eLife

Reviewed Preprint
v2 • January 24, 2025
Revised by authors

Reviewed Preprint
v1 • July 29, 2024

Neuroscience

Transdiagnostic compulsivity is associated with reduced reminder setting, only partially attributable to overconfidence

Annika Boldt ✉, **Celine A Fox, Claire M Gillan, Sam Gilbert**

University College London, London, United Kingdom • Trinity College Dublin, Dublin, Ireland

eLife Assessment

This **important** work addresses the relationship between the transdiagnostic compulsivity dimension and confidence as well as confidence-related behaviours like reminder setting. The relationship between confidence and compulsive disorders has recently received a lot of attention and has been considered to be a key cognitive change. The authors paired an elegant experimental design and pre-registration to give **convincing** evidence of the relationship between compulsivity, reminder setting, and confidence. In the revised version they thoroughly addressed the reviewers comments, in particular adding new analyses clarifying how their findings relate to prediction error based learning further strengthening the manuscript.

https://doi.org/10.7554/eLife.98114.2.sa3

Abstract

In the current study, we explored the behavioural and cognitive correlates of the transdiagnostic trait 'compulsive behaviour and intrusive thought' (CIT). CIT is associated with impaired metacognition, which in turn has been associated with cognitive offloading behaviours such as external reminder-setting that play a key role in fulfilling cognitive goals. In an online study (N=600) we investigated individual differences in compulsivity, metacognition, and external reminder-usage. Compulsive individuals had reduced preference for external reminders. This was partially, but not fully, attributable to their relative overconfidence. In contrast to previous studies, we found no evidence for an impaired confidence-action link: compulsive individuals used their metacognition to guide offloading just as much as their noncompulsive counterparts. Given the compensatory nature of cognitive offloading, our findings imply that compulsive individuals are at increased risk of inadequate external memory support. Along with transdiagnostic variation in the general population, this finding could also have implications for clinical conditions, such as OCD.

Introduction

In recent studies of clinically relevant individual differences, there has been a paradigm shift towards of the study of transdiagnostic traits, challenging the traditional, diagnostic approach. Using factor analysis, temporally stable (see Fox et al., 2023a; Sookud, Martin, Gillan, & Wise, 2024), transdiagnostic phenotypes can be extracted from extensive symptom datasets (Wise, Robinson, & Gillan, 2023). These traits are not confined to a single clinical diagnosis but instead can span a range of conditions, at the same time addressing the diagnostic heterogeneity within conditions, such as obsessive-compulsive disorder (OCD; e.g., Gillan et al., 2016; Wise & Dolan, 2020). There are obvious practical benefits of these methodologies, such as their potential to reduce the clinical burden by making the treatment of comorbid conditions more efficient and effective (Harvey, Watkins, Mansell, & Shafran, 2004). At the same time, they contribute valuable insights into mental health conditions by increasing statistical power and opening new avenues of inquiry (Dalgleish, Black, Johnston, & Bevan, 2020).

In the present study, our focus lies on the latter with the goal to investigate the downstream cognitive and behavioural correlates associated with transdiagnostic compulsivity. This symptom dimension represents a clinical concept characterized by an inability to regulate repetitive behaviours that are harmful to oneself, commonly observed in a variety of conditions, particularly obsessive-compulsive disorder (OCD), schizophrenia, addiction, and eating disorders. Previous research links transdiagnostic compulsivity to impairments in metacognition, defined as thinking about one's own thoughts, encompassing a broad spectrum of self-reflective signals, such as feelings of confidence (e.g., Rouault, Seow, Gillan & Fleming, 2018; Seow & Gillan, 2020; Benwell, Mohr, Wallberg, Kouadio, & Ince, 2022; Fox et al., 2023a; Fox et al., 2023b; Hoven, Luigjes, Denys, Rouault, van Holst, 2023a). Other studies have shown that metacognitive signals such as feelings of confidence guide cognitive offloading strategies like setting external reminders as memory aids (e.g., Gilbert, 2015; Boldt & Gilbert, 2019). Here, we aim to bridge these two literatures by investigating compulsivity, metacognition, and cognitive offloading within a single experimental paradigm. While compulsivity and cognitive offloading have both separately been linked to metacognition, the relationship between the two - with metacognition as a potential mediating factor — has not previously been systematically examined. This matters because cognitive offloading plays an integral role in our daily lives and is a key contributor to our effectiveness as cognitive agents (Gilbert et al., 2023).

Metacognition guides reminder setting

Reminders constitute an example of cognitive offloading, defined as the use of physical action to reduce the cognitive demands of a task. By offloading memory demands this way, we not only increase the likelihood of successfully completing tasks (Boldt & Gilbert, 2019), but we may also free up cognitive resources for other activities (Dupont, Zhu & Gilbert, 2022). Choosing between setting a reminder and relying on memory is not a trivial matter. Prior research has emphasized the role of metacognition in determining when individuals resort to cognitive offloading (Gilbert et al., 2023; Gilbert, 2015; Boldt & Gilbert, 2019; Sachdeva & Gilbert, 2020; Risko & Gilbert, 2016): People tend to set more reminders when they feel less confident. In other words, people tend to set reminders when they *think* that they will forget, and this effect holds even after taking into account actual memory ability (e.g., Boldt & Gilbert, 2019). The link between confidence and offloading is observed both for situational fluctuations in confidence due to varying task difficulties (*state* variable; Boldt & Gilbert, 2022), and for a general predisposition towards over- or underconfidence (*trait* variable; Boldt & Gilbert, 2019).

Metacognition, compulsivity, and checking behaviours

Given the known metacognitive impairments associated with compulsivity, changes in reminder-setting behaviour are plausible. More specifically, individuals characterized by transdiagnostic compulsivity have been consistently found to exhibit overconfidence (Rouault, Seow, Gillan & Fleming, 2018; Seow & Gillan, 2020; Benwell, Mohr, Wallberg, Kouadio, & Ince, 2022; Fox et al., 2023a; Fox et al., 2023b; Hoven et al., 2023a). If we consider the link between reminder setting and confidence, this implies a reduced likelihood of utilizing external aids, such as reminders. However, while transdiagnostic compulsivity is liked to overconfidence, the opposite pattern of underconfidence is more common in patients with OCD, a compulsive disorder (as reviewed in Hoven, Lebreton, Engelmann, Denys, Luigjes, & van Holst, 2019). Recent research suggests that metacognitive impairments in transdiagnostic compulsivity and OCD may originate from different mechanisms (Hoven, Rouault, von Holst, & Luigjes, 2023c), advising caution against broad generalizations between these groups. It should also be noted that the composite measure of transdiagnostic compulsivity includes questionnaire items linked not only with OCD but also other clinical conditions such as eating disorders (Tasca et al., 2011; Gillan et al., 2016). This results in an overlap between transdiagnostic compulsivity and other traits such as rigid perfectionism.

Despite opposite trends in metacognitive monitoring performance (under-versus overconfidence), individuals high in transdiagnostic compulsivity and those with a diagnosis of OCD show similar impairments in metacognitive control, characterized by a disrupted connection between confidence and future actions (Seow & Gillan, 2020; Vaghi, Luyckx, Sule, Fineberg, Robbins, & De Martino, 2017). Metacognitive impairments are also central to explanations of compulsive behaviours, notably in OCD patients. In such patients, compulsivity can manifest in the form of checking behaviours, for example, checking that doors are locked or that appliances are switched off (Den Ouden et al., 2022). Whilst checking behaviours are also present in other compulsive disorders (e.g., 'body checking' in eating disorders; Mountford, Haase, & Waller, 2006), in OCD, these checks are often repetitive and ritualised and are typically associated with obsessive thoughts. However, the exact function that checking compulsions serve is unclear; patients commonly report that they have the aim of reducing anxiety generally, preventing a feared consequence from taking place or that they are performed automatically and without thinking (Starcevic et al., 2011). Understanding these motivators has been challenging as studies rely on self-report of often highly individual real-world behaviours.

Some research argues that OCD patients' checking arises from low memory confidence despite intact memory (Tolin, Abramowitz, Brigidi, Amir, Street, & Foa, 2001). Our study has the potential to shed some light on the link between confidence and checking: While checking behaviours can be seen as a way of ensuring that a necessary action *was* performed in the past, reminder-setting is a way of ensuring that a necessary action *will be* performed in the future. In other words, a reminder can serve as a future checkpoint that allows us to revisit a task at an appropriate time to complete it, perhaps by setting an alarm on our phone, jotting down a note, or strategically placing a related object somewhere visible. Given these insights, one might expect an increased reliance on reminders among OCD patients as they strive to establish more checkpoints. By contrast, seeing as transdiagnostic compulsivity is associated with increased confidence, this could be associated with the opposite pattern: a decreased reliance on reminders.

Three possible mechanisms for changes in reminder setting

If, as hypothesised, compulsivity is linked with altered reminder-setting, this could be attributed to at least three underlying mechanisms. First is the *Metacognitive Control Mechanism*: Previous research has found that more compulsive individuals tend to have impaired metacognitive control (Seow & Gillan, 2020), meaning they use metacognitive signals to a lesser extent to guide future behaviour. Compulsivity is a hallmark symptom of OCD and similar deficits in metacognitive

control have been observed in a case-control studies comparing OCD patients with healthy controls examining how confidence and action are correlated (Vaghi et al., 2017; though see also Hoven, Mulder, Denys, van Holst & Luigjes, 2023b; Marzuki et al., 2022). In the context of our study, a Metacognitive Control Mechanism would be reflected in a disrupted relationship between confidence levels and their tendency to set reminders (i.e., the interaction between the bias to be over- or underconfident and transdiagnostic CIT in a regression model predicting a bias to set reminders).

Second, more compulsive individuals might conceivably differ in their reminder-setting strategies due to an altered level of confidence. We call this the *Metacognitive Monitoring Mechanism*, which suggests that the issue arises when forming the confidence signal, rather than in its behavioural application (for clarification on metacognitive monitoring versus control in cognitive offloading, see Boldt & Gilbert, 2022). Prior evidence exists for overconfidence in compulsivity (Rouault et al., 2018; Seow & Gillan, 2020; Benwell et al., 2022; Fox et al., 2023a; Fox et al., 2023b; Hoven et al., 2023a), which would therefore result in fewer reminders.

Lastly, there could be a direct link between compulsivity and reminder-usage, independent of any metacognitive influence. We refer to this as the *Direct Mechanism* and it constitutes any possible influences that affect reminder setting in highly-compulsive CIT participants outside of metacognitive mechanisms, such as perfectionism and the wish to control the task without external aids. Our study aims to differentiate between these three mechanisms. Back when we preregistered our hypotheses only a limited number of studies about confidence and transdiagnostic CIT were available. This resulted in us hypothesising to find support for the *Metacognitive Control Mechanism* and that highly compulsive individuals would offload more due to an increased need for checkpoints. Both of these hypotheses turned out to be incorrect.

Anxious-depressed transdiagnostic phenotype

As well as investigating individual differences in compulsivity, we also measured an anxious-depression factor. Based on previous findings, we predicted opposite influence of these two factors on confidence. Whereas compulsivity has been linked to increased confidence, anxious-depressed individuals typically display relative underconfidence (Rouault, Seow, Gillan & Fleming, 2018; Seow & Gillan, 2020; Benwell, Mohr, Wallberg, Kouadio, & Ince, 2022). By taking a transdiagnostic approach, we were able to jointly investigate the influence of these two factors of confidence which could potentially cancel out if they were investigated separately.

Online reminder-setting task

In the present preregistered study, we asked 600 participants drawn from the general population to complete several individual differences questionnaires. These responses were then weighted to produce both a 'compulsive behaviour and intrusive thought' (CIT) factor and an 'anxious-depression' (AD) factor (Gillan, et al., 2016; Wise & Dolan, 2020). Participants' scores on these factors were then correlated with their behaviour in a reminder-setting task, which was a modified, 20-minute version of the online reminder setting task developed by Gilbert and colleagues (2020; **Figure 6**).

Participants performed a highly-demanding, short-term memory task. On some trials they relied on internal memory alone (which typically resulted in poor accuracy); on other trials they could set external reminders (which dramatically improved accuracy). The key manipulation was the number of points associated with the two strategies. Correct responses always earned 10 points if participants used internal memory, but a lower number of points between 2-9 if they used external reminders. The latter number of points varied from trial to trial, and participants were required each time to decide which strategy they preferred (e.g., 10 points for each correct response with internal memory, or 6 points for each correct response with external reminders). The 'optimal indifference point' was that point value at which an unbiased individual would be

indifferent between the two strategies, based on their objective accuracy in the two conditions. The 'actual indifference point' was the point at which they were actually indifferent, based on all of their decisions. By comparing these two values, we obtained a 'reminder bias': the extent to which an individual had a pro- or anti-reminder bias relative to their individually-calculated optimal strategy. Note that this is different from the absolute rate of reminder usage, because the same absolute rate might reflect inadequate use of reminders in a person with poor memory and excessive reminder usage in a person with good memory ability. Along with the reminder bias, we also calculated a metacognitive bias, which represents participants' over- or underconfidence in memory ability, relative to objective performance. Our study controlled for age, gender, educational attainment, as well as cognitive ability (ICAR5; Kirkegaard & Bjerrekær, 2016), and working memory.

Previewing our results, in line with previous evidence, we found that confidence varied positively with the CIT factor and negatively with the AD factor. However, contrary to our initial expectations, more compulsive individuals offloaded less rather than more, and there was no evidence for disruption in the link between metacognition and offloading. Instead, we discovered an incomplete mediation effect: while a significant proportion of the reduced reminder setting could be attributed to overconfidence, not all the variance was accounted for by this variable. Even after controlling for it, compulsivity still predicted reduced reminder setting. This constitutes a combination of the *Metacognitive Monitoring Mechanisms* and the *Direct Mechanism*.

Results

Here we present the results of a preregistered online study on the relationship between reminder setting, metacognition and transdiagnostic compulsivity. We excluded 69 out of a total of 669 participants based on our six preregistered criteria described in the Methods and Materials section, leaving us with a final sample of 600 participants. All participants completed a previously validated reminder setting task in combination with 49 items from six mental health questionnaires. Three-hundred-and-seventy-five participants identified as male, 218 as female and 7 as other. Participants were on average 32.9 years old (min = 18; max = 76). **Figure 1** shows the included (black) and excluded (red) data, with higher average performance for included participants when reminders were used (96.1%) compared to when people had to do the task unaided (59.2%).

We calculated six key measures for each participant:

1. The first relevant measure is the optimal indifference point (OIP). The OIP describes the reward value (2-9 points) at which an unbiased, reward-maximising participant should be indifferent between the two strategies: using reminders or relying on their own memory. The OIP is calculated from their accuracy with and without reminders. Imagine a participant who achieves 60% accuracy when using their own memory or 100% accuracy when using reminders. In this case the OIP would be 6, because scoring 6 points per item with reminders (100% accuracy) would earn the same number of points as scoring 10 points with internal memory (60% accuracy). For any reward above 6, it would be optimal to choose external reminders; for any reward below 6 it would be optimal to choose internal memory.
2. In contrast, the second relevant measure is the actual indifference point (AIP), which is the number of points at which participants showed indifference between the two strategies. This measure is calculated by fitting a psychometric function to participants' choices at different levels of reward for targets when reminders were used. Please note that all choices were used to calculate the AIP, as participants only found out whether or not they would use a reminder after the decision was made.

3. Together, these variables can be used to calculate the third measure, the *reminder bias*, which is the difference between the OIP and the AIP and therefore reflects participants' tendency to over- or under-use reminders, relative to the optimal strategy. Note that the optimal strategy is calculated individually for each participant and will depend on their own level of performance when using internal memory and external reminders.
4. Fourth, we calculated a *metacognitive bias*, reflecting participants' over- or underconfidence. This is calculated by subtracting objective accuracy (percentage of targets remembered when using internal memory) from the percentage that they predicted that they would be able to remember.
5. Fifth and sixth, based on the questionnaire ratings we calculated how much someone scored on the transdiagnostic CIT and AD factors. Our analyses focus on the relationship between these key measures.

