eLife

Reviewed Preprint
v1 • September 2, 2026
Not revised

For correspondence:
morio.sinri@gmail.com
cortese.aurelio@gmail.com

Competing interests: No competing interests declared

Funding: See page 16

Reviewing editor: Nathan Faivre, Centre National de la Recherche Scientifique, France

Mapping abstraction and metacognition onto distinct transdiagnostic symptom profiles

Taiki Oka[1,2,3,4], Yoshihiko Kunisato[5], Kotaro Koizumi[1,6], Misa Murakami[1], Hugo Six[1], Jessica E Taylor[1,7], Aurelio Cortese[1,8,9]

[1]The Department of Decoded Neurofeedback, Computational Neuroscience Laboratories, Advanced Telecommunications Research Institute International (ATR), Kyoto, Japan • [2]Developmental Computational Psychiatry, University of Tübingen, Tübingen, Germany • [3]German Centre for Mental Health (DZPG), Partner Site Tübingen, Tübingen, Germany • [4]The Department of Clinical Psychology, Graduate School of Human Sciences, Osaka University, OsakaOsaka, Japan • [5]Department of Psychology, School of Human Sciences, Senshu University, Kawasaki, Japan • [6]Graduate School of Science and Technology, Division of Information Science, Nara Institute of Science and Technology, Nara, Japan • [7]Division of Behavioural Medicine and Neurosciences, Department of Psychiatry & Behavioural Sciences, School of Medicine, Duke University, Durham, United States • [8]Department of Brain Science and Engineering, Sungkyunkwan University, Suwon, Republic of Korea • [9]Centre for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Republic of Korea

eLife Assessment

This **valuable** study examines how abstraction and metacognition relate to transdiagnostic dimensions of psychopathology. It combines a reward-learning task, computational modelling and a large questionnaire. The evidence linking reduced metacognitive sensitivity to a compulsive hypersensitivity dimension is **solid**, but the evidence for the abstraction findings is currently **incomplete**. This work will be of interest to researchers studying metacognition and computational approaches to psychiatry.

https://doi.org/10.7554/eLife.112401.1.sa4

Abstract

Transdiagnostic psychiatric research on reward-guided learning has largely focused on simple associative processes, leaving it unclear whether or how higher-level processes are disrupted. Here, we studied how abstraction, the ability to extract relevant features from complex information, and metacognition, the ability to monitor and evaluate one's own mental processes, map onto specific transdiagnostic dimensions. Using an online sample (N = 249), we examined associations between these processes and three cross-culturally robust transdiagnostic dimensions derived from a large existing dataset (N = 19,505): Compulsive hypersensitivity, Social withdrawal, and Addictive behaviours. Computational modelling of an abstract representation learning task with confidence judgments revealed that Compulsive hypersensitivity was negatively associated with both abstraction ability (p_{boot} = 0.003) and metacognitive sensitivity (p_{boot} = 0.005), while Social withdrawal was positively associated with metacognitive sensitivity alone (p_{boot} = 0.002). Moreover, transdiagnostic dimensions revealed more coherent associations with higher-order cognition than symptom-level analyses, highlighting the added value of examining psychopathology at the factor rather than the symptom level. These findings portray a hierarchical view of cognitive dysfunctions in psychopathology and point to representational and metacognitive processes as potential targets for transdiagnostic intervention.

Introduction

Impairments in reward-guided learning are widely observed across psychiatric disorders and are considered a hallmark of psychopathology (Maia and Frank, 2011). While a growing number of studies have adopted transdiagnostic approaches to investigate shared mechanisms across disorders (Fried, 2022; Kist et al., 2023), this body of work has largely focused on relatively simple learning processes (Lee et al., 2025; Wise et al., 2023). Whether these findings extend to the broader cognitive architecture that supports learning in more complex scenarios remains an open question.

Here, we investigate two such candidate mechanisms — abstraction and metacognition — and examine how each is associated with transdiagnostic symptom dimensions. Abstraction is defined as the process of mentally extracting and organising, from complex information, a small set of features relevant to the task or context at hand (De Martino and Cortese, 2022; Ho et al., 2019), and is critical for generalising behaviour across diverse scenarios. Metacognition, in turn, monitors and evaluates these abstract representations (De Martino and Cortese, 2022; Lee and Hare, 2023). Both processes have been independently implicated in psychopathology. Impairments in metacognition have been reported across multiple transdiagnostic studies (Katyal et al., 2025; Seow et al., 2025a; T. Seow et al., 2021), while abstraction deficits have been documented in specific psychiatric conditions (Gjelsvik et al., 2018; Kobayashi et al., 2014). However, no previous work has examined them concomitantly, nor has it examined how they might be associated with transdiagnostic dimensions. This gap exists partly because abstraction is an inherently internal and latent process that previous studies have assessed using self-reported or behavioural measures alone, which cannot directly capture latent computational mechanisms (Eling et al., 2008). Additionally, although metacognition has been studied from a transdiagnostic perspective (Hoven et al., 2023), abstraction and metacognition have, to the best of our knowledge, not been measured and characterised within the same transdiagnostic framework and the same experiment.

Addressing this question requires both a principled computational approach to quantifying abstract processes and a transdiagnostic framework (Gillan and Seow, 2020; Wise et al., 2023). In this study, we defined abstraction as an algorithmic strategy in an associative learning task in which participants learned to identify which subset of stimulus features predicted reward. The task effectively required participants to extract a low-dimensional task-relevant structure from higher-dimensional information. We then used a computational model to derive a continuous abstraction metric at the individual level, alongside measures of metacognition based on subjective confidence ratings. These were examined across participants in relation to transdiagnostic psychiatric symptom dimension scores obtained by decomposing co-occurring psychiatric symptoms across a large online sample into a small set of latent factors, each capturing shared variance across diagnostic boundaries (Wise et al., 2023). These dimensions were derived by applying factor weights computed from an existing large-scale study (N = 19,505) to psychiatric questionnaire scores of our sample (N = 249). This computational factor modelling approach allowed us to ask how transdiagnostic symptom dimensions relate to abstraction and metacognition (Figure 1).

Results

Behavioural validation of an abstract representation learning task

We recruited participants from the general population worldwide (N = 512) via an online experimental platform (Prolific, *https://www.prolific.co/*) and analysed 249 participants after applying pre-registered exclusion criteria (e.g., failing attention checks, performing the task at chance level; see Methods for further details, and Supplementary Methods for demographics and discussion on sample characteristics). Participants completed self-report questionnaires (consisting of multiple symptom rating scales--8 questionnaires, 176 items), including those related to several types of addictions and to internalising disorders such as social anxiety, attention deficit

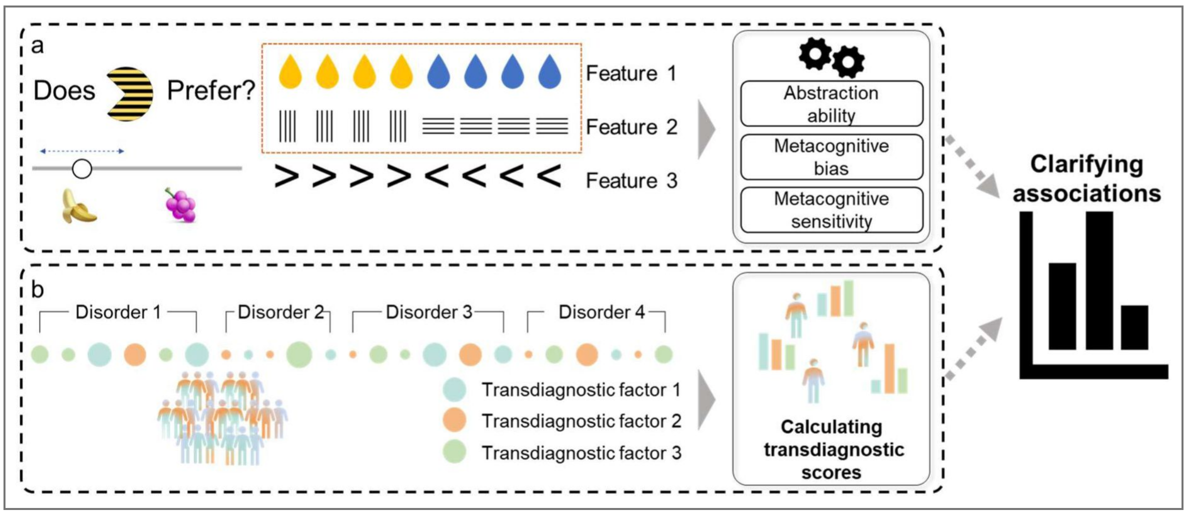


Figure 1. **Computational factor modelling framework and behavioural task.**

Computational factor modelling approach: **a.** Abstract representation learning task: Participants learned the fruit preferences of Pac-Man-like characters, with the features that predict preference changing block-by-block. Three types of feature combinations could predict preference: colour–stripe orientation, colour–mouth direction, and stripe orientation–mouth direction (see the example in the top-right box). The cursor on the slider allows participants to report both their choice and their confidence simultaneously. Latent behavioural scores are estimated through computational modelling. **b.** Transdiagnostic psychiatric dimension scores are extracted across diagnostic boundaries constructed by factor analysis. Then, extracted latent behavioural variables can be linked to each transdiagnostic psychiatric dimension. This data-driven approach enables an integrated understanding of how computational mechanisms underlying key behaviour attributes map onto transdiagnostic symptom dimensions.

hyperactivity disorder, etc., see Methods for further details) and a behavioural task. In the behavioural experiment, participants had to learn the preferred fruit of a Pac-Man-like character (Figure 1a). These characters comprised three features: colour (blue/yellow), stripe orientation (vertical/horizontal), and mouth direction (right/left). Fruit preference was determined by only two of these features at a time, with preferences randomly changing in each new block (see Supplementary Figure 2 for the details). Participants indicated their choice (selecting the character's fruit preference) as well as their confidence (e.g., more lateral position = higher confidence) using a cursor (see Figure 1a; reaction time and confidence distribution are shown in Supplementary Figures 2 and 3). The cursor's randomised starting position only marginally affected the final response (see Supplementary Figure 4).

We confirmed learning curves within blocks (Figure 2a) and across blocks (Figure 2b), observing a positive relationship between learning speed and time, although the latter effect did not reach statistical significance ($\beta = 0.0058$, $p = 0.13$; see Supplementary materials for model details). Learning speed, however, correlated with overall confidence ($\beta = 0.026$, $p = 0.008$; see Figure 2c). These results are consistent with our prior findings (Cortese et al., 2021), indicating that this online experiment is valid relative to our original lab-based experiment.

