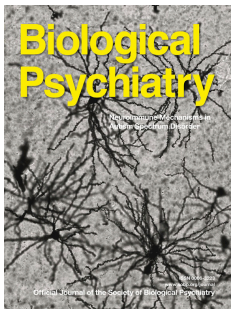
Journal Pre-proof



How local and global metacognition shape mental health

Tricia X.F. Seow, Marion Rouault, Claire M. Gillan, Stephen M. Fleming

PII: S0006-3223(21)01329-9

DOI: https://doi.org/10.1016/j.biopsych.2021.05.013

Reference: BPS 14562

To appear in: *Biological Psychiatry*

Received Date: 18 November 2020

Revised Date: 14 May 2021

Accepted Date: 16 May 2021

Please cite this article as: Seow T.X.F., Rouault M., Gillan C.M. & Fleming S.M., How local and global metacognition shape mental health, *Biological Psychiatry* (2021), doi: https://doi.org/10.1016/j.biopsych.2021.05.013.

TITLE

How local and global metacognition shape mental health

SHORT TITLE

How local and global metacognition shape mental health

AUTHOR LIST

Tricia X.F. Seow[1,2,*], Marion Rouault[3,4,*], Claire M. Gillan[5] and Stephen M. Fleming[1,2,6]

(* indicates equal contribution)

AUTHOR AFFILIATIONS

[1]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, 10-12 Russell Square, London WC1B 5EH, UK.

[2]Wellcome Centre for Human Neuroimaging, University College London, 12 Queen Square, London WC1N 3AR, UK.

[3]Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL University, 75005 Paris, France.

[4]Laboratoire de neurosciences cognitives et computationnelles, Département d'études cognitives, ENS, INSERM, PSL University, 75005 Paris, France.

[5]School of Psychology, Trinity College Dublin, Dublin, Ireland.

[6]Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, UK.

NAME AND CONTACT INFORMATION OF CORRESPONDING AUTHORS

Marion Rouault and Tricia X.F. Seow

- *Marion Rouault: Département d'Études Cognitives, École Normale Supérieure, 29 rue d'Ulm, 75005 Paris, France. Email: marion.rouault@gmail.com.*
- *Tricia X.F. Seow: Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, 10-12 Russell Square, WC1B 5EH London, UK. Email: t.seow@ucl.ac.uk*

KEYWORDS

Metacognition, self-beliefs, self-efficacy, confidence, mental health, transdiagnostic psychiatry

ABSTRACT

Metacognition is the ability to reflect on our own cognition and mental states. It is a critical aspect of human subjective experience and operates across many hierarchical levels of abstraction—encompassing “local” confidence in isolated decisions and “global” self-beliefs about our abilities and skills. Alterations in metacognition are considered foundational to neurological and psychiatric disorders, but research has mostly focused on local metacognitive computations, missing out on the role of global aspects of metacognition. Here, we first review current behavioral and neural metrics of local metacognition that lay the foundation for this research. We then address the neurocognitive underpinnings of global metacognition uncovered by recent studies. Finally, we outline a theoretical framework in which higher hierarchical levels of metacognition may help identify the role of maladaptive metacognitive evaluation in mental health conditions, particularly when combined with transdiagnostic methods.

INTRODUCTION

Metacognition, the ability to reflect on and evaluate our own thoughts and actions, is a crucial component of human behavior and subjective experience (1). A wealth of empirical studies have shown that impaired metacognition is associated with detrimental behavior and poor mental health (2,3). For instance, delusional thinking in schizophrenic patients is thought to be maintained by metacognitive deficits such as a lack of insight (4) or overconfidence in incorrect models of the world (5–7). In a range of mental health conditions, metacognition shows consistent, yet specific, individual differences (8,9) (see review (2)), findings that generalize across various tasks (10) and cognitive domains (11), and abnormalities that may be heritable (12). As researchers in psychiatry aim to develop reliable neurocognitive markers for identifying current and future mental health problems, metacognitive assessments hold promise (13).

There are several challenges in meeting this aim. First, metacognition is tightly bound to cognitive performance, such as the accuracy of visual decisions or memory recollection. Second, metacognition manifests in various hierarchical levels of abstraction, from "local" confidence in isolated decisions to more "global" metacognitive constructs like self-efficacy beliefs. Whilst most research has focused on local metacognition, we propose that global aspects of metacognition may be more closely related to daily functioning and the subjective experience of mental health symptoms. Lastly, metacognitive changes may not be readily apparent in case-control comparisons using standard diagnostic categories and instead be better captured by transdiagnostic dimensions. Here, we introduce the main behavioral and neural metrics of local metacognition, discuss the relevance of global metacognition for mental health and outline how transdiagnostic methodologies may help to unpack the role of multiple hierarchical levels of metacognition in psychiatry. Note that the disorders which we raise as examples are those with the greatest relevance to the transdiagnostic studies we discuss later.