Replication and sanity checks

In the following section, we aim, where the design allows it, to replicate four previous effects for this task. First, with Hypothesis 1, we predicted that the reminder bias and metacognitive bias are negatively correlated replicating previous findings (as reviewed in Gilbert et al., 2023). This effect tests the above-mentioned link between metacognition and cognitive offloading: the less confident someone feels, the more they use reminders. There was indeed a significant negative correlation, $r = -0.2$, $p < .001$ (**Figure 2**). Second, in replication of previous findings (e.g., Gilbert et al., 2020; Sachdeva & Gilbert, 2020; Kirk, Robinson & Gilbert, 2021; Engeler & Gilbert, 2020), Hypothesis 2 expressed our expectation to find an excessive use of reminders reflected in significantly higher optimal indifference points (OIPs) compared to actual indifference points (AIPs). In other words, we expected the reminder bias to be greater than zero, which was indeed the case, $m = 0.52$, $t = 5.1$, $p < .001$, $d = 0.21$. Third, with Hypothesis 3, we expected to replicate that participants would be underconfident in their own memory (e.g., Engeler & Gilbert, 2020), expressed in an average, negative metacognitive bias. Our data supported this hypothesis, $m = -3.64$, $t = -3.1$, $p = .001$, $d = -0.13$. Fourth, Hypothesis 4 predicts that as in previous studies, we would find evidence for compensatory reminder use. Keeping in mind that the OIP reflects the cut-off at which participants should be indifferent between offloading and not offloading and the AIP the cut-off they actually displayed, then looking at these two measures together should show that participants with poorer memory and greater benefit from reminders (lower OIP) tend to use them more (lower AIP). Indeed, the optimal and actual indifference points were positively correlated suggesting participants who benefited most from reminders were more likely to use them, $r = 0.36$, $p < .001$. Taken together, we found that participants showed the usual hallmarks of this offloading task, using their confidence to strategically decide when to offload, general tendencies for setting reminders and for underconfidence, and compensatory reminder use.

Testing our key hypotheses

Elevated confidence in CIT and reduced confidence in AD

We predicted that the metacognitive bias would correlate negatively with AD (Hypothesis 8a; more anxious-depressed individuals tend to be underconfident). For CIT, we preregistered a non-directional, significant link with metacognitive bias (Hypothesis H6a). We found support for both hypotheses, both for AD, $\beta = -0.22$, SE = 0.04, $t = -5.00$, $p < 0.001$, as well as CIT, $\beta = 0.15$, SE = 0.05, $t = 3.30$, $p = 0.001$, controlling for age, gender, and educational attainment (**Figure 3**; see also **Table S1**). Note that for CIT this effect was positive, more compulsive individuals tend to be overconfident.

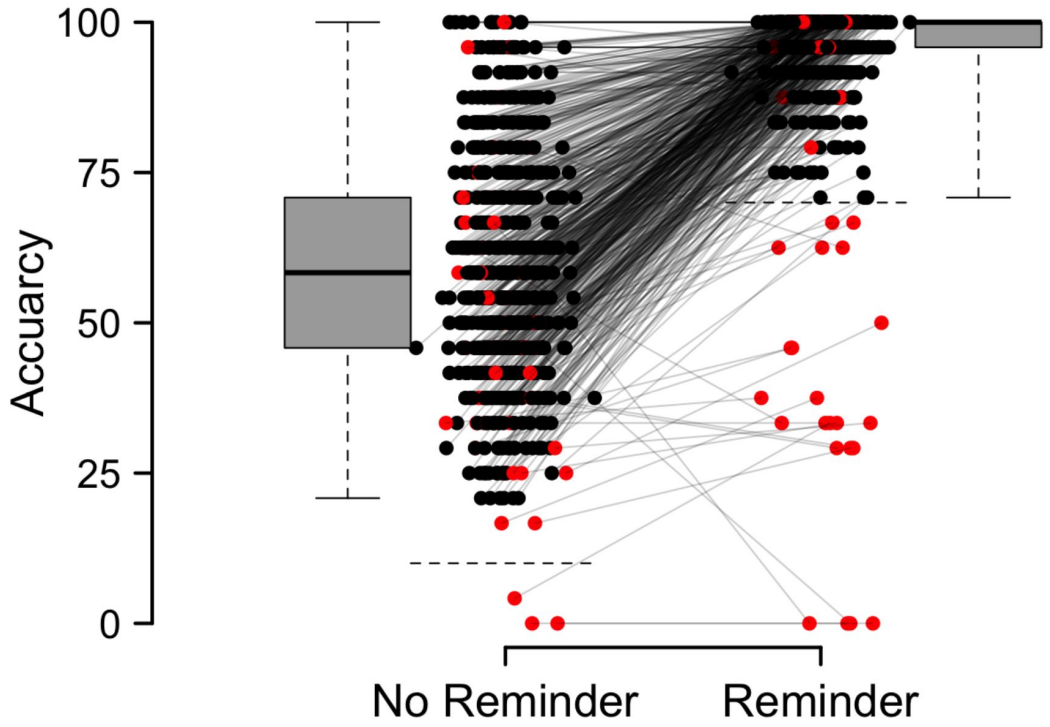


Figure 1

Average accuracy as a function of whether a reminder was used. “No Reminder”: forced internal condition; “Reminder”: forced external condition. Each pair of dots linked by a line indicates one participant. The red data points are excluded participants. The box plots indicate the median surrounded by the interquartile range (25^{th} and 75^{th} percentile). The whiskers show the minimum and maximum. The preregistered exclusion criteria for the accuracies with or without reminder are indicated as horizontal dotted lines (10% and 70% respectively).

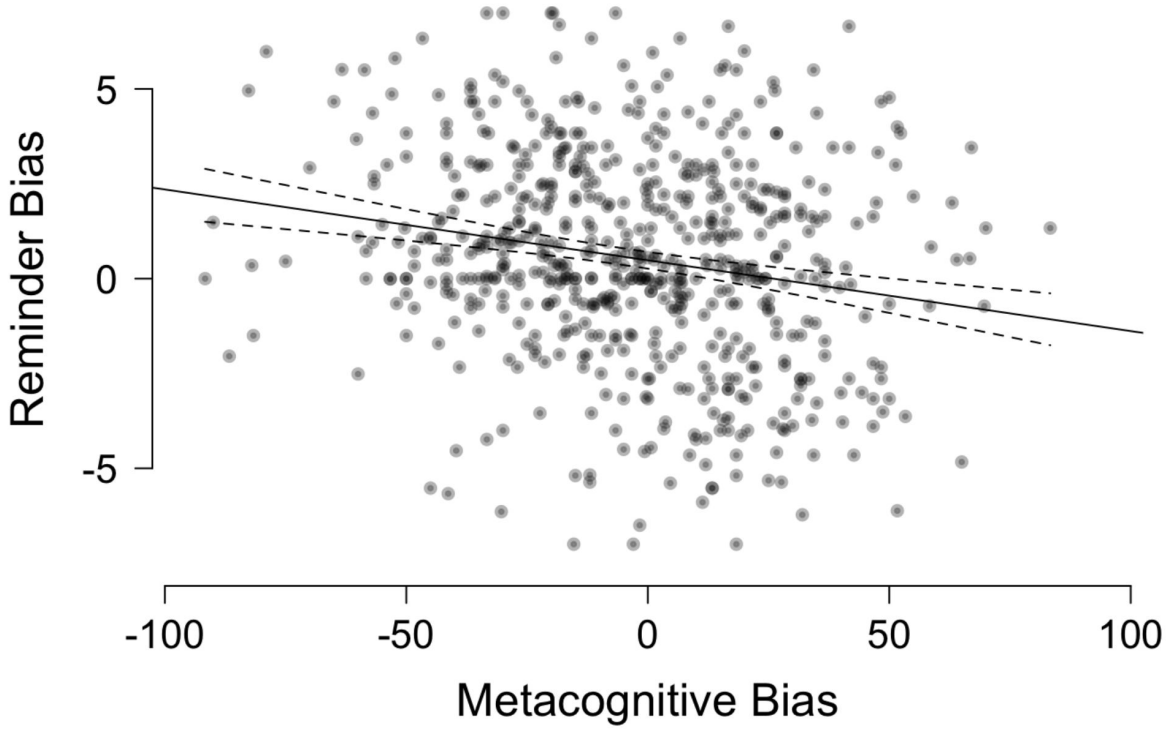


Figure 2

People's tendency to set reminders above or below the optimal offloading strategy (reminder bias) plotted against people's tendency towards over- or underconfidence (metacognitive bias). The solid line indicates the fitted relationship between both variables. The dashed line represents the 95% confidence interval around it. Each circle represents a single participant.

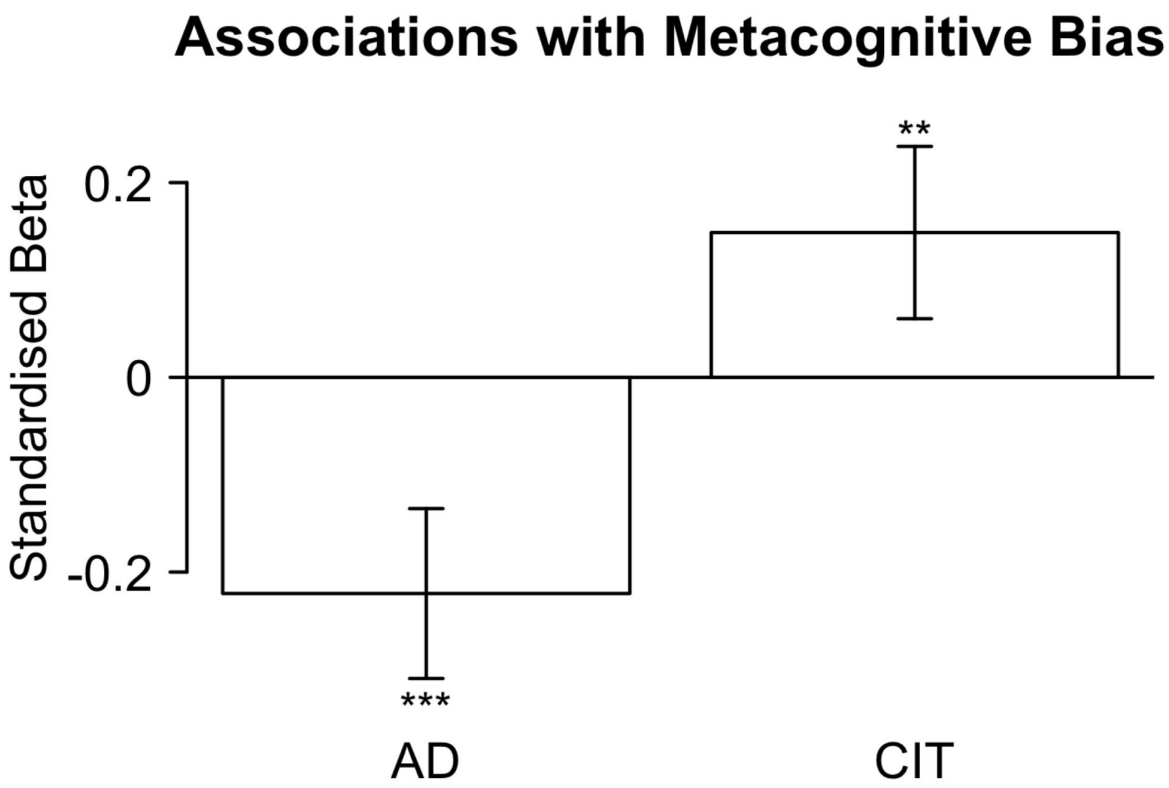


Figure 3

Standardised regression weights for the 'anxious-depression' factor (AD) and the 'compulsive behaviour and intrusive thought' factor (CIT) predicting metacognitive bias. Error bars indicate 95% confidence intervals. Asterisks indicate significance: "***": < 0.001; "**": < 0.01; "*": < 0.05.

We furthermore preregistered to also test this for raw confidence (percentage of circles participants predicted they will remember, rather than the accuracy-corrected metacognitive bias score; Hypotheses H8b and H6b). Indeed, the same patterns were found for both AD, β = −0.29, SE = 0.04, t = −6.44, p < 0.001, and CIT, β = 0.16, SE = 0.05, t = 3.50, p < 0.001 (see **Table S2**). Including scores from the cognitive ability test as an additional covariate (Hypotheses H8c and H6c respectively) furthermore did not change the results, AD, β = −0.20, SE = 0.04, t = −4.47, p < 0.001; CIT, β = 0.12, SE = 0.04, t = 2.74, p = 0.006 (see **Table S3**). Taken together, these results suggest that concordant with our hypotheses, compulsivity was linked to inflated confidence and anxiety to deflated confidence.

Contrary to Expectations, Compulsivity Reduced Pro-offloading Bias

We expected to find a positive link between CIT factor scores and reminder bias. In other words, we predicted that more compulsive individuals would show a greater pro-offloading bias, relative to the optimal strategy (Hypothesis H5a). However, our results showed the exact opposite effect with a significantly reduced reminder bias in compulsive individuals, β = −0.16, SE = 0.04, t = −3.60, p < 0.001, controlling for age, gender, and educational attainment (**Figure 4**; see also **Table S4**). This trend persisted when, instead, we predicted the absolute number of reminders chosen by the participant (Hypothesis H5b), β = −0.14, SE = 0.05, t = −3.08, p = 0.002 (see **Table S5**), as well as when predicting the AIP (Hypothesis H5c), β = 0.15, SE = 0.05, t = 3.32, p < 0.001 (see **Table S6**).

Previous studies have found reduced working memory in OCD (Harkin & Kessler, 2011), which could potentially lead to increased reminder use in compulsivity. However, the reduced reminder bias persisted if d' from the 2-back task was included as an additional covariate (Hypothesis H5d), β = −0.15, SE = 0.05, t = −3.27, p = 0.001 (see **Table S7**). Finally, we predicted that our results would persist independent of whether or not the scores from the cognitive ability test were included as an additional covariate (Hypothesis H5e), which was indeed the case, β = −0.16, SE = 0.05, t = −3.55, p < 0.001 (see **Table S8**). It should be noted that all our regression models included both CIT and AD as predictors to separate out the potentially competing influences of these predictors, as well as age, gender, and educational attainment as demographic covariates.

We furthermore preregistered to conduct the same tests for the AD factor but without any directional hypotheses. AD was not significantly linked to any changes in reminder bias, β = 0.06, SE = 0.04, t = 1.43, p = 0.15 (see **Table S4**), absolute number of reminders, β = 0.06, SE = 0.05, t = 1.32, p = 0.19 (see **Table S5**), or AIP, β = −0.08, SE = 0.04, t = −1.74, p = 0.08, controlling for age, gender, and educational attainment (see **Table S6**). This null effect did not change when working memory, β = 0.05, SE = 0.04, t = 1.21, p = 0.23 (see **Table S7**), or scores from the cognitive ability test were included as additional covariates, β = 0.06, SE = 0.04, t = 1.40, p = 0.16 (see **Table S8**).

Taken together, these results suggest that compulsive individuals are less biased towards offloading, in contrast to our hypothesised direction of the effect, but consistent with the observation of increased confidence in their ability on this task.