Extraction and validation of transdiagnostic psychiatric dimensions

To extract transdiagnostic dimensional scores, we transformed initial scores from a battery of psychiatric questionnaires (see Methods) using weights derived from a previous large-scale study (from the Japanese population, N = 19,505) (Oka et al., 2025), in line with previous methods (Seow et al., 2025a; Tricia Seow et al., 2021). We favoured this approach because our sample size (249) was limited relative to the total number of questionnaire items (176), resulting in an insufficient sample-to-item ratio for reliable de novo factor analysis (Osborne and Costello, 2004). The correlations between the questionnaires are shown in Supplementary Figure 5.

Nevertheless, we still sought to assess the validity of using the large-sample factor weights. To this end, we conducted a de novo exploratory factor analysis on the newly collected sample, identified three transdiagnostic factors, and compared the results with the original weights from the previous study. All factors showed significant correlations with moderate effect sizes ($r > 0.4$), indicating replication of the same factor structure and its cross-cultural generalisation (Factor 1: Spearman's $r = 0.63$; Factor 2: Spearman's $r = 0.44$; Factor 3: Spearman's $r = 0.46$, see Figure 3). To maintain consistency with previous work, we refer to these dimensions as follows (Oka et al., 2025). *Compulsive hypersensitivity* is defined primarily by obsessive-compulsive disorder and attention-deficit/hyperactivity disorder symptomatology, reflecting a transdiagnostic tendency toward compulsive and impulsive dyscontrol. *Social withdrawal* is characterised predominantly by social anxiety and psychological distress, capturing a disposition toward avoidance of social interaction. *Addictive behaviours* is defined largely by gaming disorder, alcohol-related problems, and nicotine dependence, reflecting a shared propensity toward behavioural and substance-related addiction.

Abstraction is associated with the Compulsive hypersensitivity dimension, extending beyond findings for individual symptoms

We first quantified individual differences in abstraction by comparing two reinforcement learning models that differed in their feature-level representations (see Methods and the original paper (Cortese et al., 2021)). In short, the "Abstract RL" model assumes that participants rely on a reduced, lower-dimensional representation of task features, whereas the "Feature RL" model assumes that participants learn the full feature combination (Figure 4a, b). We used Hierarchical Bayesian estimation (Piray et al., 2019) to perform concurrent parameter fitting (learning rate α and inverse temperature β) and model comparison across the two RL strategies in each block. The participant count for the best-fitting model in each block is shown in Supplementary Figure 6. Both parameters showed adequate recovery (α: Spearman's $r = 0.40$, β:

Figure 2. **Behaviour validation of the abstract representation learning task.**

a. Trial-by-trial ratio of correct responses served as an index of within-block learning. Each dot represents the mean across participants, error bars denote the standard error of the mean (SEM), and the shaded area indicates the 95% confidence interval. Ratio-correct values were computed for each trial between completed blocks and then averaged across participants. **b.** Relationship between learning speed and time across participants. Learning speed was defined as the inverse of the maximum-normalised number of trials required to complete a block. Thin grey lines show individual participants' data, and the bold black line represents the group-average fit. Circles represent population-level mean values, and error bars represent the SEM. **c.** Averaged confidence ratings were positively associated with learning speed across participants. Each point corresponds to a single participant, and the solid line represents the regression fit. ** p < 0.01.

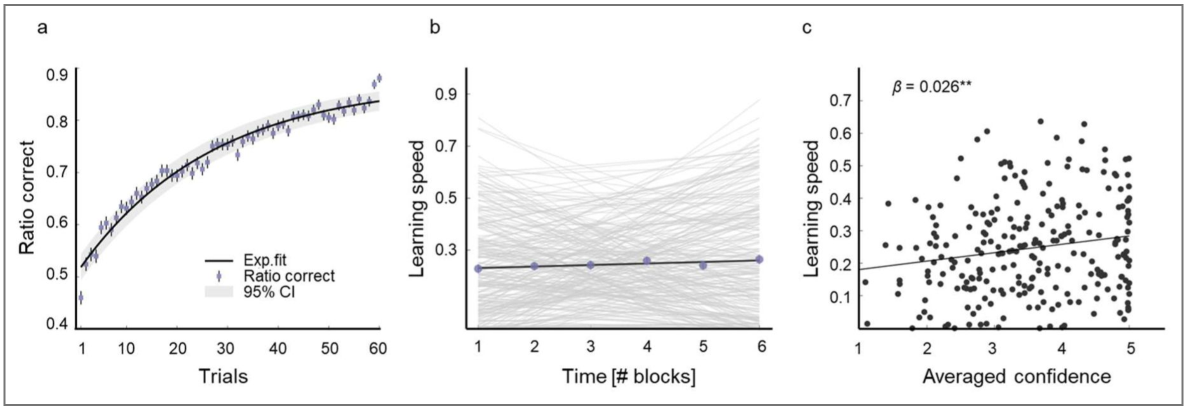
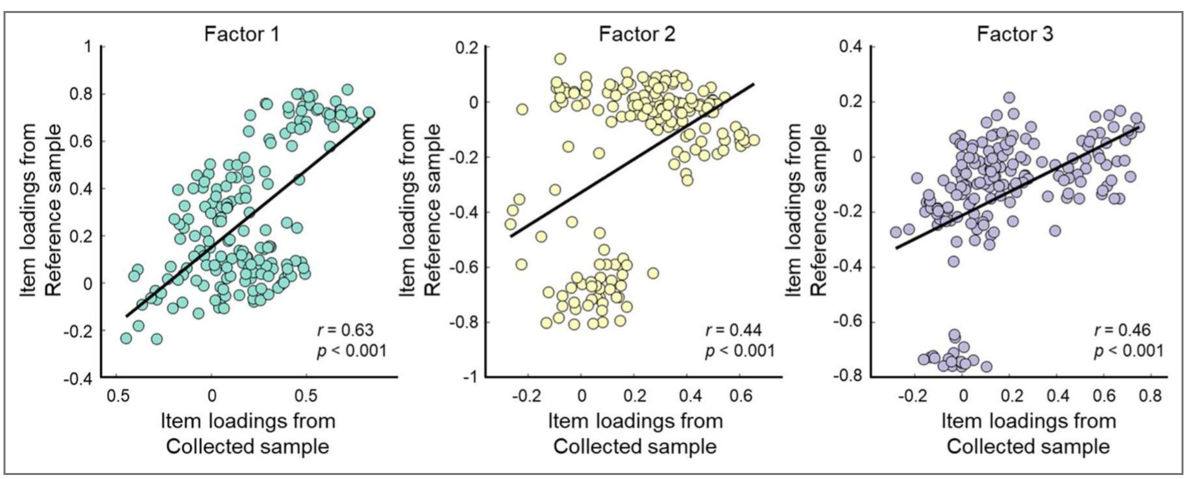


Figure 3. **Correlations between item loadings for the three-factor structure from de-novo factor analysis between a collected sample in this study (N = 249) and a larger reference dataset (N = 19,505).**

Scatter plot of weights of each transdiagnostic factor between the newly collected global sample (from this study, x-axis) and the reference sample (from a large online survey in Japan(Oka et al., 2025), y-axis). Each circle indicates a single questionnaire item. Correlation values are as follows: FA1 r = 0.63, 95% CI [0.53, 0.71], p < 0.001. FA2 r = 0.44, 95% CI [0.31, 0.55], p < 0.001, FA3: r = 0.46, 95% CI [0.34, 0.57], p < 0.001.



Spearman's r = 0.72, both p < .001; Supplementary Figure 7, see also Methods for details on parameter fitting and recovery procedures). We quantified our abstraction ability metric as the mean posterior responsibility of the Abstract RL model across blocks within each participant, as computed through HBI. This continuous metric reflects the degree to which each participant's behaviour was better explained by abstract rather than feature-level representations across the task. Although our preregistration specified the proportion of blocks in which the Abstract RL model provided a better fit as the abstraction metric, we deviated from this approach because the proportion measure takes only seven discrete values (6 blocks per participant), resulting in limited variance and reduced sensitivity to individual differences. The continuous responsibility metric retains the same theoretical interpretation while providing greater resolution (see the histogram of abstract RL use probability in Figure 4c).

We then examined the association between the abstraction metric estimated from computational modelling and psychopathological symptoms, controlling for sex, age, educational level, and processing speed [see Methods for details]. Symptom-level and transdiagnostic dimension-level models were tested independently, with the expectation that transdiagnostic dimensions would reveal more specific associations than individual symptom scores, reflecting the added value of a transdiagnostic approach over symptom-by-symptom analysis. This two-level approach was applied consistently across all subsequent regression analyses. To test these relationships, we employed bootstrap multiple linear regression (5,000 iterations with replacement), reporting the mean beta estimate, 95% confidence interval (2.5th and 97.5th percentiles of the bootstrap distribution), and a two-tailed bootstrap p-value. Although our pre-registered analysis plan specified standard multiple linear regression, we deviated from this protocol in favour of bootstrap regression to provide more robust inference without distributional assumptions (pre-registered regression analyses are reported in Supplementary Figures 8-11 and Supplementary Tables 1-4, which were almost consistent with the bootstrap results below. In particular, results using the preregistered abstraction metric based on proportion of best-fit models across blocks are reported in Supplementary Figure 9 and Supplementary Table 2 for transparency.) All statistical values for all analyses are shown in Supplementary Tables 5-7. Bootstrapping results showed negative relationships between abstraction ability and ADHD, psychological distress, and obsessive-compulsive symptoms ($\beta = -0.017$ [−0.030 −0.004], $p_{boot} = 0.032$; $\beta = -0.020$ [−0.034 −0.006], $p_{boot} = 0.024$; $\beta = -0.026$ [−0.040 −0.012], $p_{boot} = 0.003$, respectively; see Supplementary Figure 12c). Note that these ps_{boot} are adjusted for multiple comparisons (corrected for false discovery rate (FDR)). Importantly, we found a significantly negative relationship with Compulsive hypersensitivity within the transdiagnostic dimensions ($\beta_{boot} = -0.037$ [−0.061 −0.012], p_{boot} = 0.003, see Figure 5a). Notably, not all symptoms with high loadings on the Compulsive hypersensitivity factor showed significant associations at the symptom level, suggesting that the transdiagnostic dimension may capture shared variance extending beyond individual symptom scores.

