self-reported metacognitive self-beliefs (25 [↗](#)) and that metacognition can be altered through adaptive training (12 [↗](#)–15 [↗](#)).

At baseline, we replicated the previously observed bi-directional associations between metacognitive bias and anxious-depression and compulsivity and intrusive thought (7 [↗](#)–9 [↗](#), 11 [↗](#)). While higher levels of anxious-depression is associated with lower confidence, those with higher levels of compulsivity and intrusive thought have elevated confidence. This is a somewhat surprising dissociation, as compulsivity and anxious-depression are themselves positively correlated in the population. One way this can be reconciled is if the mechanisms underlying these opposing confidence biases are distinct. In anxious-depression, there appears to be more pervasive metacognitive biases that affect confidence in many domains and levels of a metacognitive hierarchy (spanning confidence in low level perceptual decisions to ideas of self-worth) (9 [↗](#), 10 [↗](#)). In contrast, inflated confidence in compulsivity may be based on more specific biases in learning and inference (8 [↗](#)).

The present study was observational and therefore did not randomly assign participants to a different treatments. To partially remediate that limitation, we included two smaller groups receiving antidepressant medication and a control group. Levels of transdiagnostic psychiatric dimensions remained stable across time among controls, while they significantly improved in the antidepressant arm. Similarly to iCBT, we found that confidence improved in the antidepressant group, but not among controls. The interaction, however, was not significant, meaning that we cannot reject the null hypothesis that confidence improved to the same degree across the three groups. As increased task difficulty among clinical groups was not significantly greater relative to controls, changes in task difficulty may simply reflect greater task familiarity at follow-up across groups, as opposed to gains in general perceptual performance among clinical arms. Examining the three groups together, the data suggests that confidence changes are unlikely to be treatment specific, rather, confidence fluctuates in tandem with anxious-depression. This was evident in an overall association between change in anxious-depression and change in confidence that was not modified by treatment arm. Additionally, levels of iCBT engagement and concurrent treatments did not bolster changes in confidence. Overall, the results indicated that metacognition fluctuates with anxious-depression state, regardless of treatment type or exposure. Future research with larger samples are required to address this definitively.

Limitations and Future Directions

Confidence change and anxious-depression change were significantly but weakly associated. Similarly, the relative change in confidence across treatment arms was small. Therefore, while tests of metacognitive confidence can inform theoretical models, like most cognitive tests, they are likely of limited utility in clinical practice, at least when used in isolation (19 [↗](#)). Given the complexity of mental health causes and presentations, multivariable models are needed to see practical value from such tests. We did not assess confidence or anxious-depression to treatment cessation and so the causal path and temporal dependence, if they exist, cannot be derived from these data. Future research should consider assessing metacognition and anxious-depression continuously through treatment, in order to elucidate the causal relationship between anxious-depression and metacognition with mediation analysis (25 [↗](#)). While this study examined changes in metacognition with iCBT generally, future research should examine if the strength of the association between confidence change and anxious-depression change is greater following iCBT modules targeting metacognition or following metacognitive intervention (4 [↗](#)). The iCBT programs in this study primarily targeted depression and anxiety, which may explain why changes in confidence did not scale with improvements in compulsivity. Future research is required to assess if treatments aimed at compulsive disorders decrease the over-confidence

commonly observed in those high in compulsivity and intrusive thought. As the antidepressant and control groups were much smaller than the iCBT arm, we were unable to compare changes in confidence across the types of antidepressant medications individuals received and we were underpowered more generally for individual differences analyses and multi-arm comparisons. Exploratory analyses were nonetheless presented and can form the basis for future investigations.

Conclusions

Our findings replicated the cross-sectional evidence that higher levels of anxious-depression are associated with under-confidence. We demonstrate that metacognitive confidence increases following four weeks of iCBT or antidepressant treatment. Overall, we observed that the greater the improvement in anxious-depression, the more confident participants became, which did not appear to be dependent on treatment type. This suggests that metacognitive biases in anxious-depression are state-dependent and might be normalised through clinical improvements.

Methods

Participants

Participants were recruited as part of the Precision in Psychiatry (PIP) study (21 [↗](#)), an observational, longitudinal study in which participants underwent a four-week course of iCBT or antidepressant medication. Further details of the PIP study procedures and ethical approval can be found in a prior publication (21 [↗](#)). A power analysis was carried out using effect sizes from a previous study examining cross-sectional associations between metacognition and anxious-depression, and compulsivity and intrusive thought (7 [↗](#)). Sample sizes of $N=454$ and $N=332$ respectively were required to detect these associations with 80% power. The sample sizes of the antidepressant arm and control group were smaller and used for secondary and more exploratory analyses.

iCBT Arm. Individuals initiating iCBT provided by SilverCloud Health were recruited from two sites: 1) the National Health Service Berkshire Foundation in the UK and 2) Aware mental health charity in Ireland. Participants included in the study either started their iCBT intervention ≤ 2 days prior to signing up, or provided a treatment start date in the near future, and scored ≥ 10 on the Work and Social Adjustment Scale (WSAS) at baseline, which indicated significant functional impairment due to clinical symptoms (34 [↗](#)). **Figure 1A** [↗](#) shows the disposition of participants throughout the study. $N=2404$ were screened, of whom $N=1496$ were eligible, $N=836$ completed baseline assessments and met inclusion criteria (detailed in the **Supplement**). A [↗](#) final $N=649$ completed and met inclusion criteria for follow-up assessments. While study follow-up data was collected after four weeks of treatment, iCBT could last up to 12 weeks (21 [↗](#)). The final sample was, on average, 32.2 years old ($SD=11.0$), mostly female ($n=501$, 77.4%), living in the United Kingdom ($n=546$, 84.4%), and had some or completed undergraduate level education ($n=342$, 52.9%) (**Table 1** [↗](#)).

Antidepressant Arm. Individuals were recruited globally using advertisements placed on Google search, in addition to social media platforms, mental health websites, local pharmacies and General Practitioner waiting rooms. Participants were included if they started or planned to start treatment ≤ 2 days of study sign-up, scored ≥ 10 on the WSAS at baseline, and provided a valid photograph of an antidepressant medication prescription. $N=270$ individuals were screened, of whom $N=174$ were eligible, $N=102$ completed and met inclusion criteria at baseline and a final $N=82$ had follow-up data (**Figure 1A** [↗](#)). Participants were mostly female ($n=60$, 73.2%), mean age=30.5 (10.5), were living in Ireland or the United Kingdom ($n=66$, 80.5%) and had some or completed undergraduate level education ($n=49$, 59.8%) (**Table 1** [↗](#)).

Characteristic	iCBT (<i>n</i> = 649) ^a	Antidepressant (<i>n</i> = 82)	Control (<i>n</i> = 88)	F/X ² (df)	<i>p</i>
Gender, No. (%)					0.414 ^b
Male	141 (21.8)	20 (24.4)	22 (25.0)		
Female	501 (77.4)	60 (73.2)	66 (75.0)		
Other	5 (0.8)	2 (2.4)	0 (0.0)		
Age, <i>M</i> (<i>SD</i>)	32.2 (11.0)	30.5 (10.5)	29.1 (12.0)	3.68 (2, 813)	0.026
Country of residence, No. (%)				211.73 (4)	<0.001
Ireland	89 (13.8)	32 (39.0)	58 (65.9)		
United Kingdom	546 (84.4)	34 (41.5)	24 (27.3)		
Other	12 (1.9)	16 (19.5)	6 (6.8)		
Highest level of education, No. (%)				7.11 (4)	0.130
Below undergraduate	147 (22.7)	12 (14.6)	13 (14.8)		
Some/completed undergraduate	342 (52.9)	49 (59.8)	57 (64.8)		
Above undergraduate	158 (24.4)	21 (25.6)	18 (20.5)		

^a Two participants in the iCBT arm were missing data for age, gender, country of residence and highest level of education.

^b Gender proportions were compared using Fisher's exact due to cell count < 5 in the 'Other' group.

Table 1

Baseline Sociodemographic Characteristics of Participants.