No Evidence for Impaired Confidence-Offloading Link

We predicted to find support for the *Metacognitive Control Mechanism*, meaning that CIT would act as a moderator on the link between confidence and offloading (Hypothesis H7a). In other words, we expected to find that the correlation between the metacognitive and the reminder bias to be weakened in highly compulsive individuals. However, the interaction between metacognitive bias and compulsivity in a model predicting the reminder bias was not significant, β = −0.01, SE = 0.04, t = −0.18, p = 0.86, controlling for age, gender, and educational attainment (see **Table S9**). This means that in our task, confidence and offloading were linked just as much as in their low compulsive counterparts. These results remained the same even if working memory performance

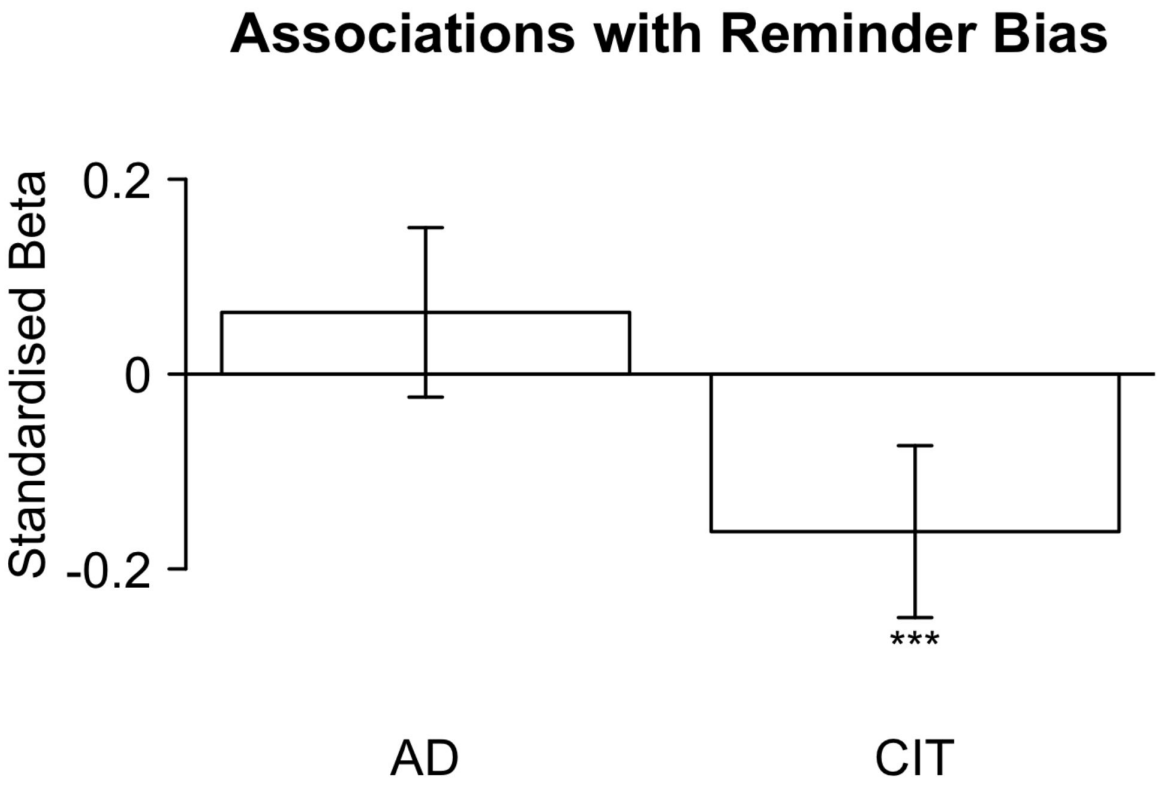


Figure 4

Standardised regression weights for the 'anxious-depression' factor (AD) and the 'compulsive behaviour and intrusive thought' factor (CIT) predicting reminder bias. Error bars indicate 95% confidence intervals. Asterisks indicate significance: "***": < 0.001; "**": < 0.01; "*": < 0.05.

(d' from the 2-back task) was included as an additional covariate (Hypothesis H7b), β = −0.01, SE = 0.04, t = −0.26, p = 0.80 (see **Table S10**), or if scores from the cognitive ability test were included as an additional covariate (Hypothesis H7c), β = −0.01, SE = 0.04, t = −0.18, p = 0.86 (see **Table S10**).

Contrary to our initial hypotheses, we found that increased CIT was associated with decreased rather than increased bias towards offloading. Seeing as CIT was also associated with increased confidence, and high confidence predicts low bias towards offloading, we tested whether the relationship between CIT and offloading was mediated via confidence: whilst parts of the reduction of reminders could be traced back to overconfidence, β = −0.18, SE = 0.04, t = −4.51, p < 0.001, there was still a significant proportion of variance that was linked to compulsivity independently of this, β = −0.14, SE = 0.05, t = −3.03, p = 0.003. **Figure 5** summarises this incomplete mediation effect. In previous sections we have already reported the total effect of CIT on the reminder bias, β = −0.16, SE = 0.05, t = −3.60, p < 0.001. Equally, we have already reported the effect of CIT on the mediator (the metacognitive bias), β = 0.12, SE = 0.05, t = 2.76, p = 0.006.

We validated this outcome through an exploratory causal mediation analysis. The indirect influence of the CIT factor on reminder bias going through the metacognitive bias was calculated to be (0.12) * (-0.18) = −0.02. To determine the significance of this influence, we implemented bootstrapping procedures. We computed unstandardized indirect effects for each of the 1,000 bootstrapped samples, followed by the calculation of the 95% confidence interval, identifying the indirect effect at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect (the average causal mediation effect) computed to be -0.02, with the 95% confidence interval ranging from -0.05 to -0.01. This indicated that the effect was statistically significant at p < 0.001.

Finally, we preregistered to run the same analysis for the AD factor without hypothesising about any specific direction for any potential effects. However, we did not find evidence for a moderation effect (an interaction between AD scores and metacognitive bias when predicting the reminder bias), β = −0.04, SE = 0.04, t = −1.13, p = 0.26, controlling for age, gender, and educational attainment (see **Table S12**).

In summary, whilst we found no support for the *Metacognitive Control Mechanism* (as would be reflected in a disrupted link between confidence and offloading), we did find support for both the *Metacognitive Monitoring Mechanism* (reduced pro-reminder bias as a downstream consequence of overconfidence) and the *Direct Mechanism* (independent contribution of CIT on offloading). The Supplement furthermore lists several additional analyses, both planned and exploratory.

Discussion

In the current study, we explored the behavioural and cognitive correlates of two transdiagnostic traits: 'compulsive behaviour and intrusive thought' (CIT) and 'anxious-depression' (AD). We focused on changes in cognitive offloading and metacognition related to transdiagnostic compulsivity. Our results replicated that more compulsive individuals were relatively overconfident, while those who were more anxious-depressed were relatively underconfident. Contrary to expectations, we observed a decreased bias towards reminders among more compulsive participants. This reduction in bias was only partially accounted for by their relative overconfidence. This partial mediation can be interpreted through both a *Metacognitive Monitoring Mechanism* (differences in the formation of the confidence signal rather than its behavioural application) and a *Direct Mechanism* (no metacognitive involvement). We found no support for a *Metacognitive Control Mechanism*, which would centre on how confidence is used to adapt behaviour (Nelson & Narens, 1990; Boldt & Gilbert, 2022).

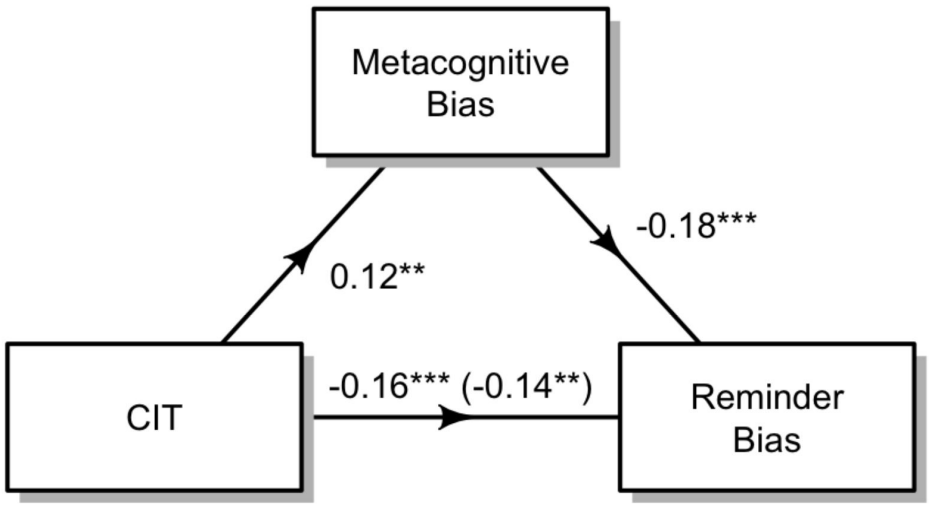


Figure 5

Diagram of the mediation analysis testing for the influence of the 'compulsive behaviour and intrusive thought' factor (CIT) on reminder bias, both directly and indirectly through the metacognitive bias. Standardised regression coefficients are given for each path. The value in parentheses indicates the influence of CIT on reminder bias controlling for the influence of the metacognitive bias. Asterisks indicate significance: "***": < 0.001; "**": < 0.01; "*": < 0.05.

Perfectionism and the need to control as potential explanations

Contrary to our hypothesis, our study revealed an inverse relationship between transdiagnostic compulsivity and offloading: the reminder bias was reduced in more compulsive individuals. One possible interpretation is perfectionism: Some compulsive individuals may avoid using reminders altogether due to rigid, perfectionistic beliefs about needing to remember everything without relying on external aids, and using reminders could trigger their anxiety or feed into their obsessions about being forgetful or unreliable. This interpretation aligns with findings suggesting that perfectionism serves as a transdiagnostic maintaining and risk factor for various mental health conditions, including compulsive disorders like eating disorders and OCD (Egan, Wade, & Shafran, 2011).

No effect of anxiety on offloading

Interestingly, we found no significant influence of the anxious-depressed (AD) transdiagnostic phenotype on offloading. This aligns with a recent study by Kirk, Robinson, & Gilbert (2021), which also found no effect of anxiety on offloading. However, their study, which used the 'trait' component of the STAI to measure anxiety (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983), found no relative underconfidence among anxious participants either. Our transdiagnostic approach likely revealed this confidence effect by separating the counteracting influences of AD and CIT factors. This distinction underscores the value of a transdiagnostic approach.

Our findings align with those reported in a recent study by Mohr, Ince, and Benwell (2024). The authors observed that while high-AD participants were underconfident in a perceptual task, this underconfidence did not lead to increased information-seeking behaviour. Future research should explore whether this is due to their pessimism regarding the effectiveness of confidence-modulated strategies (i.e., setting reminders or seeking information) or whether it stems from apathy. Another possibility is that the relevant downstream effects of anxiety were not measured in our study and instead may lie in reminder-checking behaviours.

No evidence for an impaired confidence-action link in compulsivity

Contrary to Seow and Gillan (2020) and Vaghi et al. (2017), our study did not find the impaired confidence-action link (*Metacognitive Control Mechanism*) reported for transdiagnostic compulsivity and OCD patients on this task. This may be because of differences between tasks — prior work used a reinforcement learning task with a clear learning element from trial to trial. Alternatively, it is possible our study was underpowered, as our sample size was designed to detect overconfidence in compulsivity, not the more nuanced but still psychometrically robust confidence-action link (Loosen, Seow & Hauser, 2023), which would have required a far larger sample size. Recent studies also failed to find decreased action-confidence coupling with relatively small groups of OCD patients and controls (Hoven et al., 2023b; Marzuki et al., 2022). Indeed, both our paradigm and the earlier predictive-inference task tested for an interaction effect, which is more challenging to power adequately. Future research should consider using more direct measures that ideally aim to manipulate confidence directly.

Implications

Participants in our current study were recruited from the general population through Prolific, meaning that the variance likely represents primarily subclinical sources. Consequently, caution should be exercised when extrapolating these results to clinical populations. For example, a recent study indicated that metacognitive impairments in OCD originate from different mechanisms than those observed in transdiagnostic compulsivity (Hoven et al., 2023c). Given its metacognitive impairments and the prevalent symptom of checking, OCD still remains a particularly relevant patient group for studying reminder setting and future studies need to explore this area further.

Due to their underconfidence, OCD patients might engage in more frequent reminder setting. This behaviour could serve as a compensatory mechanism, especially since OCD patients often face challenges with working memory (Harkin & Kessler, 2011) and prospective memory (Harris, Vaccaro, Jones, & Boots, 2010; Racsmány, Demeter, Csigó, Harsányi, & Németh, 2011). However, it could also worsen their checking symptoms as more reminders means more opportunities to check.

On the other hand, it is possible that the observed underconfidence in OCD populations may actually reflect the impact of an uncontrolled anxiety factor, effectively neutralizing the influence of compulsivity on confidence. This confounding issue could explain the inconsistent findings regarding confidence bias in both compulsivity and OCD. If this was the case, then future research should investigate which influences on confidence — the reductions caused by the AD factor or the increases caused by the CIT factor — are the driving force behind any changes in reminder setting in OCD.

A pivotal question remains: will the overall reduction in reminder setting, referred to as a "Direct Effect" in this study, also be observed in OCD patients and other compulsive disorders? Such findings could support the hypothesis that an inherent aspect of compulsivity leads to the decreased use of external aids, potentially due to perfectionism or a need for control.

Limitations

Our results are based on a well-validated paradigm which our lab has previously used on other, published studies (as reviewed in Gilbert et al., 2023). However, reliance on a single behavioural task also means that our results might not generalise onto cognitive offloading more broadly or even reminder setting in other contexts. As a first step, future work should aim to replicate our findings in the context of other experimental designs.

Another limitation is that in the present study, we focused solely on measuring two transdiagnostic factors: CIT and AD. We omitted the third factor, 'social withdrawal' (SW). By doing so, we were able to reduce the number of items from six clinical questionnaires to 49 (Wise & Dolan, 2020), thereby shortening the required time for completion—an essential consideration for online research (Sauter, Draschkow & Mack, 2020). Nevertheless, this focused approach could introduce variability in capturing these transdiagnostic phenotypes. A recent preprint from Hopkins and colleagues (2022) supports this approach. They used machine learning to select 71 items capable of reliably measuring all three factors, suggesting that future transdiagnostic studies might similarly adopt more concise item sets.

Conclusion

With the present study, we investigated the downstream cognitive and behavioral effects of two transdiagnostic traits, CIT and AD. In particular, we were interested in the effect these factors have on metacognition and cognitive offloading, operationalised as prospective confidence and reminder setting respectively. We replicated the finding that more compulsive individuals tend to be relatively overconfident, whereas anxious-depressed individuals tend to be relatively underconfident. Contrary to our hypotheses, however, we found that compulsivity was linked reduced offloading and that this effect was only in part explained by overconfidence.

Fulfilling delayed intentions (i.e., prospective memory) is a vital process for daily living and behavioral independence. However, this process is also highly fallible (e.g., Crawford, Smith, Maylor, Della Sala & Logie, 2003). External memory aids are highly effective and commonplace tools that compensate for these memory failures (e.g., Jones, Benge & Scullin, 2021; Scullin et al., 2022). Our findings suggest that compulsive individuals are at particular risk of inadequate external memory support and would potentially benefit from interventions that target cognitive offloading strategies.

Methods

Task and Procedure

For the present, preregistered study, we used a novel variant of an online cognitive-offloading task ("optimal reminders task"; cf. Gilbert et al., 2020). This task allowed us to measure how people set reminders in relation to their confidence. All procedures, hypotheses, and planned analyses were preregistered at *https://osf.io/kztf8* prior to the commencement of data collection.

On every trial, participants were instructed to move several numbered, yellow circles to the bottom of a square in consecutive order (see **Figure 6A**). Whenever a circle was removed, a new one appeared up to a total of 15 circles. The source of difficulty of this task stems from the 'special' circles, which constitute the delayed intentions people have to fulfil. These circles flashed in a colour (blue, orange, or magenta) when they first appear on screen before fading to yellow. Participants' task was to drag these circles to their colour-corresponding side once the time has come to remove the respective special circle (top, left or right). There were 6 special circles per trial. On some trials, participants had to rely on their own memory to complete the task and remember the target locations of the special circles. On other trials, they set spatial reminders indicating the locations to which the special circles must be moved to. More specifically, they were taught to move the special circle next to the border through which it would have to be moved out of the square later.

Every trial began with a decision: participants could choose to do the task without reminders and earn 10 points for every special circle they remembered to move to the correct border, or they could choose to use reminders but earn less for each special circle (**Figure 6B**). Critically, this lesser amount was varied between 2 and 9 points, allowing us to calculate participants' indifference point when trading off the benefit of reminders with their reduced reward. This *actual* indifference point (AIP) could then be contrasted against their *optimal* indifference point (OIP), calculated from participants' accuracy with or without reminders, see below for further details. Since our task included only 4 trials each with or without reminders, we counterbalanced the assignment of odd or even target values to these conditions.

Together, there were three key conditions in our task, presented intermixed throughout the experiment: the Forced Internal condition (FI; 4 trials) in which participants had to remember the circles unaided, the Forced External condition (FE; 4 trials) in which they had to use reminders, and the Choice Only (CO; 8 trials) condition in which they were free to choose whichever strategy they preferred but the trial ended after only six circles and without any special circles. To give participants the impression of maximum agency over the task, we only told them that their choice would be overwritten whenever there was a mismatch with the pseudorandomly assigned condition (25.3% of all trials; SD = 5.8; see **Figure 6C**). This way, participants were unable to tell which condition they were currently in and whether it would be a partial trial. Participants used reminders on average on 49.9% of trials (SD = 16.5).