Transdiagnostic dimensions are not associated with metacognitive bias

We next conducted a pre-registered analysis using the same method, except that we included participants' response bias as a covariate (Sarna et al., 2026). We first tested the association between the averaged confidence and psychopathology. If confidence was biased towards high (or low) ratings, it should reflect a metacognitive bias, defined as the degree to which the self-rating diverges directionally from reality, resulting in under- or over-confidence (Seow et al., 2025a). Our results indicated no statistically significant relationships between averaged confidence (metacognitive bias) and psychopathology after correcting for multiple comparisons (note that several associations between metacognitive bias and symptom-level scores were significant before correction, Figure 5b, Supplementary Figure 12b).

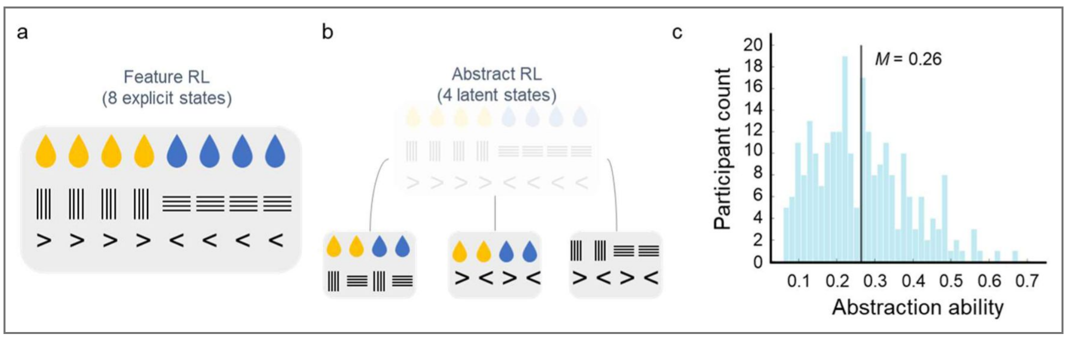


Figure 4. **Behaviour modelling framework.**

a. Feature RL model. The model has $2^3 = 8$ states, each corresponding to a combination of all three features. **b.** Abstract RL model. The model is designed with $2^2 = 4$ states, where each state is defined by the combination of two features, with three possible ways to achieve this 4-state configuration. **c.** Histogram of probability of using the abstract strategy across participants. The vertical line shows the population mean (M = 0.26).

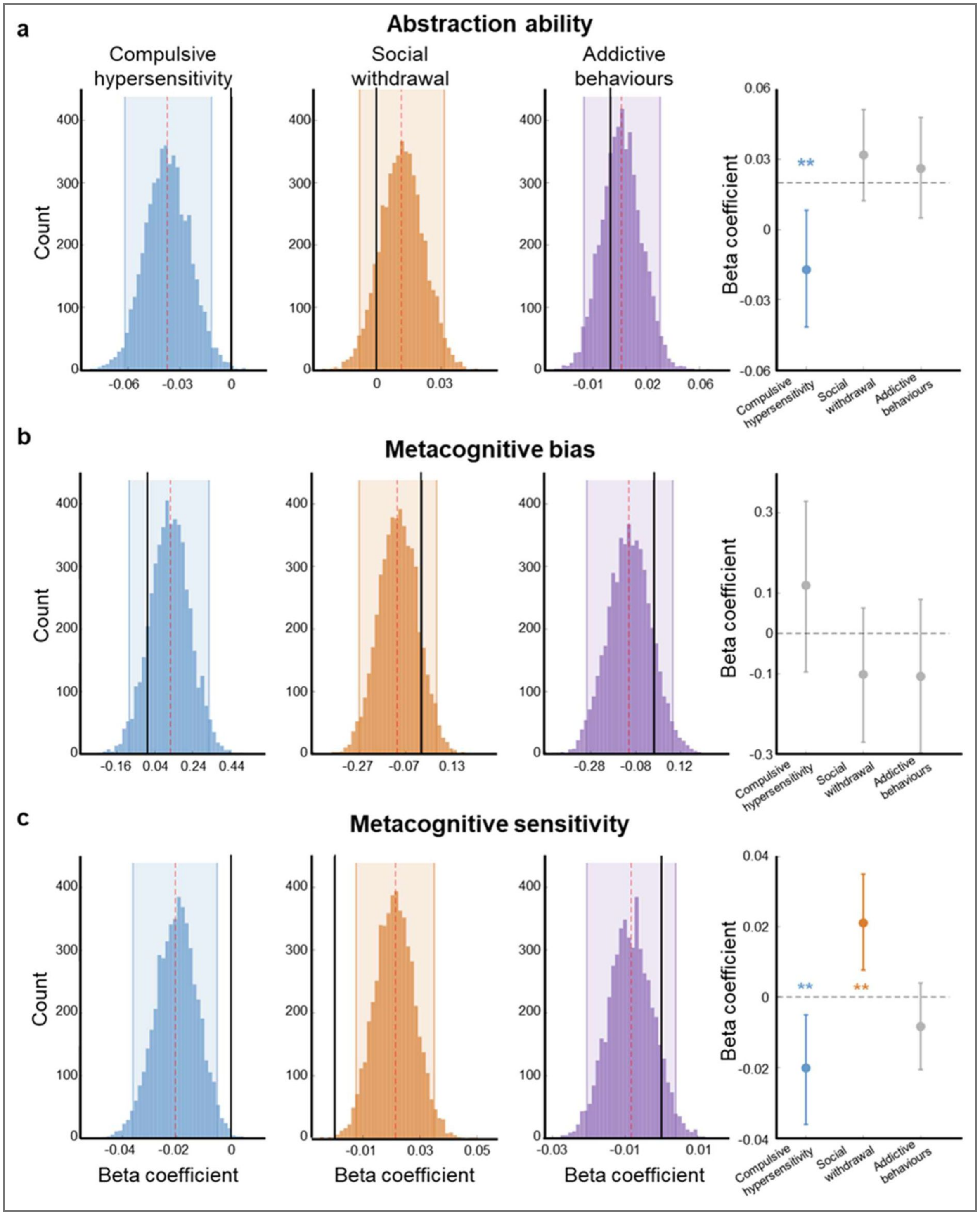


Figure 5. **Associations between task-related behaviour measures and transdiagnostic symptom dimensions.**

Regression coefficients between each metric and dimension scores based on bootstrap multiple linear regression estimation with 5000 iterations. **a.** Associations between the abstraction parameter and each dimension score. **b.** Associations between the averaged confidence and each dimension score**. c.** Associations between the metacognitive sensitivity and each dimension score. The black vertical line in each histogram represents the hypothetical null effect (coefficient = 0). The red vertical dashed line in each histogram represents the average beta coefficient. ** $p_{boot} < 0.01$

Transdiagnostic dimensions are associated with metacognitive sensitivity

Finally, we tested the relationship between psychopathology and metacognitive sensitivity. Simply speaking, this measure captures how well confidence ratings predict choice accuracy, and we measured it as the area under the type 2 ROC curve (AUROC2) (Embon et al., 2023; Jia et al., 2020; Pinkham et al., 2018). Results showed negative relationships between metacognitive sensitivity and obsessive-compulsive (β = −0.017 [−0.025 −0.008], p_{boot} = 0.004) and gaming disorder symptoms (β = −0.009 [−0.020 −0.003], p_{boot} = 0.032). Note that these p_{boot} are adjusted for multiple comparisons (See Supplementary Figure 12c). Transdiagnostically, we found a significantly negative relationship with Compulsive hypersensitivity (β_{boot} = −0.020 [−0.036 −0.005], p_{boot} = 0.005) and a positive relationship with Social withdrawal dimensions (β_{boot} = 0.021 [0.008 0.035], p_{boot} = 0.002, see Figure 5c). Models based on single multiple linear regression that included the interaction between psychopathology and metacognitive sensitivity did not provide a better fit to the data, as indicated by a likelihood-ratio test (all *ps* > 0.05, Supplementary Table 8).

Discussion

In this preregistered study, we examined the relationship between abstraction, metacognition, and transdiagnostic psychopathology in an online sample. First, we found that transdiagnostic symptom dimensions appear broadly consistent across samples (even across countries). Our computationally derived metric of abstraction ability was associated with Compulsive hypersensitivity, beyond what was observed across symptoms that collectively define this dimension. Additionally, task-related metacognitive sensitivity, extracted from confidence judgments and choice data, was distinctly associated with transdiagnostic dimensions. The Social withdrawal symptom dimension was associated with heightened metacognitive sensitivity, while the Compulsive hypersensitivity symptom dimension was associated with reduced metacognitive sensitivity. Taken together, our results indicate multilayered effects of psychopathology across both abstract task representation and metacognitive evaluation.

The selective association between reduced abstraction and the Compulsive hypersensitivity dimension provides important mechanistic insight at the transdiagnostic level. Difficulty compressing task-relevant information into lower-dimensional representations may reflect an over-reliance on detailed, feature-level encoding, consistent with the inflexible, stimulus-bound behavioural control characteristic of compulsive psychopathology (Robbins et al., 2024). Our findings are also consistent with previous converging evidence linking compulsive tendencies to less efficient computation and a preference for familiar over goal-directed action (Banca et al., 2024; Doyle et al., 2026). Alternatively, this need not necessarily imply an impaired capacity to form abstractions *per se*, but may instead reflect a tendency to perseverate within existing representational structures even when more flexible updating would be adaptive (Tricia Seow et al., 2021). However, existing evidence in compulsive psychopathology is inconsistent (Benzina et al., 2021), making it difficult to draw firm conclusions in favour of a perseveration account. Future work directly contrasting the formation of new abstractions with the flexible updating of existing ones will be needed to arbitrate between these accounts. That these associations were not observed for Social withdrawal or Addictive behaviours suggests that impaired abstraction is a dimension-specific marker of the compulsive spectrum rather than a transdiagnostic feature of psychopathology more broadly.

Our findings also show that, whereas Social withdrawal was positively associated with metacognitive sensitivity, Compulsive hypersensitivity was negatively associated. Reduced metacognitive sensitivity in the transdiagnostic compulsive dimension is consistent with previous clinical and transdiagnostic studies (Seow and Gillan, 2020; Vaghi et al., 2017). Interestingly, the positive association between Social withdrawal and metacognitive sensitivity may indicate heightened monitoring of internal performance signals, analogous to the increased insight into performance fluctuations reported for a depression-anxious dimension (Rouault et al., 2018), rather than reflecting impaired metacognitive processes.