Control Group. Participants in the no treatment control group were recruited through university mailings lists and advertisements posted online and around Trinity College Dublin. Participants included in this arm scored <10 on an adapted version of the WSAS (where they rated functional impairment from their general problems rather than mental health problems) and self-reported that they had no current mental health problems and were not undergoing treatment for any mental health problems at the time of screening. N=444 individuals were screened, of whom N=245 were eligible, N=113 had baseline data and a final N=88 completed follow-up assessments and met inclusion criteria for the study (**Figure 1A**). Participants in the control group were matched for sociodemographic characteristics in the antidepressant arm, with no significant differences across the two groups in gender (Fisher's exact test $p=0.498$), or age (Welch's $t(166.66)=-0.81$, $p=0.420$), or levels of educational attainment ($\chi^2(2) = 0.66$, $p=0.718$). Participants in the control group came from the United Kingdom and Ireland (>93%), whereas the antidepressant arm was more international, with 20% coming from other countries ($\chi^2(2)=13.59$, $p=0.001$).

When comparing sociodemographic characteristics across the three study groups, there were no significant differences between groups in gender proportions (Fisher's exact test $p=0.414$), or levels of educational attainment ($\chi^2(4)=7.11$, $p=0.130$), as reported in **Table 1**. Age differed across the three arms ($F(2, 813)=3.68$, $p=0.026$), with post hoc Tukey tests indicating that mean age was higher in the iCBT arm ($M=32.2$, $SD=11.0$) compared to the controls ($M=29.1$, $SD=12.0$) ($p_{adj}=0.034$), but not when compared to the antidepressant arm ($M=30.5$, $SD=10.5$) ($p_{adj}=0.376$) (**Table 1**). The countries participants were living in varied across the study arms ($\chi^2(4)=211.73$, $p<0.001$), as there was a higher proportion of individuals in the iCBT arm living in the UK ($n=546$, 84.4%) when compared to the antidepressant arm ($n=34$, 41.5%) and the control group ($n=24$, 27.3%) (**Table 1**).

Procedure

Figure 1B shows an overview of the study design, including the assessments involved at each timepoint. For the purposes of this study, we focused on a select set of sociodemographic characteristics (gender, age, country of residence, level of educational attainment), self-reported psychiatric questionnaires, metacognitive task performance and treatment data from the PIP study.

Self-Reported Psychiatric Questionnaires. Participants completed nine self-report questionnaires at baseline and follow-up that assess a variety of psychiatric symptoms (see **Supplement** for questionnaire details).

Metacognitive Task. Participants completed a visuo-perceptual decision-making task (26) to assess metacognition (**Figure 1C**). On each of the 210 trials, participants were asked to judge and choose the sunflower that contained more seeds (i.e., higher number of dots) and then provide a confidence rating on their decision. Mean accuracy was tightly controlled in this task using a 'two-down one-up' staircase procedure, in which equal changes in dot difference occurred after each incorrect response and after two consecutive correct responses, detailed further in the **Supplement**.

Treatment Data. Objective indicators of treatment engagement were provided by SilverCloud for 640 participants in the iCBT arm, which comprised of percentage of program viewed, time (minutes) spent in the program, and program type (listed in the **Supplement**). Information on concurrent treatment and treatment adherence across each group are reported in the **Supplement**.

Data Preparation and Analysis

Self-Reported Psychiatric Questionnaire. Individual scores on dimensions of anxious-depression, compulsivity and intrusive thought, and social withdrawal were calculated by multiplying each of the 209 item scores on the nine self-report clinical scales by the 209 corresponding item weights from a previously published factor analysis on these scales (35). Dimension scores were then scaled to centre on zero, with higher scores indicating higher levels of transdiagnostic psychopathology. We additionally tested for consistency in the factor structure in this dataset by carrying out a factor analysis on these items and testing their correlation with the prior work in unselected samples (reported in the Supplement). Careless or inattentive responding to questionnaires was detected with embedded ‘catch’ items across questionnaires, detailed in the Supplement.

Metacognition Task. The perceptual decision-making task performance was used to quantify our primary cognitive outcome measure, metacognitive bias, the mean confidence rating across trials. Other task measures include mean dot difference (task difficulty), mean accuracy and mean reaction time (in seconds), which are described further in the Supplement. As metacognitive efficiency was not previously associated with transdiagnostic dimensions cross-sectionally (7,9,11), results pertaining to efficiency are reported separately in the Supplement.

Statistical Analysis. We tested for relationships between baseline task measures and the psychiatric symptom dimensions using linear regression analysis, controlling for age, gender, and education. To examine pre-post changes, we carried out linear mixed-effects models with measures of metacognition or psychopathology as the dependent variable, time (baseline=0; follow-up=1) as the independent variable and participants as random effects. To determine the association between change in confidence and change in anxious-depression, we used Pearson correlation analyses. Exploratory linear regression analyses tested the specificity of the effects, replacing anxious-depression with each of the measures of psychopathology in turn as follows: Mean confidence ~ time*dimension/psychiatric scale score. We additionally ran regression analyses to test if concurrent treatment or the degree of objective engagement in iCBT interacted with the effect of time on mean confidence. Exploratory ANOVA analyses were also conducted to compare changes in anxious-depression, task difficulty and confidence across the three arms directly. For all tests, statistical significance was defined as $p < 0.05$, with two-tailed p -values used. All regressors were scaled as Z scores to compare the regression coefficients of independent variables within each model. For exploratory regression analyses, adjustments for multiple comparisons were not conducted (36). The code and data to reproduce statistical analyses are available at <https://osf.io/89xzq/>.

Acknowledgements

This work was funded by a fellowship awarded to CGM from MQ: transforming mental health (MQ16IP13). CGM holds additional funding from Science Foundation Ireland’s Frontiers for the Future Scheme (19/FFP/6418), and a European Research Council (ERC) Starting Grant (ERC-H2020-HABIT). The PhD studentship of CAF is funded by the Government of Ireland Postgraduate Scholarship Programme (GOIPG/2020/662). The authors thank all the participants for their involvement in this study. We thank the AWARE charity and the Berkshire foundation trust that supported recruitment for the iCBT arm. We thank the individual pharmacies and General Practitioner services for their support in recruiting the antidepressant arm.

CAF conceived the study. CTL, CMG, KES, JP, DR, VOK and SH designed the protocol. VOK provided clinical support. CAF, CTL, AH and KL acquired the data. TXFS contributed analysis tools. CAF, CTL and CMG analysed the data. CAF and CMG wrote the first draft of the paper. All authors revised the paper.

Additional File 1. Supplementary Information

Supplemental methods

Sociodemographic Characteristics

Participants in the control group were matched for sociodemographic characteristics in the antidepressant arm, with no significant differences across the two groups in gender (Fisher's exact test $p=0.498$), or age (Welch's $t(166.66)=-0.81$, $p=0.420$), or levels of educational attainment ($\chi^2(2) = 0.66$, $p=0.718$). Participants in the control group came from the United Kingdom and Ireland (>93%), whereas the antidepressant arm was more international, with 20% coming from other countries ($\chi^2(2)=13.59$, $p=0.001$).

When comparing sociodemographic characteristics across the three study arms, there were no significant differences between groups in gender proportions (Fisher's exact test $p=0.498$), or levels of educational attainment ($\chi^2(4)=7.11$, $p=0.130$) (Table S1). Age differed across the three arms ($F(2, 813)=3.68$, $p=0.026$), with post hoc Tukey tests indicating that mean age was higher in the iCBT arm ($M=32.2$, $SD=11.0$) compared to the controls ($M=29.1$, $SD=12.0$) ($p_{adj}=0.034$), but not when compared to the antidepressant arm ($M=30.5$, $SD=10.5$) ($p_{adj}=0.376$). The countries participants were living in varied across the study arms ($\chi^2(4)=211.73$, $p<0.001$), as there was a higher proportion of individuals in the iCBT arm living in the UK ($n=546$, 84.4%) when compared to the antidepressant arm ($n=34$, 41.5%) and the control group ($n=24$, 27.3%).

Self-Report Clinical Scales

Participants completed nine standard self-report questionnaires at baseline and four-week follow-up that assess a variety of psychiatric symptoms, including depression (Zung Self-Rating Depression Scale) (1), trait anxiety (State Trait Anxiety Inventory) (2), schizotypy (Short Scales for Measuring Schizotypy) (3), impulsivity (Barratt Impulsiveness Scale 11) (4), obsessive-compulsive disorder (OCD) (Obsessive-Compulsive Inventory-Revised) (5), social anxiety (Liebowitz Social Anxiety Scale) (6), eating disorders (Eating Attitudes Test) (7), apathy (Apathy Evaluation Scale) (8), and alcohol misuse (Alcohol Use Disorders Identification Test) (9). Anxious-depression, compulsive behaviour and intrusive thought, and social withdrawal were derived as psychiatric symptom dimensions by applying 209 item weights that were previously conducted from factor analysis on these scales (10). Additionally, participants completed the Working and Social Adjustment Scale (WSAS) (11) at baseline and follow-up. Individuals in the iCBT and antidepressant arms received the original WSAS scale, which asks them to rate their level of functional impairment from mental health problems. Individuals in the control group received an adapted version of the WSAS, which asks them to rate their level of impairment from 'their problems' more generally.