Participants were asked to rate their confidence once during the experiment, being asked to indicate the "percentage of the special circles [they] can correctly drag to the instructed side of the square" (**Figure 6D**). Importantly, this confidence judgement was given after the first practice trials and before the offloading strategy was introduced to ensure participants answered this question with regard to their own perceived memory capabilities. Average confidence was 55.6% (proportion of trials on which participants predicted to remember to move the special circles; SD = 24.2).

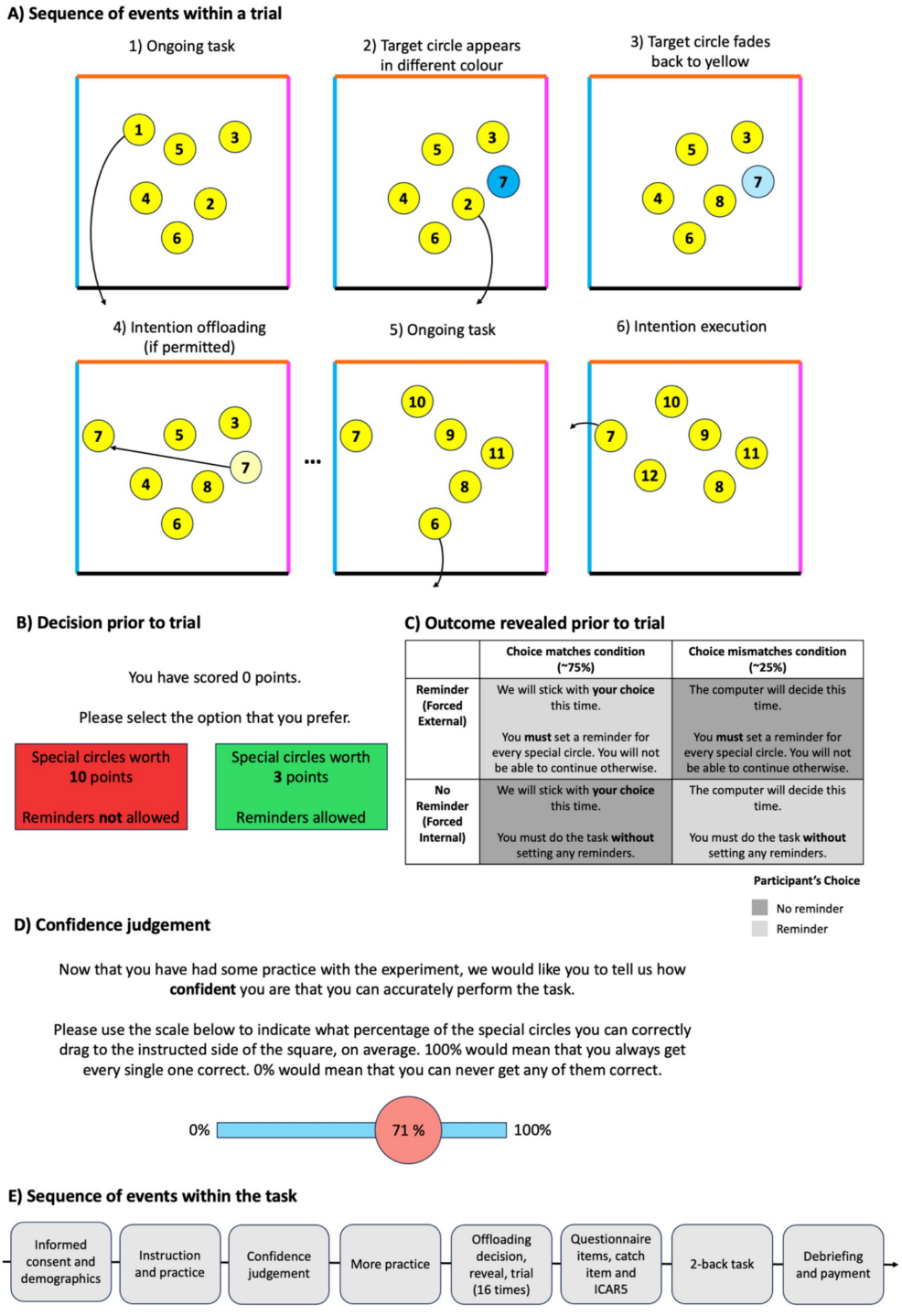


Figure 6

A) Example sequence of events within a single trial. Trajectories of movement made by a fictive participant are shown as black arrows. The blue coloured circle corresponds to the left boundary of the square and indicates that this circle must be moved to this side rather than the bottom. B) Example of an offloading decision which participants were required to make before each trial. C) After each decision, they were informed whether or not they would perform the upcoming trial with reminders. The cell's shading indicates the participant's original choice. D) Confidence was rated once before the introduction of the offloading strategy on a scale ranging from 0 to 100%. E) Sequence of events within the task. All aspects of the task were performed online in the web browser.

In addition to the reminder task, we included items from six individual differences questionnaires shortened to include only the items required to reliably measure the CIT and AD factors (Wise & Dolan, 2020). These questionnaires were presented in random order: 4 items from the Apathy Evaluation Scale (AES; Marin, Biedrzycki, & Firinciogullari, 1991), 8 items from the Zung Depression Scale (SDS; Zung, 1965), 4 items from the Eating Attitudes Test (EAT-26; Garner, Olmsted, Bohr, & Garfinkel, 1982), 12 items from the Barratt Impulsiveness Scale (BIS-11; Patton, Stanford, & Barratt, 1995), 11 items from the Obsessive Compulsive Inventory — Revised (OCIR; Foa, Huppert, Leiberg, Hajcak, Langner, et al., 2002), and 11 items from the 'trait' part of the State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). A list of all included items can be found in the Supplement.

We furthermore included a catch item in the BIS-11 ("I competed in the 1917 Summer Olympics Games.") to ensure participants were paying attention to the task as well as three covariates aimed at measuring cognitive ability (a 5-item version of the International Cognitive Ability Resource; ICAR5; Kirkegaard & Bjerrekær, 2016; Condon & Revelle, 2014); educational attainment mapped onto a 1 to 9 scale and based on the ISCED 2011 categories (see Supplement), and working memory (100 consecutive letters of the 2-back task; e.g. Kirchner, 1985). The logic behind including the latter covariate was that whilst our key dependent variables already corrected for working memory (more specifically: unaided prospective memory performance) this could tap into additional working memory components not measured already and potentially impacted in compulsivity based on the finding that they have often been found to be impaired in OCD (Harkin & Kessler, 2011). Together, these elements resulted in a total duration of approximately 35 mins. The sequence of events within the task is shown in **Figure 6E**.

Participants

Participants were invited on prolific.co to participate for £3.90. Based on points won during the main task, the upper 50% of participants were furthermore rewarded with a bonus payment of £1. We restricted our search to the Prolific standard sample, allowed participants from all countries, with a minimum of 18 years. All participants had to be fluent in English and were required to have an approval rate of over 90% based on Prolific's criteria. Moreover, we required participants to not have participated in one of the four pilots prior to this study.

We calculated our sample size based on the link between confidence and transdiagnostic compulsivity as reported in two recent studies (Rouault et al., 2018; Seow & Gillan 2020). To be able to detect a link between these variables of $\beta = 0.23$, $p < 0.001$, as in Rouault et al. (2018), we required $N = 288$ participants (two-sided testing, power = 0.8, CL = 0.95). To be able to detect a link of $\beta = 6.74$, $p < 0.001$, as in Seow and Gillan (2020), we required $N = 291$ participants (two-sided testing, power = 0.8, CL = 0.95). In both cases, the power calculation was based on a partial regression approach, excluding the effect in question from the model and comparing the explained variance compared to the full model. Since we are furthermore aiming to test a moderation effect of compulsivity on the link between the metacognitive bias and the reminder bias, we decided to collect a larger sample of $N = 600$ after exclusions.

We preregistered six exclusion criteria, based on which we excluded and replaced 69 participants: Nine participants were excluded due to a higher hit rate on forced internal than forced external trials, 22 participants were excluded due to less that 70% accuracy on FE trials, and 3 participants due to less than 10% accuracy on FI trials. We furthermore preregistered to exclude participants with a negative correlation between value and reminder choice (1 = reminder, β = no reminder), as this would indicate participants did not understand the instructions: in order to maximise points in our task, participants should preferentially choose reminders when this strategy brings a higher number of points. Based on this, we excluded 40 participants. No participants were excluded based on scoring lower or higher 3 times the median absolute deviation (MAD), calculated separately based on both the reminder bias and the metacognitive bias. Finally, we excluded 9 participants because they failed to answer *with* "Do not agree at all" to the catch item.

Figure 1 visualizes the exclusions shown in red. In total, we excluded 10.3% of all participants. There were an additional 26 participants excluded for technical reasons, raising the exclusion rate to 13.7%.

Key dependent variables

Our task allowed us to calculate several dependant variables relevant in the context of our study question. The first is the *OIP*, the optimal indifference point. The OIP describes the number of points at which an unbiased, reward-maximising participant is indifferent between the two strategies (reminders or no reminders) and is calculated as:

$$OIP = (10 * ACC_{FI})/ACC_{FE}$$

Where ACC_{FE} is the accuracy measured during trials in which participants had to solve the task using reminders (FE condition) and ACC_{FI} is the accuracy measured during trials in which participants had to solve the task without reminders (FI condition). In contrast, the *AIP* is the actual indifference point, which is the point cut-off at which participants actually were indifferent and is operationalised as the threshold parameter α from fitting a psychometric function to the choice data (target values predicting the decision whether or not to use reminders). Fitting was done using the *quickpsy* package in *R*.

Setting the OIP and the AIP in relation, we can calculate the *reminder bias*, reflecting participants' tendency to use reminders corrected for their actual performance and calculated as the difference between both indifference points:

$$bias_{rem} = OIP - AIP$$

Positive values reflecting that people set more reminders relative to the optimal strategy. The fourth measure is the *metacognitive bias*, reflecting participants' over- or underconfidence relative to their performance and was calculated as:

$$bias_{meta} = confidence - ACC_{FI}$$

Negative values can be interpreted as underconfidence.

Crucially, our study relies on the key assumption that the metacognitive bias can predict the reminder bias, but ACC_{FI} contributes to both biases. To avoid circularity, we therefore split the accuracy data to avoid potentially inflating the correlation. More specifically, we included only the even trials to calculate the ACC_{FI} for the OIP, whereas we included only the odd trials to calculate the ACC_{FI} for the metacognitive bias. All available trials from the FE condition were used to calculate the OIP.

It should be noted that we had incorrectly stated in the preregistration that accuracy from forced external trials would contribute to the calculation of the metacognitive bias. However, the metacognitive bias is a judgement given about the unaided memory performance, in fact confidence is measured before participants were even introduced to the offloading strategy (see above). We therefore used only the internal trials in calculating the metacognitive bias.

Finally, the transdiagnostic scores for the 'compulsive behaviour and intrusive thought' (CIT) factor and the 'anxious-depression' (AD) factor were calculated from participants' ratings to the individual differences questionnaires by multiplying them with the item weights from Wise and Dolan (2020) prior to summing them. The items composing the CIT and AD scores respectively were non-overlapping with 24 items forming the AD score and 25 items forming the CIT score.

Preregistered Hypotheses and Statistical Analyses

We preregistered 8 hypotheses (see **Table 1**), half of which were sanity checks (H1-H4) aimed to establish whether our task would generally lead to the same patterns as previous studies using a similar task (as reviewed in Gilbert et al., 2023). H1 was a replication of the central finding of the link between confidence and offloading. More specifically, we entered the unconfounded metacognitive bias and reminder bias into a Pearson correlation analysis. We expected to find a negative relationship between the two measures, which we planned to test for significance using a one-sided test. We furthermore expected to find that people would use more reminders than optimal. This pro-reminder bias would be reflected in a positive reminder bias (H2). We planned to test this using a one-sided paired t-test. Relatedly, we expected to find people to be generally underconfident (i.e., expecting to remember fewer special circles than they actually did when doing the task without reminders). Such underconfidence would be reflected in a negative metacognitive bias (H3), which we again planned to test this using a one-sided paired t-test. Furthermore, we expected that those who required more reminders would also be the ones to use them more, as reflected in a positive correlation between the AIP and OIP, again as a one-sided test (H4). We decided to use Spearman's rho due to the data most likely being distributed around the extremes of the scale. For H2-H4 (as well as H5, H6, and H8, see below), we used the biases and indifference points calculated from all available trials as there was no circularity issue.

Hypotheses H5-H8 were the key hypotheses of our study. Here, we address them out of order in the interest of an improved logical flow. Hypothesis H6 predicted that more compulsive individuals would show a confidence bias, reflected in a reliable predictor of the CITs scores on the metacognitive bias from the following regression model:

$$bias_{meta} \sim CIT + AD + age + gender + education + \varepsilon$$

Though we did not preregister a direction for this effect, in the light of recent findings it has now become clear that compulsivity would most likely be linked to overconfidence (Rouault et al., 2018; Seow & Gillan, 2020; Benwell et al., 2022; Fox et al., 2023a; Fox et al., 2023b; Hoven et al., 2023a). The same model was used to test hypothesis H8, predicting that more anxious-depressed individuals tend to be underconfident. This would be reflected in AD scores being negatively linked to the metacognitive bias. The model above represents the main models designed to test hypotheses H6a and H8a. We furthermore also tested these hypotheses but predicted raw confidence (percentage of circles participants predicted they would remember; H6b and H8b respectively), as well as extending the main model with the scores from the cognitive ability test (ICAR5) as an additional covariate (H6c and H8c respectively). For this as well as all following regression models, we *z*-transformed all non-binary variables prior to fitting the models.

With H5, we predicted that more compulsive individuals would show a bias towards more offloading, reflected in a positive regression coefficient when using the CIT score as a predictor of the reminder bias. This hypothesis was not a replication, consequently we decided to carry out the test two-sided. Throughout this section, whenever not explicit specified, we plan to carry out a test two-sided. Due to the diametrically opposing effects of CIT and AD, both transdiagnostic scores need to be entered into the model, alongside our demographic covariates age, gender, and educational attainment:

$$bias_{rem} \sim CIT + AD + age + gender + education + \varepsilon$$

We fitted several different versions of this model: the main model predicted the reminder bias (H5a), but we also fit one with the absolute number of reminders chosen (H5b), or the AIP (H5c). To understand whether any differences in offloading behaviour could stem from differences in working memory capacity not already captured by our correction for unaided task performance,

Number	Hypothesis	Support?
H1	The reminder bias and metacognitive bias are negatively correlated.	yes
H2	Participants use reminders excessively.	yes
H3	Participants are underconfident in their own memory.	yes
H4	OIP and AIP are positively correlated.	yes
H5a	Positive link between CIT and reminder bias.	no (significant negative effect)
H5b	Positive link between CIT and absolute number of reminders chosen.	no (significant negative effect)
H5c	Positive link between CIT and AIP.	no (significant negative effect)
H5d	Positive link between CIT and reminder bias even if working memory is included as a covariate.	no (significant negative effect)
H5e	Positive link between CIT and reminder bias even if cognitive ability is included as a covariate.	no (significant negative effect)
H6a	A significant link exists between CIT and metacognitive bias (preregistered as a two-sided test, so either more or less confident).	yes (positive)
H6b	A significant link exists between CIT and raw confidence.	yes (positive)
H6c	A significant link exists between CIT and metacognitive bias even if cognitive ability is included as a covariate.	yes (positive)
H7a	CIT acts as a moderator on the link between confidence and offloading. In other words, we expect to find that the correlation between the metacognitive and the reminder bias to be weakened in highly compulsive individuals.	no
H7b	CIT acts as a moderator on the link between confidence and offloading even if working memory is included as a covariate.	no
H7c	CIT acts as a moderator on the link between confidence and offloading even if cognitive ability is included as a covariate.	no
H8a	A significant negative link exists between AD and metacognitive bias (i.e., more anxious-depressed individuals tend to be underconfident).	yes
H8b	A significant negative link exists AD and raw confidence.	yes
H8c	A significant negative link exists between AD and metacognitive bias even if cognitive ability is included as a covariate.	yes

Table 1

List of preregistered hypotheses together with the empirical support our study found. Lighter grey shading indicates sanity check hypotheses and darker grey shading indicates key hypotheses. OIP = optimal indifference point. AIP = actual indifference point. CIT = compulsive behaviour and intrusive thought.

we furthermore extended the main model by also including the d' from a 2- back task as a covariate (H5d). Finally, we fit an extended version of the main model with scores from the cognitive ability test (ICAR5) as an additional covariate to capture cognitive ability (H5e). We ran the same analysis but for the AD factor. We included this test as a preregistered analysis but did not specify any directional hypotheses.