Section 1: Methods for quantifying local metacognition

Behavioral and computational metrics of local metacognition

Several metrics have been developed to quantify local metacognition in laboratory tasks, most of which rely on examining the correspondence between objective performance and confidence ratings (a subjective report of being correct about a decision/statement

(14)) across multiple experimental trials (***Figure 1A***). Two independent aspects of local metacognition can be distinguished: metacognitive bias and sensitivity (15) (***Figure 1B***). Metacognitive bias reflects how confident we are irrespective of actual performance, and is usually estimated as the mean confidence rating averaged over correct and incorrect judgements. In contrast, metacognitive sensitivity reflects an ability to discriminate correct from incorrect judgements. A participant who rates high confidence on correct judgements and low confidence on incorrect judgements is estimated to have high metacognitive sensitivity.

An initial wave of studies relied on simple correlation statistics, which conflated metacognitive bias and sensitivity in one measure, an issue covered previously (16,17). More recent methods (i.e., type 2 signal detection theory (SDT)) estimate a bias-free assessment of metacognitive sensitivity (18). However, metacognitive sensitivity is typically dependent on task performance, where easier tasks produce greater sensitivity (16). Instead, model-based methods reliant on SDT (e.g., meta-d' model) correct for such performance confounds, leading to the derivation of summary statistics such as metacognitive *efficiency* that represent a participant's level of metacognitive sensitivity corrected for variation in task performance (17). Another approach is to use staircase procedures (19,20) that adjust task performance at a predetermined level, allowing variation in metacognitive sensitivity to be isolated (e.g., (8)), although this method has caveats (21). Failures to replicate metamemory biases towards lowered confidence in obsessive-compulsive disorder (OCD) (22–31) or recent evidence of previously inflated effects (32) of higher confidence in errors from delusion-prone and paranoid schizophrenic patients (5,6,33–38) were ultimately explained by metacognitive sensitivity and bias not being properly separated. Future experiments should aim to minimize potential confounds in estimating metacognitive sensitivity at either the paradigm design or analysis stage.

Neural bases of local metacognition

Beyond behavioral metrics, studies have begun to reveal the neural bases of local metacognition about perception and memory (see review (11)). Strong convergent evidence highlights the importance of prefrontal cortex (PFC) for metacognition. Lesions (39) or transcranial magnetic stimulation (40) to the PFC affect perceptual metacognitive sensitivity while leaving task performance unaffected. Structural and functional MRI

studies in healthy humans have linked individual differences in anterior PFC volume, function and connectivity to metacognitive ability (8,20,41–47). Beyond PFC, a distributed network of brain regions including the medial PFC, precuneus and hippocampus (20,43,44,46–52) are also involved in metacognition. Electrophysiology studies provided convergent evidence of activity associated with metacognition in prefrontal theta oscillations (53), the P3 ERP component (54) and the error-related negativity (ERN) (55–57). Similar neural correlates are observed in relation to aberrant metacognitive processes in some psychiatric disorders. Altered metacognition about perceptual decisions in schizophrenia patients correlates with hypoactivity in fronto-parietal areas (58), and also hippocampal volume and its grey matter microstructure (59). Drug addiction, which was linked to deficits in error awareness (60) and perceptual metacognitive sensitivity (61), was linked to hypoactivity and a loss of structural integrity in the anterior cingulate cortex. Overall, the medial PFC and parietal cortex are proposed to play a domain-general role in metacognition, with other nodes of the network contributing in a domain-specific fashion (11) (***Figure 2***).

Section 2: From local confidence to global self-beliefs

Many forms of metacognition co-exist

While the psychological and neural bases of local metacognition are increasingly well characterized, its functional roles remain less clear. Local confidence has been suggested to regulate subsequent decisions by recruiting cognitive control (62), gathering information (63), controlling exploration (64) and adapting speed-accuracy trade-offs (65). However, these are all limited in scope and on short time scales. In contrast to local confidence in single decisions, global metacognitive evaluations of performance (“self-beliefs”) can span several decisions or experimental trials, allowing for a gradual formation of self-performance estimates in numerous aspects: about our ability on a given task, in a specific cognitive domain, or even how capable we feel, broadly (***Figure 3***). In turn, these self-beliefs may affect future decisions on longer time scales (66,67), such as promoting the initiation of behavioral sequences towards achieving a goal. Individuals with low self-beliefs tend to feel less in control of their environment, are less likely to believe that their decisions will affect future outcomes, and are slower to recover after setbacks (68,69). Accordingly, distorted self-beliefs may have a pervasive impact in educational and clinical settings (70), determining how people see themselves and their capabilities. But despite their recognized importance

for mental health, the cognitive and neural foundations of self-beliefs remain largely unclear.