Characteristic	iCBT (<i>n</i> = 649) ^a	Antidepressant (<i>n</i> = 82)	Control (<i>n</i> = 88)	F/X ² (df)	<i>p</i>
Gender, No. (%)					0.414 ^b
Male	141 (21.8)	20 (24.4)	22 (25.0)		
Female	501 (77.4)	60 (73.2)	66 (75.0)		
Other	5 (0.8)	2 (2.4)	0 (0.0)		
Age, <i>M</i> (<i>SD</i>)	32.2 (11.0)	30.5 (10.5)	29.1 (12.0)	3.68 (2, 813)	0.026
Country of residence, No. (%)				211.73 (4)	<0.001
Ireland	89 (13.8)	32 (39.0)	58 (65.9)		
United Kingdom	546 (84.4)	34 (41.5)	24 (27.3)		
Other	12 (1.9)	16 (19.5)	6 (6.8)		
Highest level of education, No. (%)				7.11 (4)	0.130
Below undergraduate	147 (22.7)	12 (14.6)	13 (14.8)		
Some/completed undergraduate	342 (52.9)	49 (59.8)	57 (64.8)		
Above undergraduate	158 (24.4)	21 (25.6)	18 (20.5)		

^aTwo participants in the iCBT arm were missing data for age, gender, country of residence and highest level of education.

^bGender proportions were compared using Fisher's exact due to cell count < 5 in the 'Other' group.

Table S1

Baseline Sociodemographic Characteristics of Participants.

Metacognitive Task Design

On each trial, participants were shown a fixation cross for 1000 milliseconds (ms), followed by two sunflowers with a varying number of seeds for 300 ms. Participants then had unlimited time to make a judgement on which of two sunflower stimuli contained more seeds and then rated their confidence in each judgement, on a scale from '1=Guessing' to '6=Certain'. There was a total of 210 trials, divided equally into five blocks. While participants were given feedback on which sunflower they had chosen for 500 ms, there was no trial-level feedback provided on their performance. Split-half reliability was high for baseline mean confidence, as indicated by the correlation between odd and even trials across each participant in the iCBT arm ($r(647)=0.98$, $p<0.001$).

Mean accuracy was tightly controlled in this task using a 'two-down one-up' staircase procedure, in which equally sized changes in dot difference occurred after each incorrect response and after two consecutive correct responses. This maintained objective performance across all participants at a desired level of 70% correct, crucial for estimating metacognition without any confound of real performance differences. Changes in dot difference were calculated using log-space, with a start log difference of 4.2 (+70 dots). Differences in step size changed by ± 0.4 for the first five trials, ± 0.2 for the next five trials and ± 0.1 for the remainder of the task. Dot difference on each trial could range from six dots (1.79 in log-space, the most difficult to discriminate) to 81 dots (4.39 in log-space, the easiest to discriminate).

For analyses, mean reaction time to stimuli choice across trials was measured in seconds and task accuracy was calculated as the mean proportion of correct responses across trials. Task difficulty was operationalised as the mean dot difference between the sunflower stimuli across trials, with a greater dot difference reflecting lower task difficulty.

Metacognitive Efficiency (M-Ratio)

Metacognitive sensitivity (meta- d') is the extent to which confidence ratings discriminate between correct and incorrect trials, while d' is indicated as objective performance. Metacognitive efficiency (M-Ratio) is calculated as the ratio of metacognitive sensitivity to objective performance (i.e., meta- d'/d'). M-Ratio was calculated in a hierarchical Bayesian framework using a freely available toolbox (HMeta toolbox) (12 [12](#)), <http://github.com/sm Fleming/HMM>, accessed June 2022. An M-Ratio value of 1 indicates that confidence was fully informed by accessing the total perceptual information available. In the iCBT arm, we found low split-half reliability for baseline M-Ratio, when correlating odd and even task trials ($r(647)=0.30$, $p<0.001$). The split-half reliability of M-Ratio may have been impacted by the staircase procedure, as there was a significant negative association between mean accuracy on odd and even trials ($r(647)=-0.25$, $p<0.001$). Additionally, M-Ratio values were confounded by task accuracy ($\beta=-3.19$, $SE=0.62$, $p<0.001$), which violated the assumption that M-Ratio is independent of accuracy. Given that our task has 210 trials, these findings are consistent with the recent report that existing metacognitive efficiency measures are neither independent of type 1 performance nor reliable when tasks have less than 400 trials (13 [13](#)). Thus, all analyses pertaining to M-Ratio should be considered in the context of these limitations.

Treatment Data

Every week, participants self-reported if they had started any new concurrent medications and/or psychological treatments for mental health. Participants also reported if they had adhered to iCBT or their antidepressant medications each week. Starting new medication and/or psychological treatments and treatment nonadherence during the study were not grounds for exclusion in any group.

In relation to iCBT programs undertaken by participants in the iCBT arm, *Space from Depression* was the most common iCBT program ($n=158$, 24.7%), followed by *Space from Generalised Anxiety Disorder* ($n=96$, 15.0%), *Life Skills Online* ($n=89$, 13.9%), *Space from Anxiety* ($n=84$, 13.1%), and *Space from Depression and Anxiety* ($n=77$, 12.0%). Given the overlapping content of *Space from Generalised Anxiety Disorder* and *Space from Anxiety*, these programs were merged into one category ‘*Space from Anxiety*’ ($n=180$, 28.1%) for analyses. The ‘*Other*’ category ($n=136$, 21.3%) comprised of the following iCBT programs: *Space from Stress* ($n=33$, 5.2%), *Space from Social Anxiety* ($n=20$, 3.1%), *Space for Resilience* ($n=20$, 3.1%), *Space for Perinatal Wellbeing* ($n=18$, 2.8%), *Space in Chronic Pain from Depression and Anxiety* ($n=12$, 1.9%), *Space from Health Anxiety* ($n=9$, 1.4%), *Space from OCD* ($n=7$, 1.1%), *Space from Panic* ($n=6$, 0.9%), *Space from Phobia* ($n=5$, 0.8%), *Space for Sleep* ($n=4$, 0.6%), *Space in Lung Conditions from Depression and Anxiety* ($n=1$, 0.2%) and *Space from Money Worries* ($n=1$, 0.2%), which were merged together into one category for analysis due to low N in each program.

Factor Analysis: Transdiagnostic Symptom Dimensions

Using baseline data from the iCBT arm, we aimed to replicate the three-factor structure for transdiagnostic dimensions of items from nine psychiatric questionnaires reported previously (10). A heterogeneous correlation matrix of the 209 items from the nine questionnaires was generated using the hector function in the polycor package in R. We conducted maximum likelihood factor analysis using the fa function from the psych package in R. We specified that three factors would be extracted using regression with an oblique rotation. We applied the same labels previously used for the three factors extracted from our data: 1) anxious-depression; 2) compulsivity and intrusive thought, and 3) social withdrawal (10). Items from scales measuring depression, apathy, and anxiety loaded most strongly onto the anxious-depression factor. Items relating to OCD, impulsivity, and schizotypy symptoms had the highest loading for compulsivity and intrusive thought. Items that loaded on to social withdrawal came largely from the social anxiety scale. The item weights from our data were highly correlated with item weights reported by Gillan and colleagues (10) for anxious-depression ($r(207)=0.89$, $p<0.001$), compulsivity and intrusive thought ($r(207)=0.81$, $p<0.001$), and social withdrawal ($r(207)=0.96$, $p<0.001$), which supported replication of the dimension factor structure in the present study (Figure S1).

Data Quality Checks: Task and Clinical Scales

We employed a number of exclusion criteria to ensure high data quality from the metacognitive task. Firstly, we planned to exclude participants if they selected the right or left sunflower on greater than 95% of trials, but none met this criterion. Second, we excluded participants with mean accuracy less than 0.60 and greater than 0.85, indicating the staircase procedure did not converge within acceptable bounds ($n=38$ at baseline and $n=34$ at follow-up excluded in the iCBT arm; $n=8$ at baseline and $n=7$ at follow-up excluded in the antidepressant arm; $n=6$ at baseline and $n=5$ at follow-up excluded in the control group).