The advantage of examining psychopathology at the transdiagnostic rather than symptom level was particularly evident for metacognitive sensitivity. Symptom-level associations were inconsistent in both magnitude and direction across symptoms loading on the same factor, yet coherent and significant associations emerged at the dimension level, suggesting that transdiagnostic factors capture cognitive profiles that individual symptom scores fail to resolve (Wise et al., 2023). Note, however, that our task involved sequential learning rather than independent trials, which may introduce a systematic shift in the measured metacognitive sensitivity. However, since this effect should be consistent across participants, it is unlikely to bias associations with psychiatric dimensions, although it may attenuate their magnitude. Importantly, significant effects were present even with strict screening of inattentive participants and controlling for response bias, suggesting the findings are robust to general confounding variables (Sarna et al., 2026)

The convergence of diminished abstraction use and reduced metacognitive sensitivity within the Compulsive hypersensitivity dimension points to dysfunction at two distinct but complementary levels of cognitive organisation, each of which represents a potential therapeutic target. At the representational level, the abstraction deficit identified here suggests that recalibrating how task-relevant structure is extracted and encoded may be a promising intervention target. Decoded neurofeedback (DecNef) has been proposed as one useful approach to this end in compulsive disorders (Zhang et al., 2026), operating by reinforcing distributed multivoxel neural representations rather than coarse regional signals (Taschereau-Dumouchel et al., 2022, 2021). Critically, neurofeedback has been shown to directly enhance abstraction by reinforcing neural representations of task-relevant features and increasing their utilisation for decision making (Cortese et al., 2021; Kahnt, 2022), providing a mechanistic rationale for applying DecNef to counteract the feature-level over-reliance characteristic of compulsive psychopathology. At the metacognitive level, interventions targeting maladaptive biases about one's own thoughts and performance have shown promise for compulsive psychopathology (Seow et al., 2025b). Together, these approaches may be particularly suited to transdiagnostic dimensions in which both representational and metacognitive processes are disrupted (Seow et al., 2025b). Moreover, our computational approach did not incorporate models that integrate abstract and metacognitive representations; future studies should examine whether interactive models (Cortese, 2022) better capture the mechanisms underlying metacognitive dysfunction in psychopathology.

There are some limitations to consider. First, online recruitment may have introduced sampling bias, as participants were limited to those with internet access or an interest in online research. Second, the analytic sample comprised approximately 50% of those initially recruited after applying pre-registered exclusion criteria. However, note that these exclusions followed pre-registered criteria for data quality and should have reduced measurement error (Brühlmann et al., 2020). Third, choice and confidence were reported via the same slider, potentially confounding metacognitive sensitivity, although the absence of a strong initial position effect suggests this is unlikely to account for the observed associations. Fourth, transdiagnostic scores were derived using first-order factor analysis rather than alternative approaches such as bi-factor modelling (Caspi and Moffitt, 2018); nonetheless, first-order factors have been shown to robustly capture cognitive correlates relative to alternative solutions (Fox et al., 2025).

In summary, in this study, we found that a Compulsive hypersensitivity transdiagnostic dimension was characterised by convergent impairments in both abstract task representation and metacognitive evaluation, whereas Social withdrawal was selectively associated with heightened metacognitive sensitivity. These findings support a hierarchical view of cognition in which transdiagnostic psychopathology is also expressed at the level of how task structure is represented and monitored. The dimension-specificity of these associations suggests that representational and metacognitive impairments constitute markers of compulsive psychopathology in particular, with implications for interventions targeting both levels of cognitive organisation.

Methods

Transparency, openness, and ethics

The current study has been pre-registered (Open Science Framework repository (OSF), DOI: *https://osf.io/qa4bk/*). See the supplementary methods for a summary of deviations from the original preregistration. The study has been approved by the Ethics Committee of the Advanced Telecommunication Research Institute (No. 773). Participants provided their informed consent online by clicking "Yes" on the statement "I consent to take part in this study" after reading the plain language statement and consent form.

Sample size

We used G*power 3.1.9.7 (Franz Faul, Kiel University, Germany) for the a priori power calculator. Because correlations between individual differences in transdiagnostic variables and such RL parameters (r = 0.1-0.2) in a similar previous study (Dubois and Hauser, 2022) and the correlation between the transdiagnostic factors and the degree of abstraction RL parameter in our pilot sample (r = ~0.15, N = 15) were low, we assume a one-tailed alpha-level of 0.05, a power of 90%, and a correlation effect size of r = 0.15. As a result, the analysis yielded a target sample size of 374, which we rounded to 380. Our estimated number of participants is based on previous online studies using the same approach (Hoven et al., 2023; Seow and Gillan, 2020). At least 25% of subjects in online studies are commonly excluded due to pre-defined exclusion criteria. Moreover, given the unpredictable exclusion rate (~5%), we recruited 512 subjects. After excluding participants, we obtained 249 participants for our main analysis (see the section on Participant Exclusion Criteria for details). We conducted a post hoc power analysis, and the statistical power was 77%.

General procedures

To answer our research questions, our experiment assessed a cognitive task and multiple questionnaires in a large-scale, general-population sample recruited online. We recruited the participants online through Prolific (*https://www.prolific.co/*). The inclusion criteria were that all participants must be fluent in English and aged between 20 and 60 years (note that we preregistered this as 18-65 years). Using Prolific's system, we only invited participants who reported being fluent in English. Subjects were reimbursed for their participation at an hourly rate and received a bonus based on their performance (see the "Task" section for details). We limited participants' devices to "Computer" for the task, which can also be set via the Prolific function. We separated the sample into three groups to consider the confounding effect due to time-relevant issues (e.g., chronotype (Mehrhof and Nord, 2024)). Our preregistration stated that the experiment would be conducted in two separate sessions, one in the am and one in the pm, based on UK time. However, to ensure data quality and operational stability, we allocated 250 participants each to the 9:00 AM and 5:00 PM sessions on the same day (June 20, 2025). The remaining 50 participants completed the study at 1:00 PM on the following day (June 21, 2025). This scheduling arrangement was implemented to manage the number of simultaneous participants per session, thereby minimising the potential impact of server load and concurrent access issues. This adjustment does not compromise the validity of comparisons in our main analyses.

Measures

Demographics and questionnaires

We asked about age, sex, marital status, and educational level as demographic information. We also used the following questionnaires:

-Attention/hyperactivity disorder based on Adult ADHD Self Report Scale (ASRS, 18 items) (Kessler et al., 2005)

-Fear/avoid social anxiety symptoms based on Liebowitz Social Anxiety Scale (LSAS-J, 48 items) (Baker et al., 2002)

-Impulsivity based on Barratt Impulsiveness Scale (BIS, 30 items) (Patton et al., 1995)

-Obsessive-compulsive disorder based on Obsessive-Compulsive Inventory (OCI, 42 items) (Foa et al., 1998)

-General mental health (anxiety/depression) based on Kessler Psychological Distress Scale (K6, 6 items) (Prochaska et al., 2012)

-Alcohol and tobacco use disorder based on CAGE (4 items) (Ewing, 1984)

-Tobacco use disorder based on Tobacco Dependence Screener (TDS, 10 items) (Kawakami et al., 1999)

-Gaming disorder based on Problematic Online Gaming Questionnaire (POGQ, 18 items) (Demetrovics et al., 2012)

We have selected these questionnaires, whose scores are relevant to behavioural regulation in general and are closely correlated, given their comorbidity and interrelationships (Oka et al., 2025).

Attention check

We also added an attentional check questionnaire to exclude participants who may be inappropriate. The questionnaire is located in the fourth order to check their attention at the intermediate timing. We used only one attentional check questionnaire to minimise unnecessary confounding effects on subsequent questionnaires and tasks (Hauser and Schwarz, 2015).

Items to exclude inattentive participants and control biases

Following previous research (Sarna et al., 2026; Zorowitz et al., 2023), we asked several questions to assess participants' inattentiveness and response biases. For this study, we have already modified the sentences in the original questions to fit into each questionnaire. See supplementary materials for the details.

Task

Abstraction reinforcement learning task

The abstraction task was implemented in Gorilla (Anwyl-Irvine et al., 2020). The task consists of learning the fruit preference of PAC-MAN-like characters. These characters have three features, each with two levels: colour (blue vs yellow), stripe orientation (horizontal vs vertical), and mouth direction (left vs right). On each trial, a PAC-MAN-like character composed of a unique combination of the three features was shown. The experimental session was divided into blocks, each with a specific rule that directed the association between features and preferred fruits. There were six blocks at most. There were two relevant features and one irrelevant feature, but these changed randomly at the beginning of each block. Therefore, overall, there were three types of blocks (each classified using its two relevant features): CD (colour-direction), CM (colour-mouth direction), and DM (direction-mouth orientation). For example, 'blue + vertical stripes → grapes', in which case the mouth orientation is not related to the food preference. Each rule appeared twice in each of the six blocks, and the order was randomised (see Supplementary Figure 1 for more detail). Furthermore, to avoid a simple logical deduction of the rule after one trial, the four possible combinations of two relevant features with two levels were paired with the two fruits in both a symmetric and asymmetric fashion, i.e., 2×2 or 3×1 (see Supplementary Figure 1 for the visualised tables). Moreover, the stimuli of the two fruits were also randomly changed at the beginning of each new block.

Each trial starts with a fixation screen for 2s, followed by the Pacman-like character being presented for 1s. Then, while the character remains at the centre of the screen, the two fruit options appear below it on the right and left. Participants had 5s to move a cursor to their selected position on a horizontal bar (see Supplementary Figure 2 for the average of response duration). The selected position indicates the participant's inference of the character's preference (right for the fruit to the right, left for the fruit to the left), as well as the participant's confidence in this inference (more central position = lower confidence, more lateral position = higher confidence, see Supplementary Figure 3 for the distribution).

Finally, the feedback was provided, and a tick appears on the screen if the selection is correct; otherwise, a cross appears. The feedback remains on screen for 1s, following which the trial ends with a variable ITI, bringing the trial to a fixed 8s duration.

Participants started with 40 practice trials. While each block contains up to 60 trials, a block can end earlier if participants learn the target rule. Learning is defined as a consecutive set of correct trials (ranging from 8 to 12, determined randomly in each block). Participants were instructed to receive one point for each correct choice and zero points (i.e., no reward and no penalty) for an incorrect choice. The session ended after 30 minutes, even if participants did not finish all blocks. They were further informed that bonus points were assigned, weighted by learning speed within each block. That is, the faster the rule is learned, the more points are awarded. Participants were compensated with a base payment of £6 plus a bonus (£4.2 at most) based on their performance in the abstraction task. The bonus reward was computed at the end of the session according to the formula:

$$R = \sum_{\text{block}} \left(A \cdot \frac{\sum \text{pts}}{\sum \text{tr}} - \left(\sum \text{tr} - mcs \right)^{a} \right)$$

R is the reward obtained in that block. A is the maximum available reward in a block (£0.7). ∑pts represents the sum of correct responses, and ∑tr represents the number of trials participants use in the block. mcs is the maximum length of a correct strike (12 trials), and a is a scaling factor (a = 0.005). This formula ensures time-dependent decay of the reward, which approximately follows a quadratic fit. If participants completed a block in fewer than 10 trials, the amount that would otherwise have been larger than £0.7 was rounded to £0.7. If participants do not achieve five consecutive correct answers in the practice session, they can attempt one more practice block.