Self-beliefs are related to the psychological construct of self-esteem, a global notion of self-worth that cuts across many domains (e.g. physical, social and academic) (71). Low self-esteem is a key predictor of mental health issues such as anxiety and depression (72,73). Low self-esteem has strong theoretical ties to dominant clinical psychology models of depression (74), where depressive symptoms are thought to be grounded in negative schema that persist despite alternative evidence (75). Negative schemas encompass several processes, among which confidence/self-beliefs are one critical aspect, with the proposed neural correlates of negative schemas and confidence partly overlapping, e.g. cingulate cortex (76). However, despite the strong face validity of these negative schema, their measurement with clinical scales precludes a mechanistic understanding of how these self-reports arise (77). In contrast, models of global metacognition constitute a mechanistic framework within which to define testable hypotheses, and unpack the mechanisms underpinning low self-beliefs. For instance, we can examine how shifts in processes supporting local decision confidence lead to gradual changes in global self-beliefs that likely unfold over longer timescales. The study of apathy provides a recent example—a single self-report (i.e., apathetic state) could be attributed to various computational mechanisms (reduced reward sensitivity or increased subjective perception of effort), each associated with distinct neurobiological systems (78,79).

Neurocognitive foundations of simple forms of global self-beliefs

We have begun to delineate computations underlying the formation of global self-beliefs from local confidence estimates (80,81). In these experiments, participants were asked to perform mini blocks of two interleaved perceptual tasks. At the end of each block, they selected the task which they thought they performed best—a proxy for global self-beliefs about the two tasks. Local subjective confidence ratings were found to predict global self-beliefs over and above objective performance (80). Using functional MRI, we further found that ventral striatal activity reflected the level of global self-beliefs (but not local confidence signals), while confidence-related activity in ventromedial PFC (vmPFC) was further modulated by the level of global self-belief (81). This is in line with two studies indicating that vmPFC reflects fluctuations in self-performance estimates on mini games

performed across several trials when participants monitor expected task success with (64) or without (82) external feedback. Moreover, white matter structural integrity between ventral striatum and vmPFC, estimated using DTI, shows systematic links with individual self-esteem (83). These results establish an initial link between local and global metacognition (***Figure 2***), and reveal neural representations of global self-beliefs that go beyond the tracking of local confidence (84).

It is important to acknowledge that global self-beliefs assessed in these studies were limited to the scope of a lab experiment, and to perceptual (80,84) or color/time estimation tasks (82). These tasks are well characterized in terms of how local perceptual decisions and confidence estimates are formed (e.g., (85)), which is vital for precisely quantifying how self-beliefs are constructed from local confidence and external feedback (88). However, there is a substantial gap between experimental investigations of so-called global self-beliefs and self-beliefs relevant to real-life decisions, which typically fluctuate over considerably longer time scales than those assessed in the laboratory. Additionally, many other factors contribute to the formation of real-life self-beliefs, such as feedback from other people and one's social environment (86,87). We suggest that we can bridge the gap by examining how self-beliefs generalize across different tasks and across cognitive domains (***Figure 3***). Such a generalization mechanism should normally support the formation of useful priors about expected ability in closely related tasks. But if this mechanism becomes maladaptive, leading to. e.g., excessive generalization from local experiences, it could create volatile self-beliefs. Conversely, a disruption in updating mechanisms could result in rigid self-beliefs being insufficiently updated in light of new positive experiences.

Relating global self-beliefs to functional symptoms

Adapting a framework for global metacognition may prove useful clinically because it may be more directly relevant to the subjective and functional experiences of patients as compared to local confidence in isolated decisions. For example, anosognosia, defined as a lack of awareness of cognitive deficits, particularly about memory, is a common symptom of dementia (88). A lack of self-awareness may lead to a failure to adapt to changes in cognitive abilities, for instance leading to risky behaviors such as driving long distances or traveling to unfamiliar locations (89). Anosognosia may also affect decisions about appropriate courses of treatment or prevent the implementation of

strategies to aid memory such as setting reminders (90,91). Similarly, intact global metacognition may be crucial for treatment adherence as an individual may only be willing to participate in therapeutic interventions if they have insight into their symptoms. Previous work with schizophrenic patients has indeed shown that clinical insight is predictive of medication compliance (92,93).

At present, only local confidence is routinely measured in experimental studies of metacognition. However, there is likely a complex and largely unexplored interplay between local metacognitive evaluations and global self-beliefs. Notably, anosognosia may co-exist with relatively intact local metacognition about performance on individual trials. In these studies (94–96), participants with Alzheimer's disease underwent assessments of local metacognition on memory and motor tasks, and clinicians evaluated the patients' global awareness of their deficits (95). While both local memory and motor metacognition were found to be relatively intact (89,95), there was a specific deficit of global awareness in the memory (and not motor) domain (95), suggesting that local and global metacognitive levels may dissociate in some cognitive domains, but not others. We note, however, that extended clinical interviews and/or informants' reports were used as proxies for ground-truth ability; as such, the data remains disconnected from approaches that seek to model the relationship between performance and confidence.