To determine the proportion of careless/inattentive responders on the self-report clinical questionnaire, we included ‘catch’ questions that were embedded in the OCIR (*If you are paying attention to these questions, please select “A little”*) and the WSAS (*If you are paying attention to these questions, please select “Not at all”*) at baseline and in the WSAS at all subsequent assessments. In the iCBT arm, 54 (8.3%) participants failed at least one of the checks, with just nine (1.4%) failing more than one. Additionally, eight (9.8%) individuals in the antidepressant arm and seven (8.0%) in the control group failed at least one of the catch items. Given the small number of participants that failed more than one attention checks and that all those participants passed the task exclusion criteria, these individuals were retained for subsequent analyses.

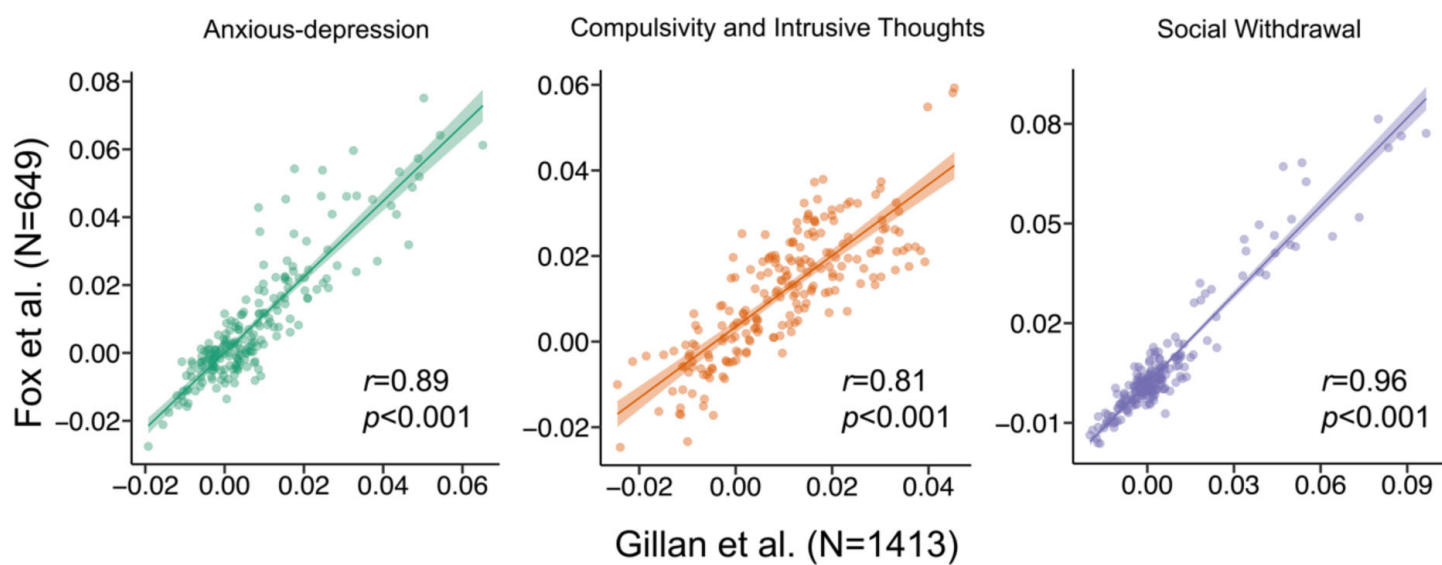


Figure S1

Correlation between Items Weights from Factor Analysis of Transdiagnostic Dimensions.

Supplemental results

Cross-sectional Baseline Findings: iCBT

At baseline, age was not significantly associated with mean confidence ($\beta=0.003$, $SE=0.003$, $p=0.401$). Males had significantly higher confidence ($M=3.91$, $SD=0.81$), than females ($M=3.75$, $SD=0.86$) ($\beta=0.08$, $SE=0.04$, $p=0.044$). Mean confidence was significantly lower among those with educational attainment above undergraduate level ($M=3.73$, $SD=0.89$), when compared to participants with attainment below undergraduate ($M=3.94$, $SD=0.87$) ($\beta=0.21$, $SE=0.10$, $p=0.036$). Mean confidence did not significantly differ when comparing those who had educational attainment above undergraduate ($M=3.73$, $SD=0.89$) to those who had some or had completed an undergraduate degree ($M=3.74$, $SD=0.83$) ($\beta=0.01$, $SE=0.08$, $p=0.921$). When comparing task measures, mean confidence was not significantly associated with mean accuracy ($\beta=-1.29$, $SE=1.42$, $p=0.366$), mean dot difference (task difficulty) ($\beta=0.002$, $SE=0.002$, $p=0.406$), mean response time ($\beta=6.47$, $SE=15.13$, $p=0.669$), or M-Ratio ($\beta=0.11$, $SE=0.09$, $p=0.218$).

Controlling for age, gender and education in the following baseline analyses, none of the transdiagnostic psychiatric dimensions were associated with overall task accuracy (anxious-depression, $\beta=0.001$, $SE=0.001$, $p=0.260$; compulsivity and intrusive thought, $\beta<-0.001$, $SE=0.001$, $p=0.948$; social withdrawal, $\beta<-0.001$, $SE=0.001$, $p=0.656$), consistent with the functioning of the staircase. While mean dot difference (task difficulty) was not associated with anxious-depression ($\beta=-0.58$, $SE=0.47$, $p=0.218$) or social withdrawal ($\beta=0.28$, $SE=0.48$, $p=0.560$), task difficulty was lower among those with higher levels of compulsivity and intrusive thought ($\beta=1.02$, $SE=0.47$, $p=0.032$). Mean response time was not associated with anxious-depression ($\beta=-0.04$, $SE=0.09$, $p=0.673$), compulsivity and intrusive thought ($\beta=-0.01$, $SE=0.09$, $p=0.882$) or social withdrawal ($\beta=-0.11$, $SE=0.09$, $p=0.214$). M-Ratio was also not associated with any of the transdiagnostic psychiatric dimensions (anxious-depression, $\beta=0.01$, $SE=0.02$, $p=0.743$; compulsivity and intrusive thought, $\beta=-0.004$, $SE=0.02$, $p=0.783$; social withdrawal, $\beta=-0.03$, $SE=0.02$, $p=0.106$).

Treatment Findings: iCBT

All scale scores and dimensions significantly improved from baseline to four-week follow-up, except for impulsivity (**Table S2** [↗](#), as referenced in the main results). Additionally, functional impairment reduced, as indicated by a decrease in WSAS score from baseline ($M=18.78$, $SD=6.35$) to follow-up ($M=17.06$, $SD=8.41$) ($\beta=-1.71$, $SE=0.30$, $p<0.001$). Mean accuracy did not change from baseline ($M=0.71$, $SD=0.02$) to follow-up ($M=0.71$, $SD=0.02$) ($\beta=0.001$, $SE=0.001$, $p=0.299$). Mean response time for stimuli choice reduced from baseline ($M=0.93$, $SD=2.22$) to follow-up ($M=0.69$, $SD=0.91$) ($\beta=-0.24$, $SE=0.09$, $p=0.010$). Metacognitive efficiency increased from baseline ($M=0.81$, $SD=0.38$) to follow-up ($M=0.86$, $SD=0.51$) ($\beta=0.05$, $SE=0.02$, $p=0.007$).

Change in confidence scaled with changes in anxious-depression ($\beta=-0.07$, $SE=0.03$, $p=0.005$), trait anxiety ($\beta=-0.08$, $SE=0.02$, $p=0.002$), depression ($\beta=-0.06$, $SE=0.02$, $p=0.011$) and alcohol misuse ($\beta=-0.05$, $SE=0.02$, $p=0.037$), but not significantly with any other psychiatric dimensions or scales (**Table S3** [↗](#), as referenced in the main results). Confidence changes did not differ across the different types of iCBT programs, when the Space from Depression and Anxiety program was compared pairwise to Space from Depression ($\beta=-0.09$, $SE=0.17$, $p=0.594$), Space from Anxiety ($\beta=-0.15$, $SE=0.17$, $p=0.376$), Life Skills ($\beta=-0.15$, $SE=0.19$, $p=0.435$) or the 'other' category comprising miscellaneous iCBT programs ($\beta=-0.03$, $SE=0.18$, $p=0.881$).