We used the comprehension questions after the practice to confirm the understanding of the confidence rating. To fit into our task style, we have modified the questions that previous studies have used:

1. If you are certain you made the correct judgment, where on the scale would you place your confidence, from around the centre ('Guessing') to either edge ('Certainly correct')?
2. If you are completely unsure whether you made a correct judgment, where on the scale would you place your confidence, from around the centre ('Guessing') to either edge ('Certainly correct')? Participants must answer both questions correctly, not just one.

Controlling for cognitive confounds

We have considered including IQ as a covariate in our analyses, given previous studies reporting associations between individual differences in model-based behaviour and general intelligence (Gillan et al., 2016). However, in the context of the present study, our primary dependent variable, the abstraction-related parameter, functionally overlaps with specific components of IQ. Controlling for IQ in this context could therefore lead to overadjustment bias, in which meaningful variance related to the cognitive processes of interest (i.e., abstraction in decision-making) is inadvertently removed. Instead, we included the Deary–Liewald Reaction Time (DLRT) task (see Supplementary methods for the details) as a covariate to account for basic processing speed, which may influence performance on our task but does not overlap with the higher-order abstraction processes central to our hypotheses. This approach allows us to control for lower-level confounds while preserving the variance in the cognitive processes we aim to investigate.

Participant Exclusion Criteria

Participants were excluded from the primary analyses based on the following criteria:

1. Exclude participants whose average performance is at or below chance (as measured by a binomial test within participants): 68

2. Exclude participants with an average reading time of the primary instruction page of <5 seconds: 25
3. Double attention check failure (Attention check + infrequency item mistake more than 2): 203
4. Participants who responded with the same value (e.g., all 1s or all 4s) to all items in a questionnaire that included reverse-coded items were excluded from analysis based on inattentive responding: 0
5. Exclude participants who fail to complete the blocks within the allocated time (30 minutes): 0
6. Exclude participants who fail comprehension questions on the confidence scale: 58
7. Exclude participants who always choose the same location 80% of the time: 0

We did not use the DLRT results for this exclusion because of many unexpected participant errors (see below), which deviated from our preregistration.

Analysis

Computational modeling

We used Hierarchical Bayesian Inference (HBI) to fit the models to individual behavioural data, enabling us to estimate hyperparameters block-wise precisely (Piray et al., 2019). Following Cortese et al. (2021), we construct a classical reinforcement learning (RL) model called Q-learning. Here, a state is defined as a combination of features. The feature RL model has 2^3 = eight states, where each state is given by the combination of all three features. The abstract RL model is designed with 2^2 = 4 states, where each state is defined by the combination of two features. These learning models create individual estimates of how participant action selection depended on features they attended and their past reward history. Both RL models share the same underlying structure and are formally described as:

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r - Q(s,a))$$

where *Q(s,a)* is the value function of selecting either fruit-option a for PAC-MAN state s. The value of the action selected in the current trial is updated based on the difference between the expected value and the actual outcome (reward or no reward), known as the reward prediction error (RPE). Expected values of the two actions are combined to compute the probability P of predicting each outcome using a SoftMax (logistic) choice rule:

$$P(S_i, A) = \frac{1}{1+\exp(-\beta(Q(S_i,a_1)-Q(S_i,a_2)))}$$

The greediness hyperparameter β controls how much the difference between the two expected values actually influences choices. That is, hyperparameters estimated through likelihood maximisation were the learning rate α and the greediness (inverse temperature) β. Abstraction ability was calculated as the average ratio of the best model fitting across blocks at each participant level. To validate the model parameters, we conducted a parameter recovery analysis following established procedures (Palminteri et al., 2017). Choice data were simulated for each participant and block using their best-fitting parameters estimated from the main analysis. To mimic trial-by-trial variability, noise was introduced during simulation, with Q-value updates occasionally applied to a randomly selected state rather than the true task state (noise probability = 0.1). Simulated data were then refitted using the same HBI procedure as in the main analysis, separately for each model. Recovered and original parameter values were pooled across participants, blocks, and models, and Spearman correlations between original and recovered parameters were computed to assess recovery quality. Our pre-registration specified a sensitivity analysis that included excluded participants, and these participants differed significantly on several psychiatric scores (e.g., ADHD, impulsivity; $p < 0.001$). Nevertheless, this analysis was not conducted because the hierarchical Bayesian model pools information across participants, meaning that including them would alter parameter estimates for the retained sample. We therefore focus on the pre-defined, strictly screened sample.

Statistical analysis

We conducted three bootstrap multiple linear regressions (5,000 iterations, case-resampling with replacement, preserving the joint covariate distribution) based on a generalised linear model to examine the relationship between task-related measures as dependent variables (parameters) and 1) psychiatric symptoms and 2) transdiagnostic dimensional scores. The entire regression model was re-estimated at each iteration, and two-tailed bootstrap *p*-values were computed as 2 × min (*p_left*, *p_right*), where *p_left* and *p_right* denote the proportions of bootstrap samples falling below and above zero, respectively. In these regression models, we also controlled for age, sex, educational level (treated as an ordinal variable), and processing speed (calculated from phase 1 of the DLRT task) as fixed effects. Note that we did not use Phase 2 data in the DLRT task because numerous participants had inadvertently skipped the instruction, and we excluded a random effect for participants from our pre-registered models because the term was meaningless for these between-participant analyses. We also applied a multiple-comparison adjustment based on the false discovery rate (Benjamini and Hochberg, 1995) to symptom-based analyses, as we repeated the model testing across scores.

Specifically, the model equations were defined as follows (Wilkinson notation),

$$\begin{aligned}&\text{Abstraction ability} \sim \text{Intercept} + \text{psychiatric score} + \text{Age} + \text{Sex}\\&+ \text{Educational level} + \text{Processing speed.}\\&\text{Abstraction ability} \sim \text{Intercept} + \text{psychiatric factors} + \text{Age} + \text{Sex}\\&+ \text{Educational level} + \text{Processing speed.}\end{aligned}$$

To avoid multicollinearity, the independent psychiatric score was input separately, e.g.:

$$\begin{gathered}\text{Abstraction ability} \sim \text{Intercept} \ + \ \text{ADHD score} \ + \ \text{Age} \ + \ \text{Sex}\\ + \ \text{Educational level} \ + \ \text{Processing speed.}\end{gathered}$$

Psychiatric factor scores were input together, i.e., Factor 1 + Factor 2 + Factor 3.

We explored the relationship between psychiatric scores/factors and confidence (i.e., based on the distance from the cursor location to the centre of the slider). The confidence level was calculated as the average across blocks. We added the averaged response to content-neutral items as a control variable to account for general response tendencies because recent research suggested that failing to control for such biases might distort the association between confidence and subjective psychopathology (Sarna et al., 2026).

Specifically, the model equations are defined as follows,

$$\begin{aligned}&\text{Confidence level} \sim \text{Intercept} + \text{psychiatric score} + \text{Age} + \text{Sex}\\&+ \text{Educational level} + \text{Processing speed}\\&+ \text{Averaged response to content}-\text{neutral.}\\&\text{Confidence level} \sim \text{Intercept} + \text{psychiatric factors} + \text{Age} + \text{Sex}\\&+ \text{Educational level} + \text{Processing speed}\\&+ \text{Averaged response to content}-\text{neutral.}\end{aligned}$$

Additionally, we calculated metacognitive sensitivity based on confidence. We applied the area under the type 2 ROC curve (AUROC2) as a metacognitive sensitivity metric and input it as the dependent variable. Note that participants (N = 5) who used only a single confidence rating across all trials were excluded from AUROC2 computation, as the Type-2 ROC curve cannot be defined in the absence of rating variance. These participants were, however, retained in analyses using mean confidence ratings, as invariant responding may itself reflect a meaningful metacognitive response style.

Specifically, the model equations are defined as follows,

$$\text{Metacognitive sensitivity} \sim \text{Intercept} + \text{psychiatricscore} + \text{Age} + \text{Sex} \\ + \text{Educationallevel} + \text{Processing speed} \\ + \text{Averaged response to content-neutral.}$$

$$\text{Metacognitive sensitivity} \sim \text{Intercept} + \text{psychiatric factors} + \text{Age} \\ + \text{Sex} + \text{Educational level} + \text{Processing speed} \\ + \text{Averaged response to content-neutral.}$$

Moreover, to explore the interaction effects between metacognitive sensitivity and psychiatric scores for abstraction ability. These models did not improve the model fit according to the likelihood ratio test (see Results and Supplementary Table 7).

Data availability

Data, relevant scripts for all analyses, and other supplementary materials can also be found on OSF (DOI: *10.17605/OSF.IO/QA4BK*).

Acknowledgements

Funders played no role in the study design, data collection, analysis, or interpretation. We thank Miho Nagata for their help scheduling the experiments. We also thank Mor Ovadia for helping with the implementation of experiments.

Additional information

Author contributions

Taiki Oka: Conceptualisation, Resources, Data curation, Software, Formal analysis, Funding acquisition, Validation, Investigation, Visualisation, Methodology, Writing - original draft, Project administration, Writing - review and editing
Yoshihiko Kunisato: Conceptualisation, Software, Formal analysis, Supervision, Validation, Methodology, Writing - review and editing
Kotaro Koizumi: Formal analysis, Validation, Visualisation, Writing - review and editing
Misa Murakami: Resources, Investigation, Writing - review and editing
Hugo Six: Methodology, Writing - review and editing
Jessica E Taylor: Formal analysis, Investigation, Methodology, Writing - review and editing
Aurelio Cortese: Conceptualisation, Software, Formal analysis, Supervision, Funding acquisition, Validation, Methodology, Writing - original draft, Project administration, Writing - review and editing

Funding

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant Number 24K22822 (T.O.), Grants-in-Aid for Scientific Research on Innovative Areas for Transformative Research Areas A Grant Number 22H05160 (A.C.), the Innovative Science and Technology Initiative for Security (ATLA) Grant Number JPJ004596 (A.C.), and the Institute for Basic Science (IBS-R015-D2) (A.C.).