Global and local metacognition also diverge in Parkinson's disease. Patients differ from healthy participants in their feeling of knowing accuracy in recognition memory tests at the item level, but not in their global prediction of accuracy (97). These examples highlight the value of a neurocognitive framework encompassing local and global metacognition, to pinpoint the origins of lack of awareness (80). It could be that symptom severity only affects upper hierarchical levels (***Figure 3***), or creates imbalances between global and local metacognitive processing within a specific domain. Similar to anosognosia, functional cognitive disorder, a condition characterized by the experience of persistent and distressing subjective cognitive difficulties in the absence of detectable objective cognitive deficit and underlying neurological disease (98,99), is thought to be explained by changes in metacognitive ability. However, it is unknown which layer(s) of the metacognitive hierarchy, if any, are affected in this condition. Likewise, patients with motor conversion disorder report difficulties in performing certain motor actions without

any apparent neurological disease. Prior work using a visuomotor task revealed that patients are just as aware and confident in trajectory deviations as control participants, but they engaged distinct brain networks when estimating their confidence (100). In this case, distortions in the formation of global self-beliefs may be central in explaining a mismatch between an internal subjective experience of poor self-ability and otherwise intact objective performance and local metacognition (***Figure 3***).

The various layers in a putative metacognitive hierarchy are likely to be more fine-grained than the local/global dichotomy highlighted here. For instance, we can make a distinction between "how well did I perform this task today at work?" and "how well am I performing at my job in general?". The levels of metacognition outlined here (***Figure 3***) partly map onto a previously proposed psychological framework for characterizing global awareness in dementia (101) that distinguishes four levels: sensory pre-registration (basic evaluation), performance monitoring (corresponding to so-called local metacognition here), evaluative judgement and meta-representation. However, in this model, the latter two constructs were defined in relation to how others see us, rather than in relation to objective experimental measurements.

Interim conclusion

Building a complete theoretical framework supported by empirical evidence of how various levels of metacognition relate to each other is important since global self-beliefs are a major determinant of our behavior. Unlike local metacognition, which is often tied to a particular task or cognitive domain, changes in global self-beliefs may generalize to other domains and to a range of daily life functions (89). In turn, global self-beliefs may be more directly relevant for understanding the mechanistic and computational bases of global aspects of subjective experience such as low mood or self-esteem characteristic of negative schemas in depression (80).

Section 3: A transdiagnostic approach for uncovering associations between metacognition and mental health symptoms

If local and global metacognition are to be neurocognitive markers for psychopathology, their robustness and specificity are important. Psychiatric research suggests that the use of the Diagnostic and Statistical Manual of Mental Disorders (DSM) categories pose a concern for these goals (102) due to high comorbidity rates, and, symptom variability

and complexity within each diagnosis (***Figure 4A & 4B***). For instance, a reduction in memory confidence is often observed in OCD individuals but this has been linked to elevated levels of other mental health symptoms in OCD patient samples (e.g., depression), rather than obsessive-compulsive symptoms per se (22). Hence, accounting for co-morbid symptoms appears crucial for understanding the precise clinical consequences of abnormalities in metacognition, and ultimately allow us to map symptoms more closely to behavior and neural circuits (102–104) (***Figure 4C***).

Transdiagnostic studies of local metacognition

Recent studies have leveraged transdiagnostic approaches to uncover links between symptom dimensions and metacognition. With self-reported symptoms in nine psychiatric questionnaires (105), we characterized large online general population samples along three symptom dimensions (anxious-depression, compulsive behavior and intrusive thought (henceforth 'compulsivity') and social withdrawal; replicated from a prior study (106)). Using a perceptual decision-making task and local confidence ratings, we found that the anxious-depression dimension was associated with lower confidence, whereas the compulsivity dimension was related to higher confidence (***Figure 5***). These results stand in contrast to classic questionnaire scores showing that OCD symptoms alone were not linked to any alterations in confidence (***Figure 5***), similar to prior findings (107,108). This is because anxiety and depression, which are both linked to lower local confidence judgments, overlapped with OCD scores (109,110), masking a positive association between confidence and compulsivity. These findings suggest that metacognitive dysfunctions previously observed may be masked by the co-occurrence of other symptoms, particularly if different families of symptoms predict opposing effects on confidence.

A transdiagnostic approach therefore provides context for interpreting prior metacognition findings in case-control studies of OCD. Vaghi and colleagues employed a reinforcement learning task where participants predicted where a particle will land and report their confidence in catching the particle (108). They observed a form of decreased metacognitive sensitivity in OCD as compared to healthy participants (smaller correlation between confidence and behavioral adjustments of their prediction), without a difference in local confidence or in how sensitive participants' confidence was to task events (e.g., sudden changes in landing location). Conducting the same paradigm in a large online

general population sample, we replicated Vaghi *et al.*'s finding of an impaired relationship between confidence and behavioral adjustments in OCD (111). However, using a dimensional approach, we found that higher confidence (as in the perceptual task (110)) (***Figure 5***) and a lower sensitivity of confidence to task events were linked to compulsivity symptoms. These studies demonstrate that transdiagnostic approaches can be crucial in delineating hidden metacognitive relationships and enhancing our understanding of psychopathology.