Treatment Adherence and Concurrent Treatment

Treatment adherence was high by week 3 in both clinical arms, with over $\geq 95\%$ of the iCBT arm still undergoing treatment (i.e., 98% at weekly check-in (WCI) 1, 97% at WCI 2 and 95% at WCI 3) and $\geq 99\%$ of the antidepressant group reported still taking antidepressant medication (i.e., 100% at

Outcome	β	SE	t	p	effect size (r^2)
Dimensions					
AD	-0.41	0.03	-12.80	<0.001	0.04
CIT	-0.28	0.02	-11.85	<0.001	0.02
SW	-0.21	0.03	-7.84	<0.001	0.01
Scale scores					
Trait Anxiety	-0.53	0.04	-14.53	<0.001	0.07
Depression	-0.47	0.03	-13.94	<0.001	0.05
OCD	-0.41	0.03	-14.80	<0.001	0.04
Apathy	-0.25	0.03	-8.34	<0.001	0.02
Social Anxiety	-0.19	0.03	-6.32	<0.001	0.009
Schizotypy	-0.14	0.03	-5.65	<0.001	0.005
Alcohol Misuse	-0.08	0.02	-3.97	<0.001	0.002
Eating Disorders	-0.05	0.02	-2.14	0.033	<0.001
Impulsivity	-0.04	0.02	-1.58	0.114	<0.001

β = standardised beta coefficient, SE = standardised error, t = t-value, p = p-value (unadjusted), AD = Anxious-depression, CIT = Compulsivity and intrusive thought, SW = Social withdrawal, OCD = Obsessive compulsive disorder.

Table S2

Changes in Psychiatric Dimensions and Scale Scores from Baseline to Follow-up in the iCBT Arm.

Outcome	β	SE	t	p
Dimensions				
AD	-0.07	0.03	-2.83	0.005
CIT	-0.01	0.03	-0.36	0.723
SW	-0.004	0.03	-0.13	0.894
Scale scores				
Trait Anxiety	-0.08	0.02	-3.18	0.002
Depression	-0.06	0.02	-2.54	0.011
OCD	-0.05	0.02	-2.09	0.037
Apathy	-0.04	0.02	-1.74	0.082
Social Anxiety	-0.03	0.02	-1.34	0.180
Schizotypy	-0.02	0.02	-0.99	0.320
Alcohol Misuse	-0.02	0.02	-0.78	0.434
Eating Disorders	-0.001	0.02	-0.05	0.959
Impulsivity	-0.001	0.02	0.02	0.981

β = standardised beta coefficient, SE = standardised error, t = t-value, p = p-value (unadjusted), AD = Anxious-depression, CIT = Compulsivity and intrusive thought, SW = Social withdrawal, OCD = Obsessive compulsive disorder.

Table S3

The Interaction Effect of Time and Psychiatric Dimension/Scale Change on Mean Confidence in the iCBT Arm.

WCI 1 and WCI 2 and 99% at WCI 3). In the antidepressant arm, four participants altered the dosage of their medication during the study participation (n=2 took less than prescribed and n=2 took more than prescribed). In terms of concurrent treatment, 175 (27.0%) in the iCBT arm were receiving another treatment during the study, of which 48 (7.4%) were taking concurrent medication for a mental health problem and 145 (22.3%) were receiving a concurrent form of psychotherapy. For the antidepressant group, 33 (40.2%) were receiving another treatment during the study, with 6 (7.3%) taking at least one other medication for a mental health problem and 2 (39.0%) were receiving some form of psychotherapy. Thus, there were partial overlaps in the treatments received across the two clinical arms. In the antidepressant arm, the effect of time on confidence was not dependent on receiving concurrent treatment ($\beta=0.13$, $SE=0.16$, $p=0.403$). Within the control group, two participants (2.3%) started taking medication during the study and none initiated psychotherapy.

Antidepressant and Control Arms: Clinical Change

All scale scores and dimensions significantly improved from baseline to follow-up in the antidepressant arm, except impulsivity (**Table S4** [↗](#), as referenced in the main results). Additionally, WSAS scores significantly reduced among those in the antidepressant arm ($\beta=-0.43$, $SE=0.12$, $p<0.001$). Among controls, OCD symptoms and alcohol misuse reduced at follow-up (**Table S4** [↗](#), as referenced in the main results). WSAS scores, which indicated impairments due to general life among controls, significantly increased from baseline to follow-up ($\beta=0.55$, $SE=0.11$, $p<0.001$).

Outcome	Antidepressant Arm				Controls			
	β	SE	t	p	β	SE	t	p
Dimensions								
AD	-0.70	0.11	-6.65	<0.001	0.13	0.08	1.68	0.096
CIT	-0.30	0.06	-5.22	<0.001	-0.12	0.07	-1.65	0.103
SW	-0.35	0.07	-5.27	<0.001	-0.16	0.09	-1.83	0.071
Scale scores								
Trait Anxiety	-0.82	0.12	-6.83	<0.001	0.11	0.08	1.32	0.189
Depression	-0.87	0.10	-8.68	<0.001	0.06	0.09	0.66	0.511
OCD	-0.43	0.07	-5.73	<0.001	-0.24	0.09	-2.58	0.012
Apathy	-0.36	0.11	-3.40	0.001	0.14	0.09	1.50	0.137
Social Anxiety	-0.32	0.07	-4.39	<0.001	-0.17	0.09	-1.90	0.060
Schizotypy	-0.13	0.06	-2.00	0.049	-0.10	0.07	-1.44	0.152
Alcohol Misuse	-0.13	0.05	-2.67	0.009	-0.12	0.05	-2.37	0.021
Eating Disorders	-0.12	0.06	-2.04	0.045	-0.05	0.08	-0.64	0.522
Impulsivity	-0.08	0.07	-1.03	0.306	0.06	0.06	1.01	0.314

β = standardised beta coefficient, SE = standardised error, t = t-value, p = p-value (unadjusted), AD = Anxious-depression, CIT = Compulsivity and intrusive thought, SW = Social withdrawal, OCD = Obsessive Compulsive Disorder.

Table S4

Changes in Psychiatric Dimensions and Scale Scores from Baseline to Follow-up in Antidepressant and Control Arms.

References

1. Fleming SM, Lau HC (2014) **How to measure metacognition** *Front Hum Neurosci* **8**
2. Fleming SM, Daw ND (2017) **Self-evaluation of decision-making: a general bayesian framework for metacognitive computation** *Psychol Rev* **124**:91–114
3. Fleming SM, Dolan RJ, Frith CD (2012) **Metacognition: computation, biology and function** *Philos Trans R Soc Lond B Biol Sci* **367**:1280–6
4. Philipp R, Kriston L, Kühne F, Härter M, Meister R (2020) **Concepts of metacognition in the treatment of patients with mental disorders** *J Ration-Emotive Cogn-Behav Ther* **38**:173–83
5. Hoven M, Lebreton M, Engelmann JB, Denys D, Luigjes J, van Holst RJ (2019) **Abnormalities of confidence in psychiatry: an overview and future perspectives** *Transl Psychiatry* **9**
6. Wise T, Robinson O, Gillan C (2022) **Identifying transdiagnostic mechanisms in mental health using computational factor modeling [Internet]** *Biol Psychiatry [Pre-proof]* **0**
7. Rouault M, Seow T, Gillan CM, Fleming SM (2018) **P Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance** *Biol Psychiatry* **84**:443–51
8. Seow TXF, Gillan CM (2020) **Transdiagnostic phenotyping reveals a host of metacognitive deficits implicated in compulsivity** *Sci Rep* **10**
9. Hoven M, Denys D, Rouault M, Luigjes J, Holst R van (2022) **Hoven M, Denys D, Rouault M, Luigjes J, Holst R van. How do confidence and self-beliefs relate in psychopathology: a transdiagnostic approach [Internet]. PsyArXiv; 2022 [cited 2022 Aug 22]. Available from: <https://psyarxiv.com/d45gn/>**
10. Seow TXF, Rouault M, Gillan CM, Fleming SM (2021) **How local and global metacognition shape mental health** *Biol Psychiatry* **90**:436–46
11. Benwell CSY, Mohr G, Wallberg J, Kouadio A, Ince RAA (2022) **Psychiatrically relevant signatures of domain-general decision-making and metacognition in the general population** *Npj Ment Health Res* **1**:1–17
12. Carpenter J, Sherman MT, Kievit RA, Seth AK, Lau H, Fleming SM (2019) **Domain-general enhancements of metacognitive ability through adaptive training** *J Exp Psychol Gen* **148**:51–64
13. Engeler NC, Gilbert SJ (2020) **The effect of metacognitive training on confidence and strategic reminder setting** *PloS One* **15**
14. Jelinek L, Van Quaquebeke N, Moritz S (2017) **Cognitive and metacognitive mechanisms of change in metacognitive training for depression** *Sci Rep* **7**
15. Lysaker PH, Gagen E, Moritz S, Schweitzer RD (2018) **Metacognitive approaches to the treatment of psychosis: a comparison of four approaches** *Psychol Res Behav Manag* **11**:341–51