Funding

Funder	Grant reference number	Author
Japan Society for the Promotion of Science	24K22822	Taiki Oka
Japan Society for the Promotion of Science	22H05160	Aurelio Cortese
Acquisition, Technology & Logistics Agency	JPJ004596	Aurelio Cortese
Institute for Basic Sciences	IBS-R015-D2	Aurelio Cortese

Author ORCID iDs

Taiki Oka: https://orcid.org/0000-0002-6403-183X
Aurelio Cortese: https://orcid.org/0000-0003-4567-0924

Additional files

Supplementary materials

References

Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, Evershed JK (2020) Gorilla in our midst: An online behavioral experiment builder. *Behav Res Methods* **52**:388-407 https://doi.org/10.3758/s13428-019-01237-x | PubMed

Baker SL, Heinrichs N, Kim H-J, Hofmann SG (2002) The liebowitz social anxiety scale as a self-report instrument: a preliminary psychometric analysis. *Behav Res Ther* **40**:701-715 https://doi.org/10.1016/s0005-7967(01)00060-2 | PubMed

Banca P, Herrojo Ruiz M, Gonzalez-Zalba MF, Biria M, Marzuki AA, Piercy T, Sule A, Fineberg NA, Robbins TW (2024) Action sequence learning, habits, and automaticity in obsessive-compulsive disorder. *eLife* **12**:RP87346 https://doi.org/10.7554/eLife.87346 | PubMed

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc* **57**:289-300 https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benzina N, N'Diaye K, Pelissolo A, Mallet L, Burguière E (2021) A cross-species assessment of behavioral flexibility in compulsive disorders. *Commun Biol* **4**:96 https://doi.org/10.1038/s42003-020-01611-y | PubMed

Brühlmann F, Petralito S, Aeschbach LF, Opwis K (2020) The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Meth Psychol* **2**:100022 https://doi.org/10.1016/j.metip.2020.100022

Caspi A, Moffitt TE (2018) All for One and One for All: Mental Disorders in One Dimension. *Am J Psychiatry* **175**:831-844 https://doi.org/10.1176/appi.ajp.2018.17121383 | PubMed

Cortese A (2022) Metacognitive resources for adaptive learning⋆. *Neurosci Res* **178**:10-19 https://doi.org/10.1016/j.neures.2021.09.003 | PubMed

Cortese A, Yamamoto A, Hashemzadeh M, Sepulveda P, Kawato M, De Martino B (2021) Value signals guide abstraction during learning. *eLife* **10**:e68943 https://doi.org/10.7554/eLife.68943 | PubMed

De Martino B, Cortese A (2022) Goals, usefulness and abstraction in value-based choice. *Trends Cogn Sci* https://doi.org/10.1016/j.tics.2022.11.001 | PubMed

Demetrovics Z, Urbán R, Nagygyörgy K, Farkas J, Griffiths MD, Pápay O, Kökönyei G, Felvinczi K, Oláh A (2012) The development of the Problematic Online Gaming Questionnaire (POGQ). *PLoS One* **7**:e36417 https://doi.org/10.1371/journal.pone.0036417 | PubMed

Doyle H, McLaughlin NCR, Garnaat SL, Desrochers TM (2026) Cognitive sequences in obsessive-compulsive disorder are supported by frontal cortex ramping activity. *Imaging Neurosci (Camb)* **4**:IMAG.a.1084 https://doi.org/10.1162/imag.a.1084 | PubMed

Dubois M, Hauser TU (2022) Value-free random exploration is linked to impulsivity. *Nat Commun* **13**:4542 https://doi.org/10.1038/s41467-022-31918-9 | PubMed

Eling P, Derckx K, Maes R (2008) On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain Cogn* **67**:247-253 https://doi.org/10.1016/j.bandc.2008.01.006 | PubMed

Embon I, Cukier S, Iorio A, Barttfeld P, Solovey G (2023) Is visual metacognition associated with autistic traits? A regression analysis shows no link between visual metacognition and Autism-Spectrum Quotient scores. *Conscious Cogn* **110**:103502 https://doi.org/10.1016/j.concog.2023.103502 | PubMed

Ewing JA (1984) Detecting alcoholism. The CAGE questionnaire. *JAMA* **252**:1905-1907 https://doi.org/10.1001/jama.252.14.1905 | PubMed

Foa EB, Kozak MJ, Salkovskis PM, Coles ME, Amir N (1998) The validation of a new obsessive-compulsive disorder scale: The Obsessive-Compulsive Inventory. *Psychological Assessment* https://doi.org/10.1037/1040-3590.10.3.206

Fox CA, Teckentrup V, Donegan KR, Seow TX, Benwell CS, Tervo-Clemmens B, Gillan CM (2025) Cognitive arbitration between candidate dimensions of psychopathology. *Mol Psychiatry* 1-14 https://doi.org/10.31234/osf.io/7waey_v2

Fried EI (2022) Studying Mental Health Problems as Systems, Not Syndromes. *Curr Dir Psychol Sci* 09637214221114089 https://doi.org/10.1177/09637214221114089

Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND (2016) Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* **5** https://doi.org/10.7554/eLife.11305 | PubMed

Gillan CM, Seow TXF (2020) Carving Out New Transdiagnostic Dimensions for Research in Mental Health. *Biol Psychiatry Cogn Neurosci Neuroimaging* **5**:932-934 https://doi.org/10.1016/j.bpsc.2020.04.013 | PubMed

Gjelsvik B, Lovric D, Williams JMG (2018) Embodied cognition and emotional disorders: Embodiment and abstraction in understanding depression. *J Exp Psychopathol* **9**:r.035714 https://doi.org/10.5127/pr.035714

Hauser DJ, Schwarz N (2015) It's a trap! Instructional manipulation checks prompt systematic thinking on “tricky” tasks. *SAGE Open* **5**:215824401558461 https://doi.org/10.1177/2158244015584617

Ho MK, Abel D, Griffiths TL, Littman ML (2019) The value of abstraction. *Curr Opin Behav Sci* **29**:111-116 https://doi.org/10.1016/j.cobeha.2019.05.001

Hoven M, Luigjes J, Denys D, Rouault M, van Holst RJ (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

Jia W, Zhu H, Ni Y, Su J, Xu R, Jia H, Wan X (2020) Disruptions of frontoparietal control network and default mode network linking the metacognitive deficits with clinical symptoms in schizophrenia. *Hum Brain Mapp* **41**:1445-1458 https://doi.org/10.1002/hbm.24887 | PubMed

Kahnt T (2022) Computationally informed interventions for targeting compulsive behaviors. *Biol Psychiatry* https://doi.org/10.1016/j.biopsych.2022.08.028

Katyal S, Huys QJ, Dolan RJ, Fleming SM (2025) Distorted learning from local metacognition supports transdiagnostic underconfidence. *Nat Commun* **16**:1854 https://doi.org/10.1038/s41467-025-57040-0 | PubMed

Kawakami N, Takatsuka N, Inaba S, Shimizu H (1999) Development of a screening questionnaire for tobacco/nicotine dependence according to ICD-10, DSM-III-R, and DSM-IV. *Addict Behav* **24**:155-166 https://doi.org/10.1016/s0306-4603(98)00127-0 | PubMed

Kessler RC, Adler L, Ames M, Demler O, Faraone S, Hiripi E, Howes MJ, Jin R, Secnik K, Spencer T, *et al.* (2005) The World Health Organization Adult ADHD Self-Report Scale (ASRS): a short screening scale for use in the general population. *Psychol Med* **35**:245-256 https://doi.org/10.1017/s0033291704002892 | PubMed

Kist JD, Vrijsen JN, Mulders PCR, van Eijndhoven PFP, Tendolkar I, Collard RM (2023) Transdiagnostic psychiatry: Symptom profiles and their direct and indirect relationship with well-being. *J Psychiatr Res* **161**:218-227 https://doi.org/10.1016/j.jpsychires.2023.03.003 | PubMed

Kobayashi H, Isohanni M, Jääskeläinen E, Miettunen J, Veijola J, Haapea M, Järvelin M-R, Jones PB, Murray GK (2014) Linking the developmental and degenerative theories of schizophrenia: association between infant development and adult cognitive decline. *Schizophr Bull* **40**:1319-1327 https://doi.org/10.1093/schbul/sbu010 | PubMed

Lee DG, Hare TA (2023) Value certainty and choice confidence are multidimensional constructs that guide decision-making. *Cogn Affect Behav Neurosci* **23**:503-521 https://doi.org/10.3758/s13415-022-01054-4 | PubMed

Lee JK, Rouault M, Wyart V (2025) Compulsivity is linked to suboptimal choice variability but unaltered reinforcement learning under uncertainty. *Nat Ment Health* 1-13 https://doi.org/10.1038/s44220-024-00364-5

Maia TV, Frank MJ (2011) From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci* **14**:154-162 https://doi.org/10.1038/nn.2723 | PubMed

Mehrhof SZ, Nord CL (2024) A common alteration in effort-based decision-making in apathy, anhedonia, and late circadian rhythm. *eLife* **13** https://doi.org/10.7554/elife.96803.3

Oka T, Sasaki A, Kobayashi N (2025) A transdiagnostic dimensional approach to behavioral dysregulation: Examining self-reported reward and punishment sensitivity across psychopathology. *J Affect Disord* **387**:119493 https://doi.org/10.1016/j.jad.2025.119493 | PubMed

Osborne JW, Costello AB (2004) Sample size and subject to item ratio in principal components analysis. *Pract Assess Res Eval* **9**:11 https://doi.org/10.7275/ktzq-jq66

Palminteri S, Wyart V, Koechlin E (2017) The importance of falsification in computational cognitive modeling. *Trends Cogn Sci* **21**:425-433 https://doi.org/10.1016/j.tics.2017.03.011 | PubMed

Patton JH, Stanford MS, Barratt ES (1995) Factor structure of the Barratt impulsiveness scale. *J Clin Psychol* **51**:768-774 https://doi.org/10.1002/1097-4679(199511)51:6<768::aid-jclp2270510607>3.0.co;2-1 | PubMed

Pinkham AE, Klein HS, Hardaway GB, Kemp KC, Harvey PD (2018) Neural correlates of social cognitive introspective accuracy in schizophrenia. *Schizophr Res* **202**:166-172 https://doi.org/10.1016/j.schres.2018.07.001 | PubMed

Piray P, Dezfouli A, Heskes T, Frank MJ, Daw ND (2019) Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Comput Biol* **15**:e1007043 https://doi.org/10.1371/journal.pcbi.1007043 | PubMed

Prochaska JJ, Sung H-Y, Max W, Shi Y, Ong M (2012) Validity study of the K6 scale as a measure of moderate mental distress based on mental health treatment need and utilization. *Int J Methods Psychiatr Res* **21**:88-97 https://doi.org/10.1002/mpr.1349 | PubMed