To our knowledge, the transdiagnostic studies presented above are the only ones applying such approaches to local metacognitive metrics. By using the same three-dimensional structure across multiple studies, we can prevent the overfitting of new psychiatric dimensions to data. Indeed, the same compulsivity dimension linked to metacognitive deficits (105,111), is also associated with goal-directed failures (106), enhanced learning from safety than threat (112), reduced avoidance of cognitive effort (113) and faulty neural representations of task structure knowledge (114). In the case of goal-directed control, deficits are seen in online (106) and in-person samples (114) alike and work in patients has shown these deficits are more strongly linked to variation in a compulsive dimension than a diagnosis of OCD (115). Although these findings are suggestive, it remains to be seen if the metacognitive abnormalities associated with these dimensions are also altered in patient samples. We also note that these dimensions may not necessarily describe cognitive alterations better than DSM-defined psychopathology or other transdiagnostic structures (116–118). Alternative dimensional or hierarchical approaches to phenotyping (119) remain to be tested in the context of metacognition, and may be superior (120–122). As psychiatry continues to improve how we define mental health and illness in the population, we can expect cycles of iterative evolution of dimensional phenotypes (both those of interest and those to be controlled for) (123).

Intersecting hierarchies of metacognition with transdiagnostic approaches

Transdiagnostic approaches have revealed that individuals with strong anxious-depression symptoms have lower local confidence, whereas those with compulsivity have higher confidence (105,111). However, the same individual can experience both anxiety and compulsivity symptoms (e.g., OCD). We argue that such opposing effects of confidence between anxious-depression and compulsivity may be unraveled by better

distinguishing between local confidence and global self-beliefs. It is likely that an individual's local belief about performance is not pure and instead involves numerous, and at least partially dissociable, neural and computational processes. Local confidence ratings in anxious-depression may be 'contaminated', i.e., driven by global estimates of self-performance unrelated to the current task, while local confidence ratings in compulsivity could reflect selective abnormalities in local evidence evaluation processes. This explanation is supported by observations that anxiety and depression symptoms are strongly linked to low self-esteem (72,73) while compulsivity is associated with difficulties in developing and using models to solve decision-making tasks (114,124). In sum, a local confidence rating could depend both on a global prior about self-ability and a local evaluation of performance.

Schizophrenic patients have frequently been reported to be overconfident about individual (local) decisions (2,37). However, recent moderation analyses suggest that this metacognitive deficit is based on studies in which other cognitive performance features vary across participants, thereby questioning whether the overconfidence effect is a central deficit (32). This issue is likely exacerbated by the inclusion of variable diagnoses (e.g., bipolar disorder or depression with psychosis) beyond schizophrenia in prior studies (32). Certain forms of schizophrenia also include high levels of apathy which could be partly linked to low global subjective expectations of success (125). As positive and negative symptoms co-exist in schizophrenia, combining a transdiagnostic perspective while considering different levels of metacognition may be fundamental to delineating the underlying psychopathology. For this reason we advocate that future studies use tasks that can distinguish, and simultaneously control for, multiple levels of metacognition (80). Cross-task comparisons might prove useful too, as we hypothesize that reductions in local confidence in depression, if driven by global self-beliefs, should be relatively impervious to task design, and generalize across domains (10). In contrast, if local confidence biases in compulsivity are the result of an issue with 'model-building', we expect the finding of over-confidence to be highly sensitive to task demands.

Clinical implications

Metacognitive beliefs have long been a therapeutic target. Metacognitive therapy (MCT) for anxiety, depression (70,126,127), OCD (128,129) and schizophrenia (130,131) focus on modifying intrusive thoughts and cognitive biases to dampen maladaptive rumination,

compulsive rituals or delusional ideation. However, MCT efficacy is not useful for all patients (132–134), and little is known about the underlying neural mechanisms facilitating symptom alleviation (135). Assessing metacognition before and after MCT treatment should help formalize a mechanistic and neural model of how clinical gains occur, and establish if it is through metacognitive processes. Meanwhile, recent studies have shown that training can improve metacognitive ability (136,137) (though with exceptions (138)). A next step is to examine if these metacognitive changes have therapeutic benefit, i.e., transfer beyond a particular training or therapeutic session and generalize to real-world functioning. Gaining an understanding of the factors promoting generalization will be critical for devising tools for improving metacognition (136,137,139) and modifying self-beliefs through psychotherapy (70,140).

The current evidence for a relationship between mental health and metacognition is correlational. Translating these insights to the clinic requires probing these associations causally and in longitudinal designs. A key question is whether abnormalities in metacognitive bias and sensitivity resolve when symptoms improve, or are relatively stable traits that may signal an overall risk for developing a mental health condition. Drawing on adjacent literature, there is some evidence to suggest that negative biases in face perception improve following antidepressant drug administration in depressed patients and predict subsequent clinical response (141). If metacognitive bias follows a similar pattern as negative biases, it may constitute a similar predictor of treatment outcome. Quantifying metacognition could therefore have clinical value if changes in metacognitive parameters help to identify individuals at risk, facilitate early intervention, guide us as to who might respond best to a given treatment, or assist in developing transdiagnostic treatment protocols that target metacognition (142–144).