16. Eisenacher S, Rausch F, Ainser F, Mier D, Veckenstedt R, Schirmbeck F (2015) **Investigation of metamemory functioning in the at-risk mental state for psychosis** *Psychol Med* **45**:3329–40
17. Gawęda Ł, Li E, Lavoie S, Whitford TJ, Moritz S, Nelson B (2018) **Impaired action self-monitoring and cognitive confidence among ultra-high risk for psychosis and first-episode psychosis patients** *Eur Psychiatry J Assoc Eur Psychiatr* **47**:67–75
18. Moeller SJ, Fleming SM, Gan G, Zilverstand A, Malaker P, d'Oleire Uquillas F (2016) **Metacognitive impairment in active cocaine use disorder is associated with individual differences in brain structure** *Eur Neuropsychopharmacol* **26**:653–62
19. Mogoșe C, Cobeianu O, David O, Giosan C, Szentagotai A (2017) **Internet-based psychotherapy for adult depression: what about the mechanisms of change?** *J Clin Psychol* **73**:5–64
20. Webb CA, Rosso IM, Rauch SL (2017) **Internet-based cognitive-behavioral therapy for depression: current progress and future directions** *Harv Rev Psychiatry* **25**:114–22
21. Lee CT, Palacios J, Richards D, Hanlon AK, Lynch K, Harty S (2023) **The Precision in Psychiatry (PIP) study: Testing an internet-based methodology for accelerating research in treatment prediction and personalisation** *BMC Psychiatry* **23**
22. Andersson G, Carlbring P, Rozental A (2019) **Response and remission rates in internet-based cognitive behavior therapy: an individual patient data meta-analysis** *Front Psychiatry* **10**
23. Eilert N, Enrique A, Wogan R, Mooney O, Timulak L, Richards D (2021) **The effectiveness of Internet-delivered treatment for generalized anxiety disorder: An updated systematic review and meta-analysis** *Depress Anxiety* **38**:196–219
24. Karyotaki E, Efthimiou O, Miguel C, Berman FM, Furukawa TA, Cuijpers P (2021) **Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis** *JAMA Psychiatry* **78**:361–71
25. Newby JM, Williams AD, Andrews G (2014) **Reductions in negative repetitive thinking and metacognitive beliefs during transdiagnostic internet cognitive behavioural therapy (iCBT) for mixed anxiety and depression** *Behav Res Ther* **59**:52–60
26. Fleming SM, Ryu J, Golfinos JG, Blackmon KE (2014) **Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions** *Brain* **137**:2811–22
27. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y (2018) **Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis** *Lancet Lond Engl* **391**:1357–66
28. Gøtzsche PC, Dinnage O (2020) **What have antidepressants been tested for? A systematic review** *Int J Risk Saf Med* **31**:157–63
29. Skapinakis P, Caldwell DM, Hollingworth W, Bryden P, Fineberg NA, Salkovskis P (2016) **Pharmacological and psychotherapeutic interventions for management of obsessive-compulsive disorder in adults: a systematic review and network meta-analysis** *Lancet Psychiatry* **3**:730–9

30. Prado CE, Watt S, Crowe SF (2018) **A meta-analysis of the effects of antidepressants on cognitive functioning in depressed and non-depressed samples** *Neuropsychol Rev* **28**:32–72
31. Rosenblat JD, Kakar R, McIntyre RS (2015) **The cognitive effects of antidepressants in major depressive disorder: a systematic review and meta-analysis of randomized clinical trials** *Int J Neuropsychopharmacol* **19**
32. Salagre E, Solé B, Tomioka Y, Fernandes BS, Hidalgo-Mazzei D, Garriga M (2017) **Treatment of neurocognitive symptoms in unipolar depression: A systematic review and future perspectives** *J Affect Disord* **221**:205–21
33. Vernon JA, Grudnikoff E, Seidman AJ, Frazier TW, Vemulapalli MS, Pareek P (2014) **Antidepressants for cognitive impairment in schizophrenia--a systematic review and meta-analysis** *Schizophr Res* **159**:385–94
34. Mundt JC, Marks IM, Shear MK, Greist JH (2002) **The Work and Social Adjustment Scale: a simple measure of impairment in functioning** *Br J Psychiatry J Ment Sci* **180**:461–4
35. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016) **Characterizing a psychiatric symptom dimension related to deficits in goal-directed control** *eLife* **5**
36. Althouse AD (2016) **Adjust for multiple comparisons? It's not that simple** *Ann Thorac Surg* **101**:1644–5
1. Zung WW (1965) **A self-rating depression scale** *Arch Gen Psychiatry* **12**:63–70
2. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G (1983) **Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2)**
3. Mason O, Linney Y, Claridge G (2005) **Short scales for measuring schizotypy** *Schizophr Res* **78**:293–6
4. Patton JH, Stanford MS, Barratt ES (1995) **Factor structure of the Barratt impulsiveness scale** *J Clin Psychol* **51**:768–74
5. Foa EB, Huppert JD, Leiberg S, Langner R, Kichic R, Hajcak G (2002) **The Obsessive-Compulsive Inventory: development and validation of a short version** *Psychol Assess* **14**:485–96
6. Liebowitz MR (1987) **Social phobia** *Mod Probl Pharmacopsychiatry* **22**:141–73
7. Garner DM, Olmsted MP, Bohr Y, Garfinkel PE (1982) **The Eating Attitudes Test: psychometric features and clinical correlates** *Psychol Med* **12**:871–8
8. Marin RS, Biedrzycki RC, Firinciogullari S (1991) **Reliability and validity of the Apathy Evaluation Scale** *Psychiatry Res* **38**:143–62
9. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M (1993) **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
10. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016) **Characterizing a psychiatric symptom dimension related to deficits in goal-directed control** *eLife* **5**

11. Mundt JC, Marks IM, Shear MK, Greist JH (2002) **The Work and Social Adjustment Scale: a simple measure of impairment in functioning** *Br J Psychiatry J Ment Sci* **180**:461–4
12. Fleming SM (2017) **HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings** *Neurosci Conscious* **2017**
13. Guggenmos M (2021) **Measuring metacognitive performance: type 1 performance dependence and test-retest reliability** *Neurosci Conscious* **2021**

Author information

Celine A Fox

Department of Psychology, Trinity College Dublin, Dublin, Ireland, Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

For correspondence: foxce@tcd.ie

Chi Tak Lee

Department of Psychology, Trinity College Dublin, Dublin, Ireland, Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

Anna K Hanlon

Department of Psychology, Trinity College Dublin, Dublin, Ireland, Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland

Tricia XF Seow

Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom

Kevin Lynch

Department of Psychology, Trinity College Dublin, Dublin, Ireland

Siobhán Harty

SilverCloud Science, SilverCloud Health Ltd, Dublin, Ireland

Derek Richards

Department of Psychology, Trinity College Dublin, Dublin, Ireland, SilverCloud Science, SilverCloud Health Ltd, Dublin, Ireland

Jorge Palacios

Department of Psychology, Trinity College Dublin, Dublin, Ireland, SilverCloud Science, SilverCloud Health Ltd, Dublin, Ireland

Veronica O’Keane

Department of Psychiatry, Trinity College Dublin, Dublin, Ireland, Tallaght Hospital, Trinity Centre for Health Sciences, Tallaght University Hospital, Tallaght, Dublin, Ireland

Klaas E Stephan

Translational Neuromodeling Unit (TNU), Institute for Biomedical Engineering, University of Zurich & ETH Zurich, Zurich, Switzerland, Max Planck Institute for Metabolism Research, Cologne, Germany

Claire M Gillan

Department of Psychology, Trinity College Dublin, Dublin, Ireland, Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland, Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland

Editors

Reviewing Editor

Alex Pike

University of York, United Kingdom

Senior Editor

Jonathan Roiser

University College London, United Kingdom

Reviewer #1 (Public Review):

It has been shown previously that there are relationships between a transdiagnostic construct of anxious-depression (AD), and average confidence rating in a perceptual decision task. This study sought to investigate these results, which have been replicated several times but only in cross-sectional studies. This work applies a perceptual decision-making task with confidence ratings and a transdiagnostic psychometric questionnaire battery to participants before and after an iCBT course. The iCBT course reduced AD scores in participants, and their mean confidence ratings increased without a change in performance. Participants with larger AD changes had larger confidence changes. These results were also shown in a separate smaller group receiving antidepressant medication. A similar sized control group with no intervention did not show changes.