Robbins TW, Banca P, Belin D (2024) From compulsivity to compulsion: the neural basis of compulsive disorders. *Nat Rev Neurosci* https://doi.org/10.1038/s41583-024-00807-z | PubMed

Rouault M, Seow T, Gillan CM, Fleming SM (2018) Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biol Psychiatry* **84**:443-451 https://doi.org/10.1016/j.biopsych.2017.12.017 | PubMed

Sarna N, Dar R, Mazor M (2026) Biased and inattentive responding contribute to apparent metacognitive biases in mental health. *Proc Natl Acad Sci U S A* **123**:e2520832123 https://doi.org/10.1073/pnas.2520832123 | PubMed

Seow T, Benoit E, Dempsey C, Jennings M, Maxwell A, O'Connell R, Gillan CM (2021) Model-Based Planning Deficits in Compulsivity Are Linked to Faulty Neural Representations of Task Structure. *J Neurosci* **41**:6539-6550 https://doi.org/10.1523/jneurosci.0031-21.2021 | PubMed

Seow T, Fleming SM, Hauser TU (2025a) Metacognitive biases in anxiety-depression and compulsivity extend across perception and memory. *PLOS Ment Health* **2**:e0000259 https://doi.org/10.1371/journal.pmen.0000259 | PubMed

Seow T, Gillan CM (2020) Transdiagnostic Phenotyping Reveals a Host of Metacognitive Deficits Implicated in Compulsivity. *Sci Rep* **10**:1-11 https://doi.org/10.1038/s41598-020-59646-4 | PubMed

Seow T, Jelinek L, Moritz S, Hauser TU (2025b) Integrating cognitive neuroscience and clinical perspectives on metacognitive mechanisms in psychopathology. *Nat Rev Psychol* https://doi.org/10.1038/s44159-025-00503-4

Seow T, Rouault M, Gillan CM, Fleming SM (2021) How local and global metacognition shape mental health. *Biol Psychiatry* https://doi.org/10.1016/j.biopsych.2021.05.013 | PubMed

Taschereau-Dumouchel V, Cortese A, Lau H, Kawato M (2021) Conducting decoded neurofeedback studies. *Soc Cogn Affect Neurosci* **16**:838-848 https://doi.org/10.1093/scan/nsaa063 | PubMed

Taschereau-Dumouchel V, Cushing CA, Lau H (2022) Real-Time Functional MRI in the Treatment of Mental Health Disorders. *Annu Rev Clin Psychol* **18**:125-154 https://doi.org/10.1146/annurev-clinpsy-072220-014550 | PubMed

Vaghi MM, Luyckx F, Sule A, Fineberg NA, Robbins TW, 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 | PubMed

Wise T, Robinson OJ, Gillan CM (2023) Identifying Transdiagnostic Mechanisms in Mental Health Using Computational Factor Modeling. *Biol Psychiatry* **93**:690-703 https://doi.org/10.1016/j.biopsych.2022.09.034 | PubMed

Zhang K, Trambaiolli L, Zhao Z, Taschereau-Dumouchel V, Feusner JD (2026) Neurofeedback interventions for obsessive-compulsive and related disorders: Current evidence and future directions. *J Psychiatr Res* **198**:1-12 https://doi.org/10.1016/j.jpsychires.2026.03.013 | PubMed

Zorowitz S, Solis J, Niv Y, Bennett D (2023) Inattentive responding can induce spurious associations between task behaviour and symptom measures. *Nat Hum Behav* **7**:1667-1681 https://doi.org/10.1038/s41562-023-01640-7 | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

This work investigated the associations between abstraction and metacognition in the context of reward-guided learning and transdiagnostic symptom dimensions in a sample of N = 249 participants. Participants completed a reward-guided learning task and confidence judgements. Transdiagnostic dimensions used to examine associations with abstraction and metacognition were based on a large existing dataset. Findings showed that in the examined sample, a Compulsive Hypersensitivity dimension was negatively associated with abstraction and metacognitive sensitivity, while a Social Withdrawal dimension was positively associated with metacognitive sensitivity. These data add to the existing literature on associations between metacognition and transdiagnostic symptom dimensions and extend previous work on the association between abstraction and these dimensions.

Strengths:

(1) The study addresses an interesting research question and uses a transdiagnostic approach. While part of the research question is a replication (metacognition), the additional inclusion of an abstraction parameter is highly valuable.

(2) Methodologically, the study is strong. Specifically, the implementation of an experimental task to estimate parameters, the hierarchical Bayesian modelling, the parameter recovery analyses, the bootstrap regression (including corrections for multiple comparisons) and the control of relevant covariates and response tendencies are quite impressive.

(3) The Open Science approach is laudable in general. The study was preregistered and provides open data and open code. Deviations from the preregistration are transparently reported (e.g., bootstrap regression, exclusion criteria). In this vein, the high number of robustness analyses provided in the supplements is very much appreciated.

(4) More generally, the extensiveness of the supplements is particularly valuable.

(5) Finally, it is very useful that the discussion takes into account potential alternative explanations of the findings.

Weaknesses:

(1) High number of exclusions:

As the authors mention (also in their limitations section), more than half of the participants had to be excluded (only 249 out of 512 participants remained), which is substantial. Specifically, 203 participants were excluded because they failed attention checks, 58 failed comprehension questions on the confidence scale, 25 had a reading time of the instruction page below 5 seconds, and 68 showed a performance that was too poor (the sum is probably higher than 263 because these overlap). This high number of excluded cases resulted in a substantially smaller analysed sample than planned (corresponding to approximately 77% power instead of 90%) and potentially limited both statistical power and generalisability. It might even be the case that highly impulsive individuals were excluded systematically. That is, the exclusion criteria may themselves be associated with psychiatric symptoms, so the analysed sample may no longer (fully) represent the target population. It would be interesting to see whether the results also hold when including the excluded participants (because excluded and included participants differ in attentional and impulsivity-related symptoms). A sensitivity analysis would be beneficial.

(2) Deviations from preregistration:

While the preregistration of the study is positive in general, some aspects raise doubts here. First, there have been quite a few deviations from the preregistration. For instance, the primary outcome variable for abstraction has been changed (initially: proportion of blocks in which the Abstract RL model had a better fit than the Feature RL model; instead: mean posterior responsibility of the Abstract RL model across blocks), and this appears to have influenced the findings (e.g., Supplementary Figures: 8 vs. 9). Generally, these deviations introduce additional researcher degrees of freedom and therefore warrant a more detailed justification (or replication in future studies). Second, it appears that the preregistration was uploaded only to an OSF folder and not formally registered with a timestamp. However, while an update to the document has been made according to the metadata, the content still seems to be the same as the original one (created on Jun 11, 2025).

(3) Combined measurement of choices and metacognition:

Choice behaviour and metacognition were not measured independently (i.e., they were measured using a single slider). This challenges the interpretation of results concerning metacognitive sensitivity, as the authors note themselves in the limitations section (even if the effect of starting position was small). It should be clarified whether responses exactly at the centre of the slider were possible or not (and what range the slider had).

(4) Generalisability of findings:

It is questionable whether transdiagnostic factors derived from a Japanese population can be transferred to the sample in this work, especially because it seems to be composed of a heterogeneous international sample comprising participants from multiple countries (South Africa, United States, United Kingdom, Poland, others). While it is appreciated that the authors validate the transdiagnostic factors through a new exploratory factor analysis and by correlating item loadings across samples, the correlations are only moderate and point at least to some degree of variation. In addition, for factor 2 (social withdrawal?), the correlation seems to be mainly existent due to two clusters of items, challenging the validation approach in itself. Also, apart from the correlations themselves, the absolute values of the item loadings appear to vary largely. Considering this, the conclusion that "transdiagnostic symptom dimensions appear broadly consistent across samples (even across countries)" (p. 9) definitely goes too far.

(5) Parameter recovery analysis: The parameter recovery analysis is appreciated; however, the recovery of the learning rate in particular seems to be rather low (r=.40). This raises concerns regarding the validity of the parameter estimates.

(6) Effect sizes: The reported regression coefficients appear relatively small in magnitude (i.e., betas of the associations between metacognition/abstraction and transdiagnostic dimensions appear to be in the range around .02-.04). If these are standardised estimates, the corresponding effect sizes are relatively small. If not, I recommend reporting standardised effect sizes in addition.

Overall, the study provides relevant replications and new insights into the association between reward-guided learning behaviour (abstraction, metacognition) and transdiagnostic symptom dimensions. It should be noted that effect sizes appear to be rather low, however. The analytic approach is generally strong. At the same time, the high number of exclusions, the limitations regarding the validation of the transferred transdiagnostic factor structure, and the limited recovery of the learning rate parameter clearly raise important questions regarding the robustness and generalizability of findings. Given this, any interpretations regarding potential therapeutic implications (e.g., p. 10) should be made with caution.

https://doi.org/10.7554/eLife.112401.1.sa3

Reviewer #2 (Public review):

Summary:

In this study, Oka and colleagues recruited an online sample to complete a previously validated abstraction task (Cortese et al., 2021) alongside confidence ratings and a large psychiatric questionnaire battery, which included a variety of methods to screen out inattentive or otherwise biased responders. Questionnaire item scores were combined with factor weights from a large dataset to estimate transdiagnostic factor scores. A computational model was then fit in a hierarchical manner to the abstraction task data, with individuals' fit to an "Abstract RL" model used as a metric of individual-level abstraction ability, and metacognitive bias and sensitivity were estimated from the confidence ratings. Associations between these task-derived measures and both dimensional and symptom-level measures of psychopathology were then estimated using multiple regression. The key findings were that, while metacognitive sensitivity and abstraction ability were associated with symptom-level scores, the associations with transdiagnostic dimensions - higher compulsivity associated with lower abstraction ability and metacognitive sensitivity; higher social withdrawal was associated with higher metacognitive sensitivity - were interpreted by the authors as more coherent.

Strengths:

(1) Robust screening for inattentive responders through catch questions (Zorowitz et al., 2023), as well as incorporating recent recommendations regarding response bias (Sarna et al., 2026).

(2) Assessed the cross-cultural generalisability of the imported factor weights by comparing item loadings from a large external sample against a de novo exploratory factor analysis in their sample.

(3) Directly compares a theory-driven model-defined abstraction metric to metacognition in relation to dimensional and symptom-level measures of psychopathology.

(4) Pre-registered analyses, with deviations from pre-registration clearly stated.

Weaknesses:

(1) The abstraction metric (mean posterior responsibility of the Abstract RL model) differed from the pre-registered metric and has not been validated here for reliability (e.g., split-half across the blocks or similar).