Our final hypothesis, H7, aimed to differentiate between the *Metacognitive Monitoring Mechanism*, the *Metacognitive Control Mechanism*, and the *Direct Mechanism*. We tested how compulsivity would affect the relationship between confidence and offloading. More specifically, we predicted that CIT scores would act as a moderator variable between the metacognitive and the reminder bias and that highly compulsive individuals would have a weaker link. We tested this by fitting the following regression model to the data:

$$bias_{rem} \sim bias_{meta} * CIT + AD + age + gender + education + \varepsilon$$

To avoid circularity, we used the unconfounded metacognitive bias and reminder bias for this analysis. The moderation of CIT is reflected in its interaction term with the $bias_{meta}$ predictor. A significant interaction term can be interpreted as support for the *Metacognitive Control Mechanism*. In addition to this main model (H7a), we furthermore also tested whether this effect would persist if working memory (2-back d'; H7b) or educational attainment (H7c) were included as additional covariates. We ran the same analysis but for the AD factor. We included this test as a preregistered analysis but did not specify any directional hypotheses.

It should be noted that whilst not explicitly preregistered, our planned models also allow testing for a mediation effect (metacognitive bias acting as a mediator on the effect of the CIT score on the reminder bias). This is done by comparing the effect of CIT on the reminder bias when the effect of the metacognitive bias is accounted for (Hypothesis 7) to when it is not (Hypothesis 5). In addition, we included a causal mediation analysis (not preregistered), using the *mediation* package in *R*. This analysis involved testing of the indirect effect using bootstrapping. More specifically, we computed unstandardised indirect effects for each of our 1000 bootstrapped samples and based on those the 95% confidence interval. To keep the information entering into the mediation analysis constant, we re-fitted the models from our sections on H5 and H6/H8 but with the unconfounded metacognitive bias and reminder bias respectively. Furthermore, we had to treat the covariate 'gender' as a continuous variable as the *mediation* package would otherwise not have been able to fit the data. We expect that this difference is unlikely to cause any issues with the interpretation of our effects. A significant mediation effect can be interpreted as support for the *Metacognitive Monitoring Mechanism.* A significant direct effect can be interpreted as support for the *Direct Mechanism*.

Supplementary Material

Table S1

Predicting metacognitive bias. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*t*	*p*
Intercept	0.01	0.05	0.21	0.831
AD	-0.22	0.04	-5.00	< 0.001
CIT	0.15	0.05	3.30	0.001
Age	-0.02	0.05	-0.43	0.666
gender1 (m vs. f)	-0.17	0.08	-2.06	0.040
gender2 (m vs. o)	-0.28	0.37	-0.76	0.447
education	-0.002	0.04	-0.05	0.960

Table S2

Predicting confidence. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*t*	*p*
Intercept	0.06	0.05	1.26	0.209
AD	-0.29	0.04	-6.44	< 0.001
CIT	0.16	0.05	3.50	< 0.001
Age	-0.12	0.05	-2.21	0.028
gender1 (m vs. f)	-0.24	0.08	-2.97	0.003
gender2 (m vs. o)	-0.20	0.37	-0.54	0.589
education	0.03	0.04	-0.72	0.474

Table S3

Predicting metacognitive bias with ICAR5 scores as an additional covariate. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*t*	*p*
Intercept	0.03	0.05	0.70	0.487
AD	-0.20	0.04	-4.47	< 0.001
CIT	0.12	0.04	2.74	0.006
Age	-0.04	0.05	-0.75	0.454
gender1 (m vs. f)	-0.21	0.08	-2.61	0.009
gender2 (m vs. o)	-0.43	0.36	-1.20	0.232
education	0.03	0.04	0.87	0.383
ICAR5	-0.19	0.04	-4.87	< 0.001

Table S4

Predicting reminder bias. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*t*	*p*
Intercept	-0.02	0.05	-0.31	0.756
AD	0.06	0.04	1.43	0.153
CIT	-0.16	0.04	-3.60	< 0.001
Age	-0.13	0.05	-2.43	0.015
gender1 (m vs. f)	0.02	0.08	0.31	0.758
gender2 (m vs. o)	0.84	0.37	2.30	0.022
education	-0.05	0.04	-1.18	0.237

Table S5

Predicting absolute number of reminders. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*t*	*p*
Intercept	-0.03	0.05	-0.65	0.518
AD	0.06	0.05	1.32	0.187
CIT	-0.14	0.05	-3.08	0.002
Age	-0.04	0.05	-0.72	0.474
gender1 (m vs. f)	0.10	0.08	1.14	0.255
gender2 (m vs. o)	0.69	0.38	1.82	0.070
education	-0.08	0.04	-2.08	0.038

Table S6

Predicting AIP. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*t*	*p*
Intercept	0.02	0.05	0.39	0.700
AD	-0.08	0.04	-1.74	0.082
CIT	0.15	0.05	3.32	< 0.001
Age	0.05	0.05	0.94	0.350
gender1 (m vs. f)	-0.06	0.08	-0.73	0.465
gender2 (m vs. o)	-0.70	0.37	-1.89	0.059
education	0.07	0.04	1.79	0.073

Table S7

Predicting reminder bias with 2-back d' as an additional covariate. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*T*	*p*
Intercept	-0.02	0.05	-0.40	0.686
AD	0.05	0.04	1.21	0.228
CIT	-0.15	0.05	-3.27	0.001
Age	-0.13	0.05	-2.46	0.014
gender1 (m vs. f)	0.02	0.08	0.31	0.758
gender2 (m vs. o)	0.81	0.36	2.24	0.026
education	-0.05	0.04	-1.31	0.190
2-back d'	0.10	0.04	2.38	0.018

Table S8

Predicting reminder bias with ICAR5 scores as an additional covariate. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other.

	β	SE	*T*	*p*
Intercept	-0.02	0.05	-0.32	0.748
AD	0.06	0.04	1.40	0.161
CIT	-0.16	0.05	-3.55	< 0.001
Age	-0.13	0.05	-2.42	0.016
gender1 (m vs. f)	0.03	0.08	0.32	0.749
gender2 (m vs. o)	0.84	0.37	2.30	0.022
education	-0.05	0.04	-1.19	0.236
2-back d'	0.01	0.04	0.13	0.899

Table S9

Predicting reminder bias with metacognitive bias as an additional covariate (i.e., testing for a moderation effect). All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other; MetaBias = metacognitive bias.

	β	SE	*T*	*p*
Intercept	-0.03	0.05	-0.60	0.552
Metacognitive Bias	-0.19	0.04	-4.56	< 0.001
AD	0.03	0.05	0.66	0.508
CIT	-0.14	0.05	-3.14	0.002
Age	-0.10	0.05	-1.93	0.054
gender1 (m vs. f)	0.03	0.08	0.35	0.725
gender2 (m vs. o)	0.94	0.37	2.58	0.010
education	-0.06	0.04	-1.64	0.101
CIT X MetaBias	-0.01	0.04	-0.18	0.860

Table S10

Predicting reminder bias with metacognitive bias and 2-back d' as additional covariates. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other; MetaBias = metacognitive bias.

	β	SE	*T*	*p*
Intercept	-0.03	0.05	-0.67	0.504
Metacognitive Bias	-0.17	0.04	-4.19	< 0.001
AD	0.02	0.05	0.55	0.582
CIT	-0.13	0.05	-2.92	0.004
Age	-0.10	0.05	-1.95	0.051
gender1 (m vs. f)	0.03	0.08	0.38	0.707
gender2 (m vs. o)	0.92	0.37	2.53	0.012
education	-0.06	0.04	-1.73	0.084
2-back d'	0.08	0.04	1.86	0.064
CIT X MetaBias	-0.01	0.04	-0.26	0.798

	β	SE	*T*	*p*
Intercept	-0.03	0.05	-0.60	0.552
Metacognitive Bias	-0.19	0.04	-4.50	< 0.001
AD	0.03	0.05	0.66	0.512
CIT	-0.14	0.05	-3.12	0.002
Age	-0.10	0.05	-1.93	0.055
gender1 (m vs. f)	0.03	0.08	0.35	0.725
gender2 (m vs. o)	0.94	0.37	2.57	0.010
education	-0.06	0.04	-1.62	0.107
ICAR5	0.001	0.04	0.03	0.975
CIT X MetaBias	-0.01	0.04	-0.18	0.860

Table S11

Predicting reminder bias with metacognitive bias and ICAR5 scores as additional covariates. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other; MetaBias = metacognitive bias.

Table S12

Predicting reminder bias with metacognitive bias and ICAR5 scores as additional covariates. All continuous variables are z-transformed. SE = standard error; m = male; f = female; o = other; MetaBias = metacognitive bias.

	β	SE	*T*	*p*
Intercept	-0.04	0.05	-0.73	0.465
Metacognitive Bias	-0.19	0.04	-4.54	< 0.001
AD	0.03	0.05	0.62	0.539
CIT	-0.14	0.05	-3.11	0.002
Age	-0.10	0.05	-1.93	0.054
gender1 (m vs. f)	0.03	0.08	0.35	0.724
gender2 (m vs. o)	0.94	0.37	2.58	0.010
education	-0.06	0.04	-1.51	0.132
AD X MetaBias	-0.04	0.04	-1.13	0.258

Questionnaire Items

Apathy Evaluation Scale (AES)

Response scores:

- Not at all characteristic (1)
- Slightly characteristic (2)
- Somewhat characteristic (3)
- Very characteristic (4)

- ***AES_2.*** I get things done during the day. (Reverse)
- ***AES_7.*** I approach life with intensity. (Reverse)
- ***AES_17.*** I have initiative. (Reverse)
- ***AES_18.*** I have motivation. (Reverse)

Barrett's Impulsivity Scale (BIS)

Response scores:

- Rarely/never (1)
- Occasionally (2)
- Often (3)
- Almost always/Always (4)

Items:

- ***BIS_1.*** I plan tasks carefully. (Reverse)
- ***BIS_6.*** I have "racing" thoughts.
- ***BIS_9.*** I concentrate easily. (Reverse)
- ***BIS_13.*** I plan for job security. (Reverse)
- ***BIS_14.*** I say things without thinking.
- ***BIS_15.*** I like to think about complex problems. (Reverse)
- ***BIS_17.*** I act "on impulse".
- ***BIS_20.*** I am a steady thinker. (Reverse)
- ***BIS_22.*** I buy things on impulse.
- ***BIS_25.*** I spend or charge more than I earn.
- ***BIS_26.*** I often have extraneous thoughts when thinking.
- ***BIS_check.*** I competed in the 1917 Summer Olympics Games.

Eating Attitudes Test (EAT)

Response scores:

- Always (3)
- Usually (2)
- Often (1)
- Sometimes (0)
- Rarely (0)
- Never (0)

Items:

- ***EAT_1.*** I am terrified about being overweight.
- ***EAT_11.*** I am preoccupied with a desire to be thinner.
- ***EAT_12.*** I think about burning up calories when I exercise.
- ***EAT_14.*** I am preoccupied with the thought of having fat on my body.

Obsessive Compulsive Inventory (OCI)

Response scores:

- Not at all (0)
- A little (1)
- Moderately (2)
- A lot (3)
- Extremely (4)

Items:

- ***OCI_1.*** I have saved up so many things that they get in the way.
- ***OCI_2.*** I check things more often than necessary.
- ***OCI_4.*** I feel compelled to count while I am doing things.
- ***OCI_6.*** I find it difficult to control my own thoughts.
- ***OCI_7.*** I collect things I don't need.
- ***OCI_9.*** I get upset if others change the way I have arranged things.
- ***OCI_11.*** I sometimes have to wash or clean myself simply because I feel contaminated.
- ***OCI_12.*** I am upset by unpleasant thoughts that come into my mind against my will.
- ***OCI_13.*** I avoid throwing things away because I am afraid I might need them later.
- ***OCI_16.*** I feel that there are good and bad numbers.
- ***OCI_18.*** I frequently get nasty thoughts and have difficulty in getting rid of them.

Self-rating Depression Scale (SDS)

Response scores:

- A little of the time (1)
- Some of the time (2)
- Good part of the time (3)
- Most of the time (4)

Items:

- ***SDS_11.*** My mind is as clear as it used to be. (Reverse)
- ***SDS_12.*** I find it easy to do the things I used to. (Reverse)
- ***SDS_13.*** I am restless and can't keep still.
- ***SDS_14.*** I feel hopeful about the future. (Reverse)
- ***SDS_16.*** I find it easy to make decisions. (Reverse)
- ***SDS_17.*** I feel that I am useful and needed. (Reverse)
- ***SDS_18.*** My life is pretty full. (Reverse)
- ***SDS_20.*** I still enjoy the things I used to do. (Reverse)

State Trait Anxiety Inventory (STAI)

Response scores:

- Almost never (1)
- Sometimes (2)
- Often (3)
- Almost always (4)

Items:

- ***STAI_1.*** I feel pleasant. (Reverse)
- ***STAI_3.*** I feel satisfied with myself. (Reverse)
- ***STAI_5.*** I feel like a failure.
- ***STAI_8.*** I feel that difficulties are piling up so that I cannot overcome them.
- ***STAI_9.*** I worry too much over something that really doesn't matter.
- ***STAI_10.*** I am happy. (Reverse)
- ***STAI_12.*** I lack self-confidence.
- ***STAI_13.*** I feel secure. (Reverse)
- ***STAI_16.*** I am content. (Reverse)
- ***STAI_19.*** I am a steady person. (Reverse)
- ***STAI_20.*** I get in a state of tension or turmoil as I think over my recent concerns and interests.

Educational Attainment Questions

Educational attainment was based on the ISCED 2011 categories and included the following options mapped onto 1 to 9 in response to the question "What is the highest level of education you have completed to this date?":

1. "Early childhood education or no formal education (e.g. early childhood education and development, play school, reception, pre-primary, pre-school, educación inicial)"
2. "Primary education (e.g. primary education, elementary education, basic education; typically ends around age 10 to 12 years)"
3. "Lower secondary education (e.g. lower grades of secondary school, junior secondary school, middle school, junior high school)"
4. "Upper secondary education (e.g. upper grades of secondary school, senior secondary school, senior high school; typically ends around age 17 to 18 years)"
5. "Post-secondary non-tertiary education (e.g. technician diploma, primary professional education, préparation aux carrières administratives; usually designed for direct labour market entry)"
6. "Short-cycle tertiary education (e.g. junior college, higher technical education, community college education, technician or advanced/higher vocational training, associate degree, bac+2; practically-based, occupationally-specific and prepare for the labour market but can also be a pathway to other tertiary education programmes)"
7. "Bachelor's or equivalent level"
8. "Master's or equivalent level"
9. "Doctoral or equivalent level"

Planned additional analyses

We preregistered several analyses that whilst planned were not thought out in detail and thus of a more exploratory nature.

First, we investigated the relationship between the transdiagnostic phenotypes and unaided task performance (accuracy on trials in which participants were not allowed to use a reminder). We predicted task performance (accuracy on FI trials) from the two transdiagnostic factors and our demographic covariates (age, gender, and education).

$$ACC_{FI} \sim CIT + AD + age + gender + education + \varepsilon$$

Neither AD, β = −0.02, SE = 0.05, t = −0.44, p = 0.66; nor CIT, β = −0.03, SE = 0.05, t = −0.66, p = 0.51, were significant predictors, suggesting there was no evidence to support meaningful performance differences for these two transdiagnostic phenotypes.