CONCLUSIONS

Theories about the role of metacognition in mental health may be enriched by adopting quantitative task-based methods for measuring metacognition across different hierarchical levels (***Figure 3***) together with robust transdiagnostic approaches (***Figure 4***). Many other aspects of metacognition have yet to be looked at in relation to mental health, and the paradigms and models described here represent a starting point. The current review serves as a framework for thinking about how different levels of metacognition (from local to global) are interrelated, possibly by generalization

mechanisms, and outlines hypotheses for how these map onto transdiagnostic dimensions of mental health.

ACKNOWLEDGEMENTS

TXFS is a post-doctoral fellow at the Max Planck UCL Centre for Computational Psychiatry and Ageing Research. The Max Planck UCL Centre for Computational Psychiatry and Ageing Research is a joint initiative supported by UCL and the Max Planck Society. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust (203147/Z/16/Z).

MR is the beneficiary of a post-doctoral fellowship from the AXA Research Fund. MR work was also supported by a department-wide grant from the Agence Nationale de la Recherche (ANR-17-EURE-0017, EUR FrontCog). This work has received support under the program «Investissements d'Avenir» launched by the French Government and implemented by ANR (ANR-10-IDEX-0001-02 PSL).

CMG is supported by a fellowship from MQ: transforming mental health (MQ16IP13) and holds grant funding from Science Foundation Ireland's Frontiers for the Future Award (19/FFP/6418).

SMF is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and Royal Society (206648/Z/17/Z).

CONFLICTS OF INTEREST

The authors report no biomedical financial interests or potential conflicts of interest.

FIGURE LEGENDS

Figure 1*. Experimental task-based measures of metacognition.*
A*) Relationship between first-order performance and second-order confidence ("local" trial-by-trial confidence judgements). On each trial, participants provide a report of their level of confidence in a decision/choice they have made, which can either be objectively correct or incorrect.*
B*) Two independent metrics of metacognition - bias and sensitivity. Each schematic graph shows a probability distribution of confidence ratings for correct and incorrect trials separately. The x-axis represents confidence reports increasing from left to right. Metacognitive sensitivity is the extent to which confidence discriminates between correct and incorrect trials, corresponding to the separation between the distributions. Metacognitive bias is the overall confidence level across both correct and incorrect trials. Confidence distributions are Gaussian for illustration purposes but are likely to take other forms depending on the generative model.*

Figure 2*. Neural correlates of metacognitive evaluation*
Schematic sagittal slice and lateral view of the human brain highlighting the role of prefrontal cortex (PFC) in metacognition. Studies of local metacognition have highlighted the ventro-medial prefrontal cortices (vmPFC) and posterior medial frontal cortex (pMFC) as central hubs reflecting confidence estimates [a: (9,145)*] and error detection [b:* (146–148)*], while the frontopolar cortex (FPC), together with the lateral prefrontal cortex (laPFC) are involved in mediating explicit metacognitive judgements, (meta)cognitive control and subsequent behavioral regulation [c:* (8,47)*; d:* (40,43)*]. Some of the neural substrates linked to local metacognition exhibit cognitive domain-specificity e.g., the precuneus (PRECU) has mostly been implicated in metamemory [e:* (9,44,46,149)*], whilst lateral-parietal areas (laPAR) are mostly implicated in metaperception [f:* (47,150)*]. Recent studies have begun to reveal that neural substrates of global metacognition only partly overlap with those of local metacognition. In particular, in vmPFC and precuneus, local confidence signals were found to be further modulated by the level of global self-belief on a perceptual task (81).*

Figure 3*. Multiple hierarchical levels of metacognitive evaluation*
Reciprocal interactions between "local" confidence judgements in isolated decisions and more global self-beliefs. Previous work has revealed that local confidence contributes to the formation of global self-beliefs, but global self-beliefs are also likely to in turn influence local confidence. Under this framework, local confidence may reflect a combination of a local component related to decision performance evaluation and a global component formed over the aggregation of multiple experiences across various tasks and domains formed through learning. On the right, examples are given to illustrate each hierarchical level in the domain of memory, though the true distinction between levels is likely to be more fine-grained. Each of these metacognitive levels is associated with dynamics unfolding across different timescales, with higher levels of the

hierarchy having slower dynamics than lower levels. Global self-beliefs may shape and be shaped by even more global constructs such as an individual's level of self-esteem.

***Figure 4**. Dimensional approaches to psychiatry addressing within- and across-diagnosis homogeneities and heterogeneities.*
***A & B**) Case-control studies comparing diagnosed patient and healthy control groups have often failed to recognize that patients have varying levels of other psychopathologies (e.g., compulsivity, anxiety, etc.) beyond the one under study. Comparing such groups (typically, ranging between 15 and 50 participants per group) have often revealed ambiguous or non-specific effects in relation to metacognition.*
***C**) Mathematical methods of dimensionality reduction allow identification of latent factors underlying various mental health conditions. These dimensions may better reflect the psychopathological complexity underlying traditional psychiatric categories, and uncover more consistent relationships with metacognition. OCD (obsessive-compulsive disorder) and GAD (generalized anxiety disorder) reflect traditional diagnostic categories. In contrast, Anxious and Compulsive dimensions reflect transdiagnostic symptom dimensions. Typically, transdiagnostic dimensions are estimated using groups of hundreds or thousands of participants.*