The major strength of the study is the elegant and well-powered data set. Longitudinal data on this scale is very difficult to collect, especially with patient cohorts, so this approach represents an exciting breakthrough. Analysis is straightforward and clearly presented. However, no multiple comparison correction is applied despite many different tests. While in general I am not convinced of the argument in the citation provided to justify this, I think in this case the key results are not borderline ($p < 0.001$) and many of the key effects are replications, so there are not so many novel/exploratory hypothesis and in my opinion the results are convincing and robust as they are. The supplemental material is a comprehensive description of the data set, which is a useful resource.

The authors achieved their aims, and the results clearly support the conclusion that the AD and mean confidence in a perceptual task covary longitudinally.

I think this study provides an important impact to the project of computational psychiatry. Specifically, it shows that the relationship between transdiagnostic symptom dimensions and behaviour is meaningful within as well as across individuals.

Reviewer #2 (Public Review):

The authors of this study investigated the relationship between (under)confidence and the anxious-depressive symptom dimension in a longitudinal intervention design. The aim was to determine whether confidence bias improves in a state-like manner when symptoms improve. The primary focus was on patients receiving internet-based CBT (iCBT; $n = 649$),

while secondary aims compared these changes to patients receiving antidepressants ($n=82$) and a control group ($n=88$).

The results support the authors' conclusions, and the authors convincingly demonstrated a weak link between changes in confidence bias and anxious-depressive symptoms (not specific to the intervention arm)

The major strength and contribution of this study is the use of a longitudinal intervention design, allowing the investigation of how the well-established link between underconfidence and anxious-depressive symptoms changes after treatment. Furthermore, the large sample size of the iCBT group is commendable. The authors employed well-established measures of metacognition and clinical symptoms, used appropriate analyses, and thoroughly examined the specificity of the observed effects.

However, due to the small effect sizes, the antidepressant and control groups were underpowered, reducing comparability between interventions and the generalizability of the results. The lack of interaction effect with treatment makes it harder to interpret the observed differences in confidence, and practice effects could conceivably account for part of the difference. Finally, it was not completely clear to me why, in the exploratory analyses, the authors looked at the interaction of time and symptom change (and group), since time is already included in the symptom change index.

This longitudinal study informs the field of metacognition in mental health about the changeability of biases in confidence. It advances our understanding of the link between anxiety-depression and underconfidence consistently found in cross-sectional studies. The small effects, however, call the clinical relevance of the findings into question. I would have found it useful to read more in the discussion about the implications of the findings (e.g., why is it important to know that the confidence bias is state-dependent; given the effect size of the association between changes in confidence and symptoms, is the state-trait dichotomy the right framework for interpreting these results; suggestions for follow-up studies to better understand the association).

Reviewer #3 (Public Review):

This study reports data collected across time and treatment modalities (internet CBT (iCBT), pharmacological intervention, and control), with a particularly large sample in the iCBT group. This study addresses the question of whether metacognitive confidence is related to mental health symptoms in a trait-like manner, or whether it shows state-dependency. The authors report an increase in metacognitive confidence as anxious-depression symptoms improve with iCBT (and the extent to which confidence increases is related to the magnitude of symptom improvement), a finding that is largely mirrored in those who receive antidepressants (without the correlation between symptom change and confidence change). I think these findings are exciting because they directly relate to one of the big assumptions when relating cognition to mental health - are we measuring something that changes with treatment (is malleable), so might be mechanistically relevant, or even useful as a biomarker?

This work is also useful in that it replicates a finding of heightened confidence in those with compulsivity, and lowered confidence in those with elevated anxious-depression.

One caveat to the interest of this work is that it doesn't allow any causal conclusions to be drawn, and only measures two timepoints, so it's hard to tell if changes in confidence might drive treatment effects (but this would be another study). The authors do mention this in the limitations section of the paper.

Another caveat is the small sample in the antidepressant group.

Some thoughts I had whilst reading this paper: to what extent should we be confident that the changes are not purely due to practice? I appreciate there is a relationship between improvement in symptoms and confidence in the iCBT group, but this doesn't completely rule out a practice effect (for instance, you can imagine a scenario in which those whose symptoms have improved are more likely to benefit from previously having practiced the task).

Relatedly, to what extent is there a role for general task engagement in these findings? The paper might be strengthened by some kind of control analysis, perhaps using (as a proxy for engagement) the data collected about those who missed catch questions in the questionnaires.

I was also unclear what the findings about task difficulty might mean. Are confidence changes purely secondary to improvements in task performance generally - so confidence might not actually be 'interesting' as a construct in itself? The authors could have commented more on this issue in the discussion.

To make code more reproducible, the authors could have produced an R notebook that could be opened in the browser without someone downloading the data, so they could get a sense of the analyses without fully reproducing them.

Rather than reporting full study details in another publication I would have found it useful if all relevant information was included in a supplement (though it seems much of it is). This avoids situations where the other publication is inaccessible (due to different access regimes) and minimises barriers for people to fully understand the reported data.

Author Response:

Reviewer #1 (Public Review):

[...] The major strength of the study is the elegant and well-powered data set. Longitudinal data on this scale is very difficult to collect, especially with patient cohorts, so this approach represents an exciting breakthrough. Analysis is straightforward and clearly presented. However, no multiple comparison correction is applied despite many different tests. While in general I am not convinced of the argument in the citation provided to justify this, I think in this case the key results are not borderline ($p < 0.001$) and many of the key effects are replications, so there are not so many novel/exploratory hypothesis and in my opinion the results are convincing and robust as they are. The supplemental material is a comprehensive description of the data set, which is a useful resource.

The authors achieved their aims, and the results clearly support the conclusion that the AD and mean confidence in a perceptual task covary longitudinally. I think this study provides an important impact to the project of computational psychiatry. Specifically, it shows that the relationship between transdiagnostic symptom dimensions and behaviour is meaningful within as well as across individuals.

Response: We thank the reviewer for their appraisal of our paper and positive feedback on the main manuscript and supplementary information. We agree with the reviewer that the lack of multiple comparison corrections can also be justified by key findings being replications and not borderline significance. We have added this additional justification to the manuscript (Methods, Statistical Analyses, page 15, line 568: "Adjustments for multiple comparisons were not conducted for analyses of replicated effects")

Reviewer #2 (Public Review):

[...] The major strength and contribution of this study is the use of a longitudinal intervention design, allowing the investigation of how the well-established link between underconfidence and anxious-depressive symptoms changes after treatment. Furthermore, the large sample size of the iCBT group is commendable. The authors employed well-established measures of metacognition and clinical symptoms, used appropriate analyses, and thoroughly examined the specificity of the observed effects.

However, due to the small effect sizes, the antidepressant and control groups were underpowered, reducing comparability between interventions and the generalizability of the results. The lack of interaction effect with treatment makes it harder to interpret the observed differences in confidence, and practice effects could conceivably account for part of the difference. Finally, it was not completely clear to me why, in the exploratory analyses, the authors looked at the interaction of time and symptom change (and group), since time is already included in the symptom change index.

Response: We thank the reviewer for their succinct summary of the main results and strengths of our study. We apologise for the confusion in how we described that analysis. We examine state-dependence, i.e. the relationship between symptom change and metacognition change, in two ways in the paper – perhaps somewhat redundantly. (1) By correlating change indices for both measures (e.g. as plotted in Figure 3D) and (2) by doing a very similar regression-based repeated-measures analysis, i.e. mean confidence ~ time*anxious-depression score change. Where mean confidence is entered with two datapoints – one for pre- and one for post-treatment (i.e. within-person) and anxious-depression change is a single value per person (between-person change score). This allowed us to test if those with the biggest change in depression had a larger effect of time on confidence. This has been added to the paper for clarification (Methods, Statistical Analysis, page 14, line 553-559: “To determine the association between change in confidence and change in anxious-depression, we used (1) Pearson correlation analysis to correlate change indices for both measures and, (2) regression-based repeated-measures analysis: mean confidence ~ time*anxious-depression score change, where mean confidence is entered with two datapoints (one for pre- and one for post-treatment i.e., within-person) and anxious-depression change is a single value per person (between-person change score)”).