Second, we planned to investigate whether participants would show a response "stickiness" in their reminder use as reported by Scarampi & Gilbert (2020) and whether such response perseverance would correlate with their CIT (e.g., Shahar et al., 2021) and AD scores. We conducted this analysis based on a subset of participants: only participants whose first trial was not a partial trial were included (choice condition, as no strategy was performed that could later be repeated). Furthermore, only participants who indicated with their strategy choice that they would have chosen the strategy they were randomly assigned were included. This resulted in a subset of N=157 participants. We then calculated the proportion of the remaining trials in which participants chose to repeat this strategy and compared the resulting proportion to 50%.

Participants repeated this strategy on only 31.6% of the remaining trials, significantly lower than 50%, $t(156)$ = −8.43, p < 0.001, d = 0.67. We need to keep in mind that the strategy in question was the one they did not choose on the first trial. Most people have an overall preference for or against reminders in this task independent of the manipulation of reward. A person who shows numbers lower than 50% might thus have repeatedly chosen the same strategy they had overwritten on that first trial, reflecting stable biases for their preferred response strategies. We furthermore tested whether this response perseverance was predicted by CIT or AD (and demographic covariates) fitting the following regression model to the data:

$$perseverance \sim CIT + AD + age + gender + education + \varepsilon$$

We indeed found that the effect was modulated by CIT: Compulsive individuals showed more response perseverance, β = 0.21, SE = 0.09, t = 2.17, p = 0.03, whereas there was no significant effect of AD, β = −0.08, SE = 0.09, t = −0.85, p = 0.40. As a conclusion, we can say that the more compulsive individuals show a tendency to repeat that first forced/overwritten trial later in the remaining experiment. However, caution should be advised to not over-interpret this effect as only a small subset of our sample was included in the analysis.

Exploratory analyses

During our analysis, several additional questions arose, which we aimed to address with exploratory analyses.

First, we aimed to understand whether highly compulsive individuals would approach our task differently, potentially even struggling with it. For instance, some OCD patients prefer ordered sequences, and it might therefore have been aversive for our compulsive individuals to move the target circles out of the numbered order to set reminders. Additionally, they might have been put

off by the scattered nature of the visual display and might have spent time rearranging circles, for example in a grid-like fashion. We therefore tested whether the transdiagnostic phenotypes were reliable predictors of response times (RTs), depending on condition with the following linear mixed model with random intercepts for participants:

$$RT \sim condition * (CIT + AD) + age + gender + education + \varepsilon$$

Where *condition* denoted whether or not participants did the task with (FE) or without (FI) reminders. CIT did not predicted RT at our significance cut-off, β = −0.04, SE = 0.02, $t(1142) = -1.91$, $p = 0.06$. The same was true for AD, β = −0.04, SE = 0.02, $t(1176) = -1.89$, $p = 0.06$.

However, there was a crucially significant interaction between CIT and *condition*, β = 0.08, SE = 0.02, $t(9566) = 3.35$, $p < 0.001$. In other words, when reminders were possible, compulsive individuals were slower (0.04); but when reminders were not possible, compulsive individuals were faster (-0.04). There was no such interaction effect for the AD factor, β = −0.02, SE = 0.02, $t(9561) = -0.96$, $p = 0.34$.

We followed up this analysis by fitting two additional linear mixed models with random intercepts for participants to gain insight into what actions highly compulsive individuals might have been performing during the reminder trials that might have led to the increase in RT. The first model predicted the trial-wise number of times that participants re-arranged a circle they had previously already moved:

$$circles_{movedagain} \sim condition * CIT + AD + age + gender + education + \varepsilon$$

Compulsive individuals showed a tendency to this more often, however, this effect did not reach our level of significance, β = 0.03, SE = 0.02, t (1299) = 1.91, $p = 0.06$.

The second model focused on a smaller subset of trials (m = 4.2 trials per participant; min = 4, max = 12) in which more than one circle was moved and expressed the extent to which these circles were moved in their numbered order:

$$circles_{movedinorder} \sim condition * CIT + AD + age + gender + education + \varepsilon$$

Contrary to expectation, highly compulsive individuals showed a reduced tendency to move circles in their numbered order, β = −0.10, SE = 0.04, t (943.7) = −2.69, $p = 0.007$. Taken together, high CIT individuals took significantly longer on reminder trials, but we cannot say with certainty why and this will therefore need to be the focus of future studies.

Second, we asked whether the transdiagnostic phenotypes affected the compensatory nature of reminder use (cf. Hypothesis 4), meaning people who need reminders more tend to be the ones who use them more. The motivation for this analysis was the compulsive individual's tendency towards a reduced reminder bias. To this end, we fit a regression model to predict the AIP from the OIP, the transdiagnostic phenotypes, and the demographic covariates:

$$AIP \sim OIP * (CIT + AD) + age + gender + education + \varepsilon$$

The key effects of interest were the interaction terms between the transdiagnostic phenotypes and the OIP.

Indeed, this seems to be the case reflected numerically in the interaction between CIT and OIP when predicting AIP, in other words, there was a stronger link between AIP and OIP in low compulsive individuals, compared to high compulsive individuals. However, this effect did not

reach our required level of significance, β = −0.08, SE = 0.04, t = −1.83, p = 0.07. There was no such interaction effect for the AD factor, β = −0.002, SE = 0.04, t = 0.06, p = 0.95. In a follow-up analysis, we then investigated the influence of compulsivity on accuracy in the FE condition by fitting the following regression model:

$$ACC_{FE} \sim CIT + AD + age + gender + education + \varepsilon$$

Compulsive individuals were found to be less accurate on reminder trials, β = −0.10, SE = 0.5, t = −2.06, p = 0.04, pointing towards a picture of not only impaired reminder setting but also impaired reminder use. We note that whilst this is an interesting finding, it certainly needs follow-up in future studies to understand the mechanisms at play.

References

Boldt A., Gilbert S. J. (2019) **Confidence guides spontaneous cognitive offloading** *Cognitive Research: Principles and Implications* **4** https://doi.org/10.1186/s41235-019-0195-y

Boldt A., Gilbert S. J. (2022) **Partially Overlapping Neural Correlates of Metacognitive Monitoring and Metacognitive Control** *The Journal of Neuroscience* **42**:3622–3635 https://doi.org/10.1523/JNEUROSCI.1326-21.2022

Benwell C. S. Y., Mohr G., Wallberg J., Kouadio A., Ince R. A. A. (2022) **Psychiatrically relevant signatures of domain-general decision-making and metacognition in the general population** *Npj Mental Health Research* **1** https://doi.org/10.1038/s44184-022-00009-4

Condon D. M., Revelle W. (2014) **The international cognitive ability resource: Development and initial validation of a public-domain measure** *Intelligence* **43**:52–64 https://doi.org/10.1016/j.intell.2014.01.004

Crawford J., Smith G., Maylor E., Della Sala S., Logie R. (2003) **The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample** *Memory* **11**:261–275 https://doi.org/10.1080/09658210244000027

Den Ouden L., Suo C., Albertella L., Greenwood L. M., Lee R. S. C., Fontenelle L. F., Yücel M. (2022) **Transdiagnostic phenotypes of compulsive behavior and associations with psychological, cognitive, and neurobiological affective processing** *Translational Psychiatry* **12**:1–11 https://doi.org/10.1038/s41398-021-01773-1

Dupont D., Zhu Q., Gilbert S. J. (2022) **Value-Based Routing of Delayed Intentions Into Brain-Based Versus External Memory Stores** *Journal of Experimental Psychology: General* **152**:175–187 https://doi.org/10.1037/xge0001261

Egan S. J., Wade T. D., Shafran R. (2011) **Perfectionism as a transdiagnostic process: A clinical review** *Clinical Psychology Review* **31**:203–212 https://doi.org/10.1016/j.cpr.2010.04.009

Engeler N. C., Gilbert S. J. (2020) **The effect of metacognitive training on confidence and strategic reminder setting** *PLoS ONE* **15**:1–14 https://doi.org/10.1371/journal.pone.0240858

Foa E. B., Huppert J. D., Leiberg S., Langner R., Kichic R., Hajcak G., Salkovskis P. M. (2002) **The obsessive-compulsive inventory: Development and validation of a short version** *Psychological Assessment* **14**:485–496 https://doi.org/10.1037/1040-3590.14.4.485

Fox C. A., Lee C. T., Hanlon A. K., Seow T. X. F., Lynch K., Harty S., Gillan C. M. (2023) **An observational treatment study of metacognition in anxious-depression** *ELife* **12**:1–17 https://doi.org/10.7554/eLife.87193

Fox C. A., McDonogh A., Donegan K. R., Teckentrup V., Crossen R. J., Hanlon A. K., Gillan C. M. (2024) **Reliable, rapid, and remote measurement of metacognitive bias** *Scientific Reports* **14** https://doi.org/10.1038/s41598-024-64900-0

Garner D. M., Bohr Y., Garfinkel P. E. (1982) **The Eating Attitudes Test: Psychometric Features and Clinical Correlates** *Psychological Medicine* **12**:871–878 https://doi.org/10.1017/S0033291700049163

Gilbert S. J. (2015) **Strategic use of reminders: Influence of both domain-general and taskspecific metacognitive confidence, independent of objective memory ability** *Consciousness and Cognition* **33**:245–260 https://doi.org/10.1016/j.concog.2015.01.006

Gilbert S.J., Bird A., Carpenter J., Fleming S.M., Sachdeva C., Tsai P.C. (2020) **Optimal use of reminders: Metacognition, effort, and cognitive offloading** *Journal of Experimental Psychology: General* **149**:501–517

Gilbert S. J., Boldt A., Sachdeva C., Scarampi C., Tsai P. C. (2023) **Outsourcing Memory to External Tools: A Review of 'Intention Offloading'** *Psychonomic Bulletin & Review* https://doi.org/10.3758/s13423-022-02139-4

Gillan C. M., Kosinski M., Whelan R., Phelps E. A., Daw N. D. (2016) **Characterizing a psychiatric symptom dimension related to deficits in goaldirected control** *ELife* **5**:1–24 https://doi.org/10.7554/eLife.11305

Harkin B., Kessler K. (2011) **The role of working memory in compulsive checking and OCD: A systematic classification of 58 experimental findings** *Clinical Psychology Review* **31**:1004–1021 https://doi.org/10.1016/j.cpr.2011.06.004

Harris L. M., Vaccaro L., Jones M. K., Boots G. M. (2010) **Evidence of impaired event-based prospective memory in clinical obsessive-compulsive checking** *Behaviour Change* **27**:84–92 https://doi.org/10.1375/bech.27.2.84

Hoven M., Lebreton M., Engelmann J. B., Denys D., Luigjes J., van Holst R. J. (2019) **Abnormalities of confidence in psychiatry: an overview and future perspectives** *Translational Psychiatry* **9** https://doi.org/10.1038/s41398-019-0602-7

Hoven M., Luigjes J., Denys D., Rouault M., van Holst R. J. (2023) **How do confidence and self-beliefs relate in psychopathology: a transdiagnostic approach** *Nature Mental Health* **1**:337–345 https://doi.org/10.1038/s44220-023-00062-8

Hoven M., Mulder T., Denys D., Holst R. J. Van, Luigjes J. (2023) **OCD patients show lower confidence and higher error sensitivity while learning under volatility compared to healthy and highly compulsive samples from the general population** *PsyArXiv* https://doi.org/10.31234/osf.io/37nad

Hoven M., Rouault M., van Holst R., Luigjes J. (2023) **Differences in metacognitive functioning between obsessive-compulsive disorder patients and highly compulsive individuals from the general population** *Psychological Medicine* :1–10 https://doi.org/10.1017/S003329172300209X

Jones W. E., Benge J. F., Scullin M. K. (2021) **Preserving prospective memory in daily life: A systematic review and meta-analysis of mnemonic strategy, cognitive training, external memory aid, and combination interventions** *Neuropsychology* **35**:123–140 https://doi.org/10.1037/neu0000704

Kirchner W. K. (1958) **Age differences in short-term retention of rapidly changing information** *Journal of Experimental Psychology* **55**:352–358 https://doi.org/10.1037/h0043688

Kirk P. A., Robinson O. J., Gilbert S. J. (2021) **Trait anxiety does not correlate with metacognitive confidence or reminder usage in a delayed intentions task** *Quarterly Journal of Experimental Psychology* **74**:634–644 https://doi.org/10.1177/1747021820970156

Kirkegaard E. O. W., Bjerrekær J. D. (2016) **ICAR5: design and validation of a 5-item public domain cognitive ability test** *Open Differential Psychology* https://doi.org/10.26775/odp.2016.07.11

Loosen A., Seow T. X. F., Hauser T. U. (2022) **Consistency within change: Evaluating the psychometric properties of a widely-used predictive-inference task** *PsyArXiv* https://doi.org/10.31234/osf.io/qkf7j

Marin R. S., Biedrzycki R. C., Firinciogullari S. (1991) **Reliability and validity of the apathy evaluation scale** *Psychiatry Research* **38**:143–162 https://doi.org/10.1016/0165-1781(91)90040-V

Marzuki A. A., Vaghi M. M., Conway-Morris A., Kaser M., Sule A., Apergis-Schoute A., Robbins T. W. (2022) **Atypical action updating in a dynamic environment associated with adolescent obsessive-compulsive disorder** *Journal of Child Psychology and Psychiatry and Allied Disciplines* **63**:1591–1601 https://doi.org/10.1111/jcpp.13628

Mohr G., Ince R.A.A., Benwell C.S.Y. (2024) **Information search under uncertainty across transdiagnostic psychopathology and healthy ageing** *Transl Psychiatry* **14** https://doi.org/10.1038/s41398-024-03065-w

Mountford V., Haase A., Waller G. (2006) **Body checking in the eating disorders: Associations between cognitions and behaviors** *International Journal of Eating Disorders* **39**:708–715 https://doi.org/10.1002/eat.20279

Patton J. H., Stanford M. S., Barratt E. S. (1995) **Factor structure of the Barratt Impulsiveness Scale** *Journal of Clinical Psychology* **51**:768–774 https://doi.org/10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Racsmány M., Demeter G., Csigó K., Harsányi A., Németh A. (2011) **An experimental study of prospective memory in obsessive-compulsive disorder** *Journal of Clinical and Experimental Neuropsychology* **33**:85–91 https://doi.org/10.1080/13803395.2010.493147

Risko E. F., Gilbert S. J. (2016) **Cognitive Offloading** *Trends in Cognitive Sciences* **20**:676–688 https://doi.org/10.1016/j.tics.2016.07.002

Rouault M., Seow T., Gillan C. M., Fleming S. M. (2018) **Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance** *Biological Psychiatry* :1–9 https://doi.org/10.1016/j.biopsych.2017.12.017

Sachdeva C., Gilbert S. J. (2020) **Excessive use of reminders: Metacognition and effortminimisation in cognitive offloading** *Consciousness and Cognition* **85** https://doi.org/10.1016/j.concog.2020.103024

Sauter M., Draschkow D., Mack W. (2020) **Building, hosting and recruiting: A brief introduction to running behavioral experiments online** *Brain Sciences* **10**:1–11 https://doi.org/10.3390/BRAINSCI10040251

Scarampi C., Gilbert S. J. (2020) **The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task** *Memory* **28**:677–691 https://doi.org/10.1080/09658211.2020.1764974

Scullin M. K., Jones W. E., Phenis R., Beevers S., Rosen S., Dinh K., Benge J. F. (2022) **Using smartphone technology to improve prospective memory functioning: A randomized controlled trial** *Journal of the American Geriatrics Society* **70**:459–469 https://doi.org/10.1111/jgs.17551

Seow T. X. F., Gillan C. M. (2020) **Transdiagnostic Phenotyping Reveals a Host of Metacognitive Deficits Implicated in Compulsivity** *Scientific Reports* **10**:1–11 https://doi.org/10.1038/s41598-020-59646-4

Shahar N., Hauser T. U., Moran R., Moutoussis M., Bullmore E. T., Dolan R. J. (2021) **Assigning the right credit to the wrong action: compulsivity in the general population is associated with augmented outcome-irrelevant value-based learning** *Translational Psychiatry* **11** https://doi.org/10.1038/s41398-021-01642-x

Sookud S., Martin I., Gillan C., Wise T. (2024) **Impaired goal-directed planning in transdiagnostic compulsivity is explained by uncertainty about learned task structure** https://doi.org/10.31234/osf.io/zp6vk