***Figure 5**. Relationships of confidence and psychiatric symptoms (standard approach), or with psychiatric dimensions (transdiagnostic approach), across two different paradigms. AD: anxious-depression dimension, CIT: compulsive behavior and intrusive thought ("compulsivity") dimension, SW: social withdrawal dimension. Confidence abnormalities linked to psychiatric symptoms using the standard approach are inconsistent across studies. However, with a transdiagnostic approach, the finding of lowered confidence with anxious-depression and higher confidence with compulsivity replicates across tasks. The y axis indicates the change in z-scored confidence for each change of 1 standard deviation of symptom/dimension scores. Error bars denote standard error. *p < .05, **p < .01, ***p < .001 corrected for multiple comparisons, ^{o}p < .05, uncorrected. Figures are reproduced from their original studies* (105,111). *Note that performance was controlled for using a staircase procedure in the perceptual discrimination task and was not related to symptom dimensions* (105). *Task performance also showed no relationship with symptom dimensions in the reinforcement learning task* (111).

REFERENCES

1. Flavell JH. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. Am Psychol. 1979;34(10):906.
2. Hoven M, Lebreton MMM, Engelmann JB, Denys D, Luigjes J, van Holst RJ. Abnormalities of confidence in psychiatry: an overview and future perspectives. Transl Psychiatry. 2019;9(1):1–18.
3. Sun X, Zhu C, So SHW. Dysfunctional metacognition across psychopathologies: a meta-analytic review. Eur Psychiatry. 2017;45:139–53.
4. Engh JA, Friis S, Birkenaes AB, Jónsdóttir H, Klungsøyr O, Ringen PA, et al. Delusions are associated with poor cognitive insight in schizophrenia. Schizophr Bull. 2010;36(4):830–5.
5. Moritz S, Woodward TS, Ruff CC. Source monitoring and memory confidence in schizophrenia. Psychol Med. 2003;33(1):131–9.
6. Moritz S, Woodward TS, Whitman JC, Cuttler C. Confidence in errors as a possible basis for delusions in schizophrenia. J Nerv Ment Dis. 2005;193(1):9–16.
7. Moritz S, Ramdani N, Klass H, Andreou C, Jungclaussen D, Eifler S, et al. Overconfidence in incorrect perceptual judgments in patients with schizophrenia. Schizophr Res Cogn. 2014;1(4):165–70.
8. Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. Relating introspective accuracy to individual differences in brain structure. Science (80-). 2010;329(5998):1541–3.
9. Morales J, Lau H, Fleming SM. Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. J Neurosci. 2018;38(14):3534–46.
10. Ais J, Zylberberg A, Barttfeld P, Sigman M. Individual consistency in the accuracy and distribution of confidence judgments. Cognition. 2016;146:377–86.
11. Rouault M, McWilliams A, Allen MG, Fleming SM. Human metacognition across domains: insights from individual differences and neuroimaging. Personal Neurosci. 2018;1.
12. Cesarini D, Lichtenstein P, Johannesson M, Wallace B. Heritability of overconfidence. J Eur Econ Assoc. 2009;7(2–3):617–27.
13. Paulus MP, Huys QJMM, Maia T V. A Roadmap for the Development of Applied Computational Psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging. 2016;
14. Pouget A, Drugowitsch J, Kepecs A. Confidence and certainty: Distinct

probabilistic quantities for different goals. Nat Neurosci. 2016;19(3):366.

15. Fleming SM, Lau HC. How to measure metacognition. Front Hum Neurosci. 2014;8:443.
16. Maniscalco B, Lau H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious Cogn. 2012;
17. Fleming SM. HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. Neurosci Conscious. 2017;2017(1):nix007.
18. Guggenmos M. Validity and reliability of metacognitive performance measures. PsyArXiv March. 2021;23.
19. Song C, Kanai R, Fleming SM, Weil RS, Schwarzkopf DS, Rees G. Relating inter-individual differences in metacognitive performance on different perceptual tasks. Conscious Cogn. 2011;20(4):1787–92.
20. Allen M, Glen JC, Müllensiefen D, Schwarzkopf DS, Fardo F, Frank D, et al. Metacognitive ability correlates with hippocampal and prefrontal microstructure. Neuroimage. 2017;149:415–23.
21. Rahnev D, Fleming SM. How experimental procedures influence estimates of metacognitive ability. Neurosci Conscious. 2019;2019(1):niz009.
22. Moritz S, Jacobsen D, Willenborg B, Jelinek L, Fricke S. A check on the memory deficit hypothesis of obsessive–compulsive checking. Eur Arch Psychiatry Clin Neurosci. 2006;256(2):82–6.
23. Moritz S, Ruhe C, Jelinek L, Naber D. No deficits in nonverbal memory, metamemory and internal as well as external source memory in obsessive-compulsive disorder (OCD). Behav Res Ther. 2009;47(4):308–15.
24. Moritz S, Kloss M, von Eckstaedt FV, Jelinek L. Comparable performance of patients with obsessive-compulsive disorder (OCD) and healthy controls for verbal and nonverbal memory accuracy and confidence:. Time to forget the forgetfulness hypothesis of OCD? Psychiatry Res. 2009;166(2–3):247–53.
25. Tekcan Aİ, Topçuoğlu V, Kaya B. Memory and metamemory for semantic information in obsessive–compulsive disorder. Behav Res Ther. 2007;45(9):2164–72.
26. McNally RJ, Kohlbeck PA. Reality monitoring in obsessive-compulsive disorder. Behav Res Ther. 1993;31(3):249–53.
27. Cougle JR, Salkovskis PM, Wahl K. Perception of memory ability and confidence in recollections in obsessive-compulsive checking. J Anxiety Disord.