(2) Model recovery is not shown, so it's not clear whether the Abstract RL and Feature RL models are fully dissociable in this task design.

(3) Metacognitive measures are behaviourally defined (AUROC2 for sensitivity and mean confidence for bias), but models do not correct for task accuracy, which may be related to both.

(4) Dimensional and symptom regressions differ: the dimensions are entered in one model, but the symptom measures are entered into separate regressions and the marginal effects corrected for multiple comparisons. If I've understood this correctly, this means that the dimensional coefficients are partial associations adjusting for the other two factors, whereas the symptom-level coefficients are marginal and FDR-corrected, making it difficult to directly compare them.

Additional questions and context:

(1) Could split-half reliability (e.g. odd vs even blocks) be reported for the abstraction measure? Relatedly, a model recovery/confusion analysis for the two abstraction models, and/or posterior predictive checks showing that the two models generate behaviour resembling that of participants would help establish that the responsibility metric is able to dissociate the different abstraction strategies.

(2) Supplementary Table 2 shows the results for the pre-registered discrete proportion metric - here, there is limited evidence (p=0.220) of an association between abstraction and compulsivity, so saying they are "almost consistent" is perhaps a little overstated. Though the argument for using the alternative continuous metric is justified in the text, it's not quite clear whether the difference is due to the inference method or the abstraction metric itself - the bootstrapped analysis of the pre-registered metric is not reported, nor is the analysis without bootstrapping of the continuous metric (I think this may have been what Supplementary Table 1 was meant to report, but currently it's identical to Supplementary Table 5). In addition, it might be helpful if the correlation between the two metrics were presented graphically.

(3) In the Methods and Supplement, the authors mention that they had pre-registered running a sensitivity analysis including excluded participants. This might be interesting given the high exclusion rate, and given that most exclusions were not based on task behaviour (chance-level choosing). If there is concern about shifting group-level parameter distributions, then this could be explicitly included in the model by including an offset on group-level parameters (i.e., interaction term) on excluded participants, which would allow them to systematically differ in model parameters. Alternatively, one could at least estimate the abstraction and metacognition metrics in the excluded sample (perhaps restricted to those excluded on questionnaire-based criteria rather than task performance) to see whether they do indeed differ.

(4) How do factor scores relate to task accuracy - do those with higher compulsivity perform worse, and is this plausibly related to less abstraction?

(5) How do the factors extracted here compare to those in other studies, such as those from Gillan et al. (2016, eLife)? In particular, it'd be interesting to know what

questionnaires/symptoms in the "Compulsive hypersensitivity" factor in the present study overlap with the Compulsive behaviour/intrusive thoughts factor from that earlier work, as the latter has been strongly associated with metacognitive measures - higher metacognitive efficiency for anxious/depression, lower metacognitive efficiency for compulsive behaviour - in previous work (Rouault et al., 2018). That three-factor structure also included a factor they labelled "Social withdrawal" - is it similar to the one presented here, or is the one here (including distress) more like their anxious depressive factor?

(6) In the Discussion, the authors state "Our findings are also consistent with previous converging evidence linking compulsive tendencies to less efficient computation and a preference for familiar over goal-directed action". That the Feature RL model might fairly be called less efficient is reasonable, but I'm not sure how the reduction of features in the Abstract RL model relates to goal-directed action (they're both model-free RL algorithms).

(7) The Discussion also mentions "models that integrate abstract and metacognitive representations" - was there a reason these could not be applied in the present study?

https://doi.org/10.7554/eLife.112401.1.sa2

Reviewer #3 (Public review):

Summary:

This is an interesting study investigating the relationship between transdiagnostic symptom profiles and computational parameters of abstraction and metacognition. The authors found that a transdiagnostic dimension they term compulsive hypersensitivity was negatively related to abstraction ability and metacognitive sensitivity, and a dimension termed social withdrawal was positively related to metacognitive sensitivity. Overall, the question of whether and how higher-level cognitive processes relate to transdiagnostic psychiatric factors is interesting and addresses a relevant gap in the literature.

Strengths:

This is an overall well-designed study, and the authors have clearly given considerable thought to data quality and validation during the study design. This is evident, for instance, in the use of multiple approaches to assess inattentive responding and acquiescence tendencies. In addition, the use of advanced modelling approaches for the task data represents a strength and allows the authors to capture individual differences in task behaviour in a more nuanced and mechanistic manner in comparison to what would have been possible with model-agnostic analyses.

Weaknesses:

Nevertheless, I would like to raise several concerns, detailed below.

(1) My most pressing concern regards the exclusion rate of roughly half the sample. Of the 512 participants who were tested, only 249 (48.6%) were included in the final analyses, meaning that more than half of the recruited participants were excluded. Although the authors state that these exclusions were based on preregistered criteria, this high exclusion rate still warrants further investigation. Pre-registering exclusion criteria does not eliminate the potential for selection bias or establish that the resulting sample is representative of the recruited sample. In fact, the authors even report that the excluded participants differed significantly on several psychiatric scores. I believe that a detailed account of the exclusion, including how included and excluded participants differed on relevant demographic or study variables, would be beneficial.

Additionally, the authors state that a sensitivity analysis including participants excluded from the primary analysis was preregistered but was not conducted because the excluded and included participants differed significantly. This does not seem a sufficient justification for omitting a preregistered sensitivity analysis; in fact, systematic differences between included and excluded participants warrant this kind of sensitivity analysis. I agree with the authors that this may lead to changes in task parameter estimates. However, I believe this change would be an informative result of the sensitivity analyses rather than a methodological problem to avoid.

Finally, I would like the authors to clarify some inconsistencies in the preregistered exclusion criteria. In the pre-registration, they first state that all participants who fail any infrequency question will be excluded. Later in the pre-registration, they state that participants with two or more failed questions will be excluded and that a sensitivity analysis will be done, in which participants with only one failed question are retained. Neither of these criteria matches what is reported in the manuscript ("Attention check + infrequency item mistake more than 2").

(2) The current sample size of n = 249 participants is not sufficient for a factor analysis with 176 questionnaire items, and I believe that the solution of using factor weights from a previous larger sample is generally sensible. I also commend the authors for wanting to assess the validity of this approach. However, both the methods and results reported for the comparison between the factor analysis based on the current, smaller sample and the large, previous sample are insufficient and warrant substantially more detail. Firstly, it is unclear to me whether the three-factor solution in the factor analysis with the current, smaller sample was empirically grounded or whether three factors were extracted to match the three-factor solution from the larger sample. For both factor analyses, I would recommend reporting the factor extraction methods, the criterion used to determine the number of factors, and whether alternative factor solutions were considered. Secondly, it is unclear what exactly was correlated across the two factor analyses (e.g., factor scores, item loadings, factor weights?). Finally, I do not believe that the result of significant correlation is sufficient to conclude that the dimensions replicate between samples, particularly given they indicate only moderate correspondence (r = 0.46, for instance, corresponds to only 21% shared variance), thereby providing very limited evidence for replication of the factor structure.

These concerns are particularly pressing given that the authors state the consistency of transdiagnostic symptom dimensions across samples and countries as a primary result in the discussion.

(3) It is also unclear to me whether any exclusion was applied based on the first part of the Deary-Liewald reaction time task. The supplement implies that it was ("This task result was used to [...] exclude participants who exhibit problematic behaviours"), but I could not find a corresponding report in the manuscript.

(4) The manuscript refers to an attentional check questionnaire used to identify and exclude inattentive participants, but does not provide a reference for this measure or describe the questions included. Given the substantial overall exclusion rate, it is particularly important that all exclusion procedures and criteria are described in sufficient detail to allow readers to assess their appropriateness and reproducibility. The authors should therefore provide the relevant reference and/or report the specific questions and criteria used to determine inattention.

Relatedly, when describing the control-neutral items in the Supplementary Materials, the authors state that further details are provided in the Supplementary Materials. However, I cannot find another, separate Supplementary Material that may contain this information

https://doi.org/10.7554/eLife.112401.1.sa1

Author response:

We thank the editors for the eLife Assessment and the reviewers for their thorough and constructive evaluation of our manuscript.

We are glad that the evidence for reduced metacognitive sensitivity in relation to the compulsive hypersensitivity dimension was considered solid. In hindsight, we agree that the evidence for the abstraction findings is currently incomplete, and we plan to address this with additional analyses as outlined below (to clarify the robustness of the abstraction metric and its association with symptom dimensions).

Below, we outline how we plan to address the main points raised by the reviewers, grouped thematically, given the overlap across the three reviews.

A full, detailed point-by-point response accompanied by the corresponding analyses will follow in our formal revision response.

(1) Exclusion rate and lack of relevant sensitivity analysis

In the revision, we will report a detailed comparison of included versus excluded participants on demographic and symptom variables, and we will conduct the originally preregistered sensitivity analysis to estimate abstraction and metacognition metrics, including the excluded sample, and compare these metrics with psychopathological variables, rather than omitting this analysis.

(2) Abstraction metric and its deviation from the preregistered definition

We recognise that our primary measure of abstraction differs from the preregistered metric (the proportion of blocks better fit by the Abstract RL model), and that this change appears to affect the pattern of results. In the revision, we will report split-half reliability for the abstraction metric, present the bootstrapped analysis of the preregistered metric, and provide a clearer justification for the switch, making the source of the discrepancy (metric versus inference method) transparent.

(3) Generalisability of the transdiagnostic factor structure

We acknowledge that our claim of consistency between samples of the transdiagnostic dimensions requires more support and more careful framing. In the revision, we will provide full methodological detail on both factor analyses (extraction method, criteria for the number of factors, etc.), and we will moderate our interpretation, especially in the discussion, to reflect the actual strength of correspondence rather than describing the structure as broadly consistent.

(4) Parameter and model recovery

We will extend our recovery analyses to include a model recovery/confusion analysis between the Abstract RL and Feature RL models, and we will investigate the source of the relatively low learning-rate recovery (e.g., by testing recovery stratified by block length and without injected noise), reporting these results in the revision.

(5) Documentation of the attention-check procedure

We will provide all details on the attention-check items and criteria used to identify inattentive participants, including a reference where applicable, and we will standardise terminology (e.g., infrequency vs. inattention items) throughout the manuscript and supplementary information.

(6) Other methodological and presentational clarifications

We will review the manuscript, methods, and supplementary materials to resolve the remaining methodological ambiguities and presentational issues raised by the reviewers. This includes, among other points, clarifying whether the reported regression coefficients are standardised and correcting labelling inconsistencies, missing references/DOIs, and other minor textual issues throughout.

https://doi.org/10.7554/eLife.112401.1.sa0