Spielberger C. D., Gorsuch R. L., Lushene R., Vagg P. R., Jacobs G. A. (1983) **Manual for the State-Trait Anxiety Inventory** Palo Alto, CA: Consulting Psychologists Press

Starcevic V., Berle D., Brakoulias V., Sammut P., Moses K., Milicevic D., Hannan A. (2011) **Functions of compulsions in obsessive-compulsive disorder** *Australian and New Zealand Journal of Psychiatry* **45**:449–457 https://doi.org/10.3109/00048674.2011.567243

Tasca G. A., Presniak M. D., Demidenko N., Balfour L., Krysanski V., Trinneer A., Bissada H. (2011) **Testing a maintenance model for eating disorders in a sample seeking treatment at a tertiary care center: a structural equation modeling approach** *Comprehensive Psychiatry* **52**:678–687 https://doi.org/10.1016/j.comppsych.2010.12.010

Tolin D. F., Abramowitz J. S., Brigidi B. D., Amir N., Street G. P., Foa E. B. (2001) **Memory and memory confidence in obsessive-compulsive disorder** *Behaviour Research and Therapy* **39**:913–927 https://doi.org/10.1016/S0005-7967(00)00064-4

Vaghi M. M., Luyckx F., Sule A., Fineberg N. A., Robbins T. W., De Martino B. (2017) **Compulsivity Reveals a Novel Dissociation between Action and Confidence** *Neuron* **96**:348–354 https://doi.org/10.1016/j.neuron.2017.09.006

Wise T., Dolan R. J. (2020) **Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample** *Nature Communications* **11**:1–13 https://doi.org/10.1038/s41467-020-17977-w

Wise T., Robinson O. J., Gillan C. M. (2023) **Identifying Transdiagnostic Mechanisms in Mental Health Using Computational Factor Modeling** *Biological Psychiatry* **93**:690–703 https://doi.org/10.1016/j.biopsych.2022.09.034

Zung W. W. (1965) **A self-rating depression scale** *Archives of general psychiatry* **12**:63–70

Author information

Annika Boldt
University College London, London, United Kingdom

For correspondence: a.boldt@ucl.ac.uk

Celine A Fox
Trinity College Dublin, Dublin, Ireland

Claire M Gillan
Trinity College Dublin, Dublin, Ireland

Sam Gilbert
University College London, London, United Kingdom

Editors

Reviewing Editor
Nils Kolling
Stem-cell and Brain Institute (SBRI), U1208 Inserm, Bron Cedex, France

Senior Editor
Michael Frank
Brown University, Providence, United States of America

Reviewer #1 (Public review):

Summary:

Boldt et al test several possible relationships between trandiagnostically-defined compulsivity and cognitive offloading in a large online sample. To do so, they develop a new and useful cognitive task to jointly estimate biases in confidence and reminder-setting. In doing so, they find that over-confidence is related to less utilization of reminder-setting, which partially mediates the negative relationship between compulsivity and lower reminder-setting. The paper thus establishes that, contrary to the over-use of checking behaviors in patients with OCD, greater levels of transdiagnostically-defined compulsivity predicts less deployment of cognitive offloading. The authors offer speculative reasons as to why (perhaps it's perfectionism in less clinically-severe presentations that lowers the cost of expending memory resources), and sets an agenda to understand the divergence in cognitive between clinical and nonclinical samples. Because only a partial mediation had robust evidence, multiple effects may be at play, whereby compulsivity impacts cognitive offloading via overconfidence and also by other causal pathways.

Strengths:

The study develops an easy-to-implement task to jointly measure confidence and replicates several major findings on confidence and cognitive offloading. The study uses a useful measure of cognitive offloading - the tendency to set reminders to augment accuracy in the presence of experimentally manipulated costs. Moreover, the utilizes multiple measures of presumed biases -- overall tendency to set reminders, the empirically estimated indifference point at which people engage reminders, and a bias measure that compares optimal

indifference points to engage reminders relative to the empirically observed indifference points. That the study observes convergenence along all these measures strengthens the inferences made relating compulsivity to the under-use of reminder-setting. Lastly, the study does find evidence for one of several a priori hypotheses and sets a compelling agenda to try to explain why such a finding diverges from an ostensible opposing finding in clinical OCD samples and the over-use of cognitive offloading.

Weaknesses:

Although I think this design and study are very helpful for the field, I felt that a feature of the design might reduce the tasks's sensitivity to measuring dispositional tendencies to engage cognitive offloading. In particular, the design introduces prediction errors, that could induce learning and interfere with natural tendencies to deploy reminder-setting behavior. These PEs comprise whether a given selected strategy will be or not be allowed to be engaged. We know individuals with compulsivity can learn even when instructed not to learn (e.g., Sharp, Dolan and Eldar, 2021, Psychological Medicine), and that more generally, they have trouble with structure knowledge (eg Seow et al; Fradkin et al), and thus might be sensitive to these PEs. Thus, a dispositional tendency to set reminders might be differentially impacted for those with compulsivity after an NPE, where they want to set a reminder, but aren't allowed to. After such an NPE, they may avoid moreso the tendency to set reminders. Those with compulsivity likely have superstitious beliefs about how checking behaviors lead to a resolution of catastrophes, that might in part originate from inferring structure in the presence of noise or from purely irrelevant sources of information for a given decision problem.
It would be good to know if such learning effects exist, if they're modulated by PE (you can imagine PEs are higher if you are more incentivized - e.g., 9 points as opposed to only 3 points - to use reminders, and you are told you cannot use them), and if this learning effect confounds the relationship between compulsivity and reminder-setting.

A more subtle point, I think this study can be more said to be an exploration than a deductive of test of a particular model -> hypothesis -> experiment. Typically, when we test a hypothesis, we contrast it with competing models. Here, the tests were two-sided because multiple models, with mutually exclusive predictions (over-use or under-use of reminders) were tested. Moreover, it's unclear exactly how to make sense of what is called the direct mechanism, which is supported by the partial (as opposed to complete) mediation.

https://doi.org/10.7554/eLife.98114.2.sa2

Reviewer #2 (Public review):

Summary:

Boldt et al., investigated whether previously established relationships between transdiagnostic psychiatric symptom dimensions and confidence distortions would result in downstream influences on the confidence-related behaviour of reminder setting. 600 individuals from the general population completed a battery of psychiatric symptom questionnaires and an online reminder-setting task. In line with previous studies, individuals high in compulsivity (CIT) showed over-confidence in their task performance, whereas individuals high in anxious-depression (AD) tended to be under-confident. Crucially, the over-confidence associated with CIT partially mediated a decreased tendency to use external reminders during task performance, whereas the under-confidence associated with AD did not result in any alteration in external reminder setting. The authors suggest that metacognitive monitoring is impaired in CIT which has a knock-on effect on reminder setting behaviour, but that a direct link also exists between CIT and reduced reminder setting independently of confidence.

Strengths:

The study combines the latest advances in transdiagnostic approaches to psychopathology with a cleverly designed external reminder-setting task. The approach allows for investigation of what some of the downstream consequences associated with impaired metacognition in sub-clinical psychopathology may be.

The experimental design and hypotheses were pre-registered prior to data collection.

The manuscript is well written and rigorous analysis approaches are used throughout.

Weaknesses:

Participants only performed a single task so it remains unclear if the observed effects would generalise to reminder setting in other cognitive domains.

The sample consisted of participants recruited from the general population. Future studies should investigate whether the effects observed extend to individuals with the highest levels of symptoms (including clinical samples).

https://doi.org/10.7554/eLife.98114.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Reviewer #1:***
>
> *(1) You claim transdiagnostic phenotypes are temporally stable -- since they're relatively new constructs, do we know how stable? In what order?*

This is an important question. We have added two recent references to support this claim on page 1 and cite these studies in the references on pages 25 and 28:

"Using factor analysis, temporally stable (see Fox et al., 2023a; Sookud, Martin, Gillan, & Wise, 2024), transdiagnostic phenotypes can be extracted from extensive symptom datasets (Wise, Robinson, & Gillan, 2023)."

Fox, C. A., McDonogh, A., Donegan, K. R., Teckentrup, V., Crossen, R. J., Hanlon, A. K., … Gillan, C. M. (2024). Reliable, rapid, and remote measurement of metacognitive bias. *Scientific Reports*, *14*(1), 14941. https://doi.org/10.1038/s41598-024-64900-0

Sookud, S., Martin, I., Gillan, C., & Wise, T. (2024, September 5). Impaired goal-directed planning in transdiagnostic compulsivity is explained by uncertainty about learned task structure. https://doi.org/10.31234/osf.io/zp6vk

More specifically, Sookud and colleagues found the intraclass correlation coefficient (ICC) for both factors to be high after a 3- or 12 month period (ICC_{AD_3} = 0.87; ICC_{AD_12} = 0.87; ICC_{CIT_3} = 0.81; ICC_{CIT_3}= 0.76; see Tables S41 and S50 in Sookud et al., 2024).

> *(2) On hypotheses of the study:*
>
> *I didn't understand the logic behind the hypothesis relating TDx Compulsivity -> Metacognition > Reminder-setting*
>
> *It seems that (a) Compulsivity relates to overconfidence which should predict less remindersetting*

> *Compulsivity has an impaired link between metacognition and action, breaking the B->C link in the mediation described above in (a). What would this then imply about how Compulsivity is related to reminder-setting?*
>
> *"In the context of our study, a Metacognitive Control Mechanism would be reflected in a disrupted relationship between confidence levels and their tendency to set reminders." What exactly does this predict - a lack of a correlation between confidence and remindersetting, specifically in high-compulsive subjects?*
>
> *Lastly, there could be a direct link between compulsivity and reminder-usage, independent of any metacognitive influence. We refer to this as the Direct Mechanism Why though theoretically would this be the case?*
>
> *"We initially hypothesised to find support for the Metacognitive Control Mechanism and that highly compulsive individuals would offload more".*
>
> *The latter part here, "highly compulsive individuals would offload more" is I think the exact opposite prediction of the Metacognitive control mechanism hypothesis (compulsive individuals offload less). How could you possibly have tried to find support, then, for both?*
>
> *Is the hypothesis that compulsivity positively predicts reminder setting the "direct mechanism" - if so, please clarify that, and if not, it should be added as a distinct mechanism, and additionally, the direct mechanism should be specified.*
>
> *There's more delineation of specific hypotheses (8 with caveats) in Methods.*
>
> *"We furthermore also tested this hypothesis but predicted raw confidence (percentage of circles participants predicted they would remember; H6b and H8b respectively)," What is the reference of "this hypothesis" given that right before this sentence two hypotheses are mentioned? To keep this all organized, it would be good to simply have a table with hypotheses listed clearly.*

We agree with the reviewer that there is room to improve the clarity of how our hypotheses are presented. The confusion likely arises from the fact that, since we first planned and preregistered our study, several new pieces of work have emerged, which might have led us to question some of our initial hypotheses. We have taken great care to present the hypotheses as they were preregistered, while also considering the current state of the literature and organizing them in a logical flow to make them more digestible for the reader. We have clarified this point on page 4:

“Back when we preregistered our hypotheses only a limited number of studies about confidence and transdiagnostic CIT were available. This resulted in us hypothesising to find support for the *Metacognitive Control Mechanism* and that highly compulsive individuals would offload more due to an increased need for checkpoints.”

The biggest improvement we believe comes from our new Table 1, which we have included in the Methods section in response to the reviewer’s suggestion (pp. 21-22):

“We preregistered 8 hypotheses (see Table 1), half of which were sanity checks (H1-H4) aimed to establish whether our task would generally lead to the same patterns as previous studies using a similar task (as reviewed in Gilbert et al., 2023).”

We furthermore foreshadowed more explicitly how we would test the Metacognitive Control Mechanism in the Introduction section on page 4, as requested by the reviewer:

"In the context of our study, a Metacognitive Control Mechanism would be reflected in a disrupted relationship between confidence levels and their tendency to set reminders (i.e., the interaction between the bias to be over- or underconfident and transdiagnostic CIT in a regression model predicting a bias to set reminders)."

To avoid any confusion regarding the term 'direct' in the 'Direct Mechanism', we now explicitly clarify on page 4 that it refers to any non-metacognitive influences. Additionally, we had already emphasized in the Discussion section the need for future studies to specify these influences more directly.

Page 4: "We refer to this as the *Direct Mechanism* and it constitutes any possible influences that affect reminder setting in highly-compulsive CIT participants outside of metacognitive mechanisms, such as perfectionism and the wish to control the task without external aids."

The reviewer was correct in pointing out that, in the Methods section, we incorrectly referred to 'this hypothesis' when we actually meant both of the previously mentioned hypotheses. We have corrected this on page 23:

"We furthermore also tested these hypotheses but predicted raw confidence (percentage of circles participants predicted they would remember; H6b and H8b respectively), as well as extending the main model with the scores from the cognitive ability test (ICAR5) as an additional covariate (H6c and H8c respectively)."

Finally, upon revisiting our Results section, we noticed that we had not made it sufficiently clear that hypothesis H6a was preregistered as non-directional. We have now clarified this on page 9:

"We predicted that the metacognitive bias would correlate negatively with AD (Hypothesis 8a; more anxious-depressed individuals tend to be underconfident). For CIT, we preregistered a non-directional, significant link with metacognitive bias (Hypothesis H6a). We found support for both hypotheses, both for AD, β = -0.22, SE = 0.04, t = -5.00, p < 0.001, as well as CIT, β = 0.15, SE = 0.05, t = 3.30, p = 0.001, controlling for age, gender, and educational attainment (Figure 3; see also Table S1). Note that for CIT this effect was positive, more compulsive individuals tend to be overconfident."

> *(3) You say special circles are red, blue, or pink. Then, in the figure, the colors are cyan, orange, and magenta. These should be homogenized.*

Apologies, this was not clear on our screens. We have corrected this now but used the labels "blue", "orange" and "magenta" as our shade of blue is much darker than cyan:

Page 16: "These circles flashed in a colour (blue, orange, or magenta) when they first appear on screen before fading to yellow."

> *(4) The task is not clearly described with respect to forced choice. From my understanding, "forced choice" was implicitly delivered by a "computer choosing for them". You should indicate in the graphic that this is what forced choice means in the graphic and description more clearly.*

This is an excellent point. On pages 17 and 18 we now include a slightly changed Figure 6, which includes improved table row names and cell shading to indicate the choice people gave. Hopefully this clarifies what "forced choice" means.

> *(5) If I have point (4) right, then a potential issue arises in your design. Namely, if a participant has a bias to use or not use reminders, they will experience more or less prediction errors during their forced choice. This kind of prediction error could introduce*

> *different mood impacts on subsequent performance, altering their accuracy. This will have an asymmetric effect on the different forced phases (ie forced reminders or not). For this reason, I think it would be worthwhile to run a version of the experiment, if feasible, where you simply remove choice prior to revealing the condition. For example, have a block of choices where people can "see how well you do with reminders" -- this removes expectation and PE effects.*
>
> *[See also this point from the weaknesses listed in the public comments:]*
>
> *Although I think this design and study are very helpful for the field, I felt that a feature of the design might reduce the tasks's sensitivity to measuring dispositional tendencies to engage cognitive offloading. In particular, the design introduces prediction errors, that could induce learning and interfere with natural tendencies to deploy reminder-setting behavior. These PEs comprise whether a given selected strategy will be or not be allowed to be engaged. We know individuals with compulsivity can learn even when instructed not to learn (e.g., Sharp, Dolan, and Eldar, 2021, Psychological Medicine), and that more generally, they have trouble with structure knowledge (eg Seow et al; Fradkin et al), and thus might be sensitive to these PEs. Thus, a dispositional tendency to set reminders might be differentially impacted for those with compulsivity after an NPE, where they want to set a reminder, but aren't allowed to. After such an NPE, they may avoid more so the tendency to set reminders. Those with compulsivity likely have superstitious beliefs about how checking behaviors leads to a resolution of catastrophes, which might in part originate from inferring structure in the presence of noise or from purely irrelevant sources of information for a given decision problem.*
>
> *It would be good to know if such learning effects exist if they're modulated by PE (you can imagine PEs are higher if you are more incentivized - e.g., 9 points as opposed to only 3 points - to use reminders, and you are told you cannot use them), and if this learning effect confounds the relationship between compulsivity and reminder-setting.*

We would like to thank the reviewer for providing this interesting perspective on our task. If we understand correctly, the situation most at risk for such effects occurs when participants choose to use a reminder. Not receiving a reminder in the following trial can be seen as a negative prediction error (PE), whereas receiving one would represent the control condition (zero PE). Therefore, we focused on these two conditions in our analysis.