2007;21(1):118–30.

28. Ecker W, Engelkamp J. Memory for actions in obsessive-compulsive disorder. Behav Cogn Psychother. 1995;23(4):349–71.
29. Foa EB, Amir N, Gershuny B, Molnar C, Kozak MJ. Implicit and explicit memory in obsessive-compulsive disorder. J Anxiety Disord. 1997;11(2):119–29.
30. Macdonald PA, Antony MM, Macleod CM, Richter MA. Memory and confidence in memory judgments among individuals with obsessive compulsive disorder and non-clinical controls. Behav Res Ther. 1997;35(6):497–505.
31. Tolin DF, Abramowitz JS, Brigidi BD, Amir N, Street GP, Foa EB. Memory and memory confidence in obsessive-compulsive disorder. Behav Res Ther. 2001;39(8):913–27.
32. Rouy M, Saliou P, Nalborczyk L, Pereira M, Roux P, Faivre N. Systematic review and meta-analysis of the calibration of confidence judgments in individuals with schizophrenia spectrum disorders. Neurosci Biobehav Rev. 2021;(in press).
33. Bhatt R, Laws KR, McKenna PJ. False memory in schizophrenia patients with and without delusions. Psychiatry Res. 2010;178(2):260–5.
34. Gawęda Ł, Moritz S, Kokoszka A. Impaired discrimination between imagined and performed actions in schizophrenia. Psychiatry Res. 2012;195(1–2):1–8.
35. Kircher TTJ, Koch K, Stottmeister F, Durst V. Metacognition and reflexivity in patients with schizophrenia. Psychopathology. 2007;40(4):254–60.
36. Moritz S, Woodward TS. Memory confidence and false memories in schizophrenia. J Nerv Ment Dis. 2002;190(9):641–3.
37. Moritz S, Woodward TS, Rodriguez-Raecke R. Patients with schizophrenia do not produce more false memories than controls but are more confident in them. Psychol Med. 2006;36(5):659.
38. Kwok SC, Xu X, Duan W, Wang X, Tang Y, Allé MC, et al. Autobiographical and episodic memory deficits in schizophrenia: A narrative review and proposed agenda for research. Clin Psychol Rev. 2020;101956.
39. Fleming SM, Ryu J, Golfinos JG, Blackmon KE. Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. Brain. 2014;137(10):2811–22.
40. Shekhar M, Rahnev D. Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. J Neurosci. 2018;38(22):5078–87.
41. Hilgenstock R, Weiss T, Witte OW. You'd better think twice: post-decision

perceptual confidence. Neuroimage. 2014;99:323–31.

42. Yokoyama O, Miura N, Watanabe J, Takemoto A, Uchida S, Sugiura M, et al. Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. Neurosci Res. 2010;68(3):199–206.
43. Qiu L, Su J, Ni Y, Bai Y, Zhang X, Li X, et al. The neural system of metacognition accompanying decision-making in the prefrontal cortex. PLoS Biol. 2018;16(4):e2004037.
44. McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, De Lange FP, Lau H. Anatomical coupling between distinct metacognitive systems for memory and visual perception. J Neurosci. 2013;33(5):1897–906.
45. Sinanaj I, Cojan Y, Vuilleumier P. Inter-individual variability in metacognitive ability for visuomotor performance and underlying brain structures. Conscious Cogn. 2015;36:327–37.
46. Baird B, Smallwood J, Gorgolewski KJ, Margulies DS. Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. J Neurosci. 2013;33(42):16657–65.
47. Vaccaro AG, Fleming SM. Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. Brain Neurosci Adv. 2018;2:2398212818810591.
48. Ren Y, Nguyen VT, Sonkusare S, Lv J, Pang T, Guo L, et al. Effective connectivity of the anterior hippocampus predicts recollection confidence during natural memory retrieval. Nat Commun. 2018;9(1):1–10.
49. Kim H, Cabeza R. Trusting our memories: dissociating the neural correlates of confidence in veridical versus illusory memories. J Neurosci. 2007;27(45):12190–7.
50. Moritz S, Gläscher J, Sommer T, Büchel C, Braus DF. Neural correlates of memory confidence. Neuroimage. 2006;33(4):1188–93.
51. Chua EF, Schacter DL, Rand-Giovannetti E, Sperling