The analyses have also been reported as regression in the Results for consistency (Treatment Findings: iCBT, page 5, line 197-204: “To test if changes in confidence from baseline to follow-up scaled with changes in anxious-depression, we ran a repeated measure regression analyses with per-person changes in anxious-depression as an additional independent variable. We found this was the case, evidenced by a significant interaction effect of time and change in anxious-depression on confidence ($b=-0.12$, $SE=0.04$, $p=0.002$)... This was similarly evident in a simple correlation between change in confidence and change in anxious-depression ($r(647)=-0.12$, $p=0.002$)”).

This longitudinal study informs the field of metacognition in mental health about the changeability of biases in confidence. It advances our understanding of the link between anxiety-depression and underconfidence consistently found in cross-sectional studies. The small effects, however, call the clinical relevance of the findings into question. I would have found it useful to read more in the discussion about the implications of the findings (e.g., why is it important to know that the confidence bias is state-dependent; given the effect size of the association between changes in confidence and symptoms, is the state-trait dichotomy the right framework for interpreting these results; suggestions for follow-up studies to better understand the association).

Response: Thank you for this comment. We have elaborated on the implications of our findings in the Discussion, including the relevance of the state-trait dichotomy to future research and how more intensive, repeated testing may inform our understanding of the state-like nature of metacognition (Discussion, Limitations and Future Directions, page 10, line 378-380: “More intensive, repeating testing in future studies may also reveal the temporal window at which metacognition has the propensity to change, which could be more momentary in nature.”).

Reviewer #3 (Public Review):

[...] I think these findings are exciting because they directly relate to one of the big assumptions when relating cognition to mental health - are we measuring something that changes with treatment (is malleable), so might be mechanistically relevant, or even useful as a biomarker?

This work is also useful in that it replicates a finding of heightened confidence in those with compulsivity, and lowered confidence in those with elevated anxious-depression.

One caveat to the interest of this work is that it doesn't allow any causal conclusions to be drawn, and only measures two timepoints, so it's hard to tell if changes in confidence might drive treatment effects (but this would be another study). The authors do mention this in the limitations section of the paper.

Another caveat is the small sample in the antidepressant group.

Some thoughts I had whilst reading this paper: to what extent should we be confident that the changes are not purely due to practice? I appreciate there is a relationship between improvement in symptoms and confidence in the iCBT group, but this doesn't completely rule out a practice effect (for instance, you can imagine a scenario in which those whose symptoms have improved are more likely to benefit from previously having practiced the task).

Response: We thank the reviewer for commenting on the implications of our findings and we agree with the caveats listed. We thank the reviewer for raising this point about practice effects. A key thing to note is that this task does not have a learning element with respect to the core perceptual judgement (i.e., accuracy), which is the target of the confidence judgment itself. While there is a possibility of increased familiarity with the task instructions and procedures with repeated testing, the task is designed to adjust the difficulty to account of any improvements, so accuracy is stable. We see that we may not have made this clear in some of our language around accuracy vs. perceptual difficulty and have edited the Results to make this distinction clearer (Treatment Findings: iCBT, pages 4-5, lines 184-189: “Although overall accuracy remained stable due to the staircasing procedure, participants’ ability to detect differences between the visual stimuli improved. This was reflected as the overall increase in task difficulty to maintain the accuracy rates from baseline (dot difference: $M=41.82$, $SD=11.61$) to follow-up (dot difference: $M=39.80$, $SD=12.62$), ($b=-2.02$, $SE=0.44$, $p<0.001$, $r^2=0.01$ ”).)

However, it is true that there can be a ‘practice’ effect in the sense that one may feel more confident (despite the same accuracy level) due to familiarity with a task. One reason we do not subscribe to the proposed explanation for the link between anxious-depression change and confidence change is that the other major aspect of behaviour that improved with practice did so in a manner unrelated to clinical change. As noted above in the quoted text, participants’ discrimination improved from baseline to follow-up, reflected in the need for higher difficulty level to maintain accuracy around 70%. Crucially, this was not associated with symptom change. This speaks against a general mechanism where symptom

improvement leads to increased practice effects in general. Only changes in confidence specifically are associated with improved symptoms. We have provided more detail on this in the Discussion (page 9, lines 324-326: “This association with clinical improvements was specific to metacognitive changes, and not changes in task performance, suggesting that changes in confidence do not merely reflect greater task familiarity at follow-up.”).

Relatedly, to what extent is there a role for general task engagement in these findings? The paper might be strengthened by some kind of control analysis, perhaps using (as a proxy for engagement) the data collected about those who missed catch questions in the questionnaires.

Response: Thank you for your comment. We included the details of data quality checks in the Supplement. Given the small number of participants that failed more than one attention checks (1% of the iCBT arm) and that all those participants passed the task exclusion criteria, we made the decision to retain these individuals for analyses. We have since examined if excluding these small number of individuals impacts our findings. Excluding those that failed more than one catch item did not affect the significance of results, which has now been added to the Supplementary Information (Data Quality Checks: Task and Clinical Scales, page 5, lines 181-185: “Additionally, excluding those that failed more than one catch item in the iCBT arm did not affect the significance of results, including the change in confidence ($b=0.16$, $SE=0.02$, $p<0.001$), change in anxious-depression ($b=-0.32$, $SE=0.03$, $p<0.001$), and the association between change in confidence and change in anxious-depression ($r(638)=-0.10$, $p=0.011$)”).

I was also unclear what the findings about task difficulty might mean. Are confidence changes purely secondary to improvements in task performance generally - so confidence might not actually be 'interesting' as a construct in itself? The authors could have commented more on this issue in the discussion.

Response: Thank you for this comment and sorry it was not clear in the original paper. As we discussed in a prior reply, accuracy – i.e. proportion of correct selections (the target of confidence judgements) are different from the difficulty of the dot discrimination task that each person receives on a given trial. We had provided more details on task difficulty in the Supplement. Accuracy was tightly controlled in this task using a ‘two-down one-up’ staircase procedure, in which equally sized changes in dot difference occurred after each incorrect response and after two consecutive correct responses. The task is more difficult when the dot difference between stimuli is lower, and less difficult when the dot difference between stimuli is greater. Therefore, task difficulty refers to the average dot difference between stimuli across trials. Crucially, task accuracy did not change from baseline to follow-up, only task difficulty. Moreover, changes in task difficulty were not associated with changes in anxious-depression, while changes in confidence were, indicating confidence is the clinically relevance construct for change in symptoms.

We appreciate that this may not have been clear from the description in the main manuscript, and have added more detail on task difficulty to the Methods (Metacognition Task, page 14, lines 540-542: “Task difficulty was measured as the mean dot difference across trials, where more difficult trials had a lower dot difference between stimuli.”) and Results (Treatment Findings: iCBT, pages 4-5, lines 184-186: “Although overall accuracy remained stable due to the staircasing procedure, participants’ ability to detect differences between the visual stimuli improved.”). We have also elaborated more on how improvements in symptoms are associated with change in confidence, not task performance in the Discussion (page 9, lines 324-326: “This association with clinical improvements was specific to metacognitive changes, and not changes in task performance, suggesting that changes in confidence do not merely reflect greater task familiarity at follow-up”).

To make code more reproducible, the authors could have produced an R notebook that could be opened in the browser without someone downloading the data, so they could get a sense of the analyses without fully reproducing them.

Response: Thank you for your comment. We appreciate that an R notebook would be even better than how we currently share the data and code. While we will consider using Notebooks in future, we checked and converting our existing R script library into R Notebooks would require a considerable amount of reconfiguration that we cannot devote the time to right now. We hope that nonetheless the commitment to open science is clear in the extensive code base, commenting and data access we are making available to readers.

Rather than reporting full study details in another publication I would have found it useful if all relevant information was included in a supplement (though it seems much of it is). This avoids situations where the other publication is inaccessible (due to different access regimes) and minimises barriers for people to fully understand the reported data.

Response: We agree this is good practice – the Precision in Psychiatry study is very large, with many irrelevant components with respect to the present study (Lee et al., BMC Psychiatry, 2023). For this reason, we tried to provide all that was necessary and only refer to the Precision in Psychiatry study methods for fine-grained detail. Upon review, the only thing we think we omitted that is relevant is information on ethical approval in the manuscript, which we have now added (Methods, Participants, page 11, lines 412-417: “Further details of the PIP study procedures that are not specific to this study can be found in a prior publication (21). Ethical approval for the PIP study was obtained from the Research Ethics Committee of School of Psychology, Trinity College Dublin and the Northwest-Greater Manchester West Research Ethics Committee of the National Health Service, Health Research Authority and Health and Care Research Wales”). If any further information is lacking, we are happy to include it here also.