We indeed found that participants had a slightly higher tendency to choose reminders again after trials where they successfully requested them compared to after trials where they were not allowed reminders (difference = 4.4%). This effect was statistically significant, $t(465) = 2.3$, $p = 0.024$. However, it is important to note that other studies from our lab have reported a general, non-specific response 'stickiness,' where participants often simply repeat the same strategy in the next trial (Scarampi & Gilbert, 2020), which could have contributed to this pattern.

When we used CIT to predict this effect in a simple linear regression model, we did not find a significant effect ($\beta = -0.05$, $SE = 0.05$, $t = -1.13$, $p = 0.26$).

To further investigate this and potentially uncover an effect masked by the influence of the points participants could win in a given trial, we re-ran the model using a logistic mixed-effects regression model. This model predicted the upcoming trial's choice (reminder or no reminder) from the presence of a negative prediction error in the current trial (dummy variable), the ztransformed number of points on offer, and the z-transformed CIT score (between-subject covariate), as well as the interaction of CIT and negative PE. In this model, we replicated the previous 'stickiness' effect, with a negative influence of a negative PE on the upcoming choice, $\beta = -0.24$, $SE = 0.07$, $z = -3.44$, $p < 0.001$. In other words, when a negative PE was encountered in the current trial, participants were less likely to choose reminders in the

next trial. Additionally, there was a significant negative influence of points offered on the upcoming choice, β = -0.28, *SE* = 0.03, *z* = -8.82, *p* < 0.001. While this might seem counterintuitive, it could be due to a contrast effect: after being offered high rewards with reminders, participants might be deterred from using the reminder strategy in consecutive trials where lower rewards are likely to be offered, simply due to the bounded reward scale. CIT showed a small negative effect on upcoming reminder choice, β = -0.06, *SE* = 0.04, *z* = -1.69, *p* = 0.09, indicating that participants scoring higher on the CIT factor tended to be less likely to choose reminders, thus replicating one of the central findings of our study. It is unclear why this effect was not statistically significant, but this is likely due to the limited data on which the model was based (see below). Finally, and most importantly, the interaction between the current trial's condition (negative PE or zero PE) and CIT was not significant, contrary to the reviewer's hypothesis, β = 0.04, *SE* = 0.07, *z* = 0.57, *p* = 0.57.

It should also be noted that this exploratory analysis is based on a limited number of data points: on average, participants had 2.5 trials (min = 0; max = 4) with a negative PE and 6.7 trials (min = 0; max = 12) with zero PE. There were more zero PE trials simply because to maximise the number of trials included in this analysis, each participant's 8 choice-only trials were included and on those trials the participant always got what they requested (the trial then ended prematurely). Due to the fact that not all cells in the analysed design were filled, only 466 out of 600 participants could be included in the analysis. This may have caused the fit of the mixed model to be singular.

In summary, given that these results are based on a limited number of data points, some models did not fit without issues, and no evidence was found to support the hypotheses, we suggest not including this exploratory analysis in the manuscript. However, if we have misunderstood the reviewer and should conduct a different analysis, we are happy to reconsider.

Unfortunately, conducting an additional study without the forced-choice element is not feasible, as this would create imbalances in trial numbers for the design. The advantage of the current, condensed task is the result of several careful pilot studies that have optimized the task's psychometric properties.

Scarampi, C., & Gilbert, S. J. (2020). The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task. Memory, 28(5), 677–691. https://doi.org/10.1080/09658211.2020.1764974

> *(6) One can imagine that a process goes on in this task where a person must estimate their own efficacy in each condition. Thus, individuals with more forced-choice experience prior to choosing for themselves might have more informed choice. Presumably, this is handled by your large N and randomization, but could be worth looking into.*

We would like to thank the reviewer for pointing this out, as we had not previously considered this aspect of our task. However, we believe it is not the experience with forced trials per se, but rather the frequency with which participants experience both strategies (reminder vs. no reminder), that could influence their ability to make more informed choices. To address this, we calculated the proportion of reminder trials during the first half of the task (excluding choiceonly trials, where the reminder strategy was not actually experienced). We hypothesized that the absolute distance of this 'informedness' parameter should correlate positively with the absolute reminder bias at the end of the task, with participants who experienced both conditions equally by the midpoint of the task being less biased towards or away from reminders. However, this was not the case, *r* = 0.05, *p* = 0.21.

Given the lengthy and complex nature of our preregistered analysis, we prefer not to include this exploratory analysis in the manuscript.

> *(7) Is the Actual indifference calculated from all choices? I believe so, given they don't know only till after their choice whether it's forced or not, but good to make this clear.*

Indeed, we use all available choice data to calculate the AIP. We now make this clear in two places in the main text:

Page 5: “The ‘actual indifference point’ was the point at which they were actually indifferent, based on all of their decisions.”

Page 6: “Please note that all choices were used to calculate the AIP, as participants only found out whether or not they would use a reminder after the decision was made.”

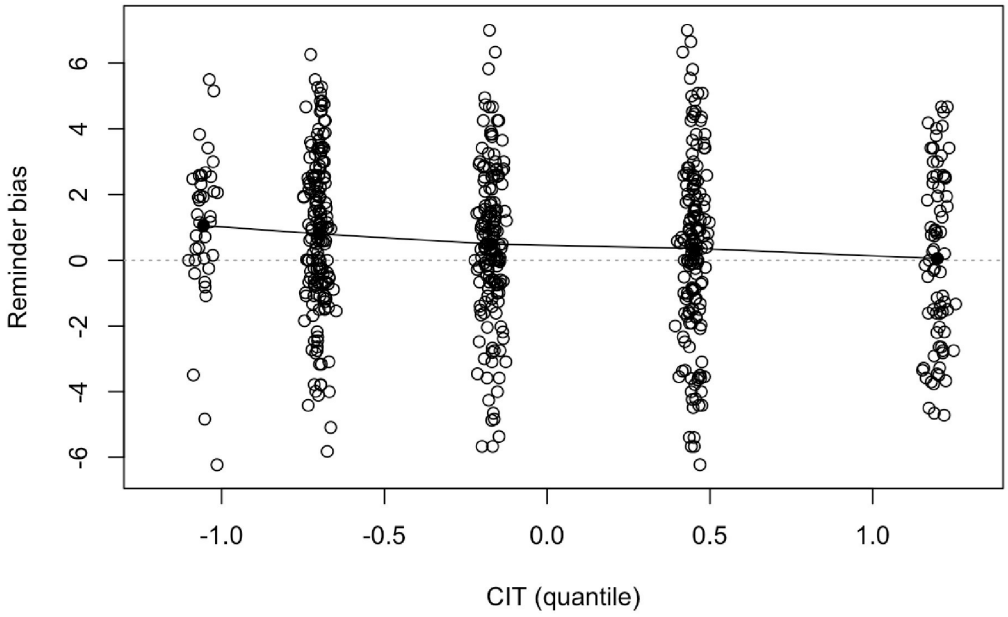
> *(8) Related to 7, I believe this implies that the objective and actual indifference points are not entirely independent, given the latter contains the former.*

Yes, the OIP and AIP were indeed calculated in part from events that happened within the same trials. However, since these events are non-overlapping (e.g., the choice from trial 6 contributes to the AIP but the accuracy measured several seconds later from that trial contributes to the OIP) and since our design dictates whether or not reminders can be used on those trials in question (by randomly assigning them to the forced internal/forced external condition) this could not induce circularity.

> *(9) I thought perfectionism might be a trait that could explain findings and it was nice to see convergence in thinking once I reached the conclusion. Along these lines, I was thinking that perhaps perfectionism has a curvilinear relationship with compulsivity (this is an intuition I'm not sure if it's backed up empirically). If it's really perfectionism, do you see that, at the extreme end of compulsivity, there's more reminder-setting? Ie did you try to model this relationship using a nonlinear function? You might clues simply by visual inspection.*

It is interesting to note that the reviewer reached a similar interpretation of our results. We considered this question during our analysis and conducted an additional exploratory analysis to examine how CIT quantile relates to reminder bias (see Author response image 1). Each circle reflects a participant. As shown, no clear nonlinearities are evident, which challenges this interpretation. We believe that adding this to the already lengthy manuscript may not be necessary, but we are of course happy to reconsider if Reviewer 1 disagrees.

Author response image 1.



> *(10) [From the weaknesses listed in the public comments.] A more subtle point, I think this study can be more said to be an exploration than a deductive test of a particular model -> hypothesis > experiment. Typically, when we test a hypothesis, we contrast it with competing models. Here, the tests were two-sided because multiple models, with mutually exclusive predictions (over-use or under-use of reminders) were tested. Moreover, it's unclear exactly how to make sense of what is called the direct mechanism, which is supported by partial (as opposed to complete) mediation.*

The reviewer's observation is accurate; some aspects of our study did take on a more exploratory nature, despite having preregistered hypotheses. This was partly due to the novelty of our research questions. We appreciate this feedback and will use it to refine our approach in future studies, aiming for more deductive testing.

> ***Reviewer #2:***
>
> *(1) Regarding the lack of relationship between AD and reminder setting, this result is in line with a recent study by Mohr et al (2023:https://osf.io/preprints/psyarxiv/vc7ye) investigating relationships between the same transdiagnostic symptom dimensions, confidence bias and another confidence-related behaviour: information seeking. Despite showing trial-by-trial under-confidence on a perceptual decision task, participants high in AD did not seek information any more than low AD participants. Hence, the under-confidence in AD had no knock-on effect on downstream information-seeking behaviour. I think it is interesting that converging evidence from your study and the Moher et al (2023) study suggest that high AD participants do not use the opportunity to increase their confidence (i.e., through reminder setting or information seeking). This may be because they do not believe that doing so will be effective or because they lack the motivation (i.e., through anhedonia and/or apathy) to do so.*

This is indeed an interesting parallel and we would like to thank the reviewer for pointing out this recently published study, which we unfortunately have missed. We included it in the Discussion section, extending our sub-section on the missing downstream effects of the AD factor, as well as listing it in the references on page 27.

Page 14: "Our findings align with those reported in a recent study by Mohr, Ince, and Benwell (2024). The authors observed that while high-AD participants were underconfident in a perceptual task, this underconfidence did not lead to increased information-seeking behaviour. Future research should explore whether this is due to their pessimism regarding the effectiveness of confidence-modulated strategies (i.e., setting reminders or seeking information) or whether it stems from apathy. Another possibility is that the relevant downstream effects of anxiety were not measured in our study and instead may lie in reminder-checking behaviours."

Mohr, G., Ince, R.A.A. & Benwell, C.S.Y. Information search under uncertainty across transdiagnostic psychopathology and healthy ageing. Transl Psychiatry 14, 353 (2024). https://doi.org/10.1038/s41398-024-03065-w

> *(2) Fox et al 2023 are cited twice at the same point in the second paragraph of the intro. Not sure if this is a typo or if these are two separate studies?*

Those are indeed two different studies and should have been formatted as such. We have corrected this mistake in the following places and furthermore also corrected one of the references as the study has recently been published:

P. 2 (top): "Previous research links transdiagnostic compulsivity to impairments in metacognition, defined as thinking about one's own thoughts, encompassing a broad spectrum of self-reflective signals, such as feelings of confidence (e.g., Rouault, Seow, Gillan & Fleming, 2018; Seow & Gillan, 2020; Benwell, Mohr, Wallberg, Kouadio, & Ince, 2022; Fox et al., 2023a;

Fox et al., 2023b; Hoven, Luigjes, Denys, Rouault, van Holst, 2023a)."

P. 2 (bottom): "More specifically, individuals characterized by transdiagnostic compulsivity have been consistently found to exhibit overconfidence (Rouault, Seow, Gillan & Fleming, 2018; Seow & Gillan, 2020; Benwell, Mohr, Wallberg, Kouadio, & Ince, 2022; Fox et al., 2023a; Fox et al., 2023b; Hoven et al., 2023a)."

P. 4: "Prior evidence exists for overconfidence in compulsivity (Rouault et al., 2018; Seow & Gillan, 2020; Benwell et al., 2022; Fox et al., 2023a; Fox et al., 2023b; Hoven et al., 2023a), which would therefore result in fewer reminders."

P. 23: "Though we did not preregister a direction for this effect, in the light of recent findings it has now become clear that compulsivity would most likely be linked to overconfidence (Rouault et al., 2018; Seow & Gillan, 2020; Benwell et al., 2022; Fox et al., 2023a; Fox et al., 2023b; Hoven et al., 2023a)."

P. 24: "Fox, C. A., Lee, C. T., Hanlon, A. K., Seow, T. X. F., Lynch, K., Harty, S., … Gillan, C. M. (2023a). An observational treatment study of metacognition in anxious-depression. ELife, 12, 1–17. https://doi.org/10.7554/eLife.87193"

P. 24: "Fox, C. A., McDonogh, A., Donegan, K. R., Teckentrup, V., Crossen, R. J., Hanlon, A. K., … Gillan, C. M. (2024). Reliable, rapid, and remote measurement of metacognitive bias. Scientific Reports, 14(1), 14941. https://doi.org/10.1038/s41598-024-64900-0"

> *(3) Typo in the Figure 1 caption: "The preregistered exclusion criteria for the for the accuracies with....".*

Thank you so much for pointing this out. We haved changed the sentence in the caption of Figure 1 to read “The preregistered exclusion criteria for the accuracies with or without reminder are indicated as horizontal dotted lines (10% and 70% respectively).”

> *Typo in the Figure 5 caption: "Standardised regression coefficients are given for each pat".*

Thank you so much for pointing this out to us, we have corrected the typo and the sentence in the caption of Figure 5 now reads “Standardised regression coefficients are given for each path.”

> *[From the weaknesses listed in the public comments.] Participants only performed a single task so it remains unclear if the observed effects would generalise to reminder-setting in other cognitive domains.*

We appreciate the reviewer’s concern regarding the use of a single cognitive task in our study, which is indeed a common limitation in many cognitive neuroscience studies. The cognitive factors underlying offloading decisions are still under active debate. Notably, a previous study found that intention fulfilment in an earlier version of our task correlates with real-world behaviour, lending validity to our paradigm by linking it to realistic outcomes (Gilbert, 2015). Additionally, recent unpublished work (Grinschgl, 2024) has shown a correlation between offloading across two lab tasks, though a null effect was reported in another study with a smaller sample size by the same team (Meyerhoff et al., 2021), likely due to insufficient power. In summary, we agree that future research should replicate these findings with alternative tasks to enhance robustness.

Gilbert, S. J. (2015). Strategic offloading of delayed intentions into the external environment. Quarterly Journal of Experimental Psychology, 68(5), 971–992. https://doi.org/10.1080/17470218.2014.972963

Grinschgl, S. (2024). Cognitive Offloading in the lab and in daily life. 2nd Cognitive Offloading Meeting. [Talk]

Meyerhoff, H. S., Grinschgl, S., Papenmeier, F., & Gilbert, S. J. (2021). Individual differences in cognitive offloading: a comparison of intention offloading, pattern copy, and short-term memory capacity. Cognitive Research: Principles and Implications, 6(1), 34. https://doi.org/10.1186/s41235-021-00298-x

> *(6) [From the weaknesses listed in the public comments.] The sample consisted of participants recruited from the general population. Future studies should investigate whether the effects observed extend to individuals with the highest levels of symptoms (including clinical samples).*

We agree that transdiagnostic research should ideally include clinical samples to determine, for instance, whether the subclinical variation commonly studied in transdiagnostic work differs qualitatively from clinical presentations. However, this approach poses challenges, as transdiagnostic studies typically require large sample sizes, and recruiting clinical participants can be more difficult. With advancements in online sampling platforms, such as Prolific, achieving better availability and targeting may make this more feasible in the future. We intend to monitor these developments closely and contribute to such studies whenever possible.

https://doi.org/10.7554/eLife.98114.2.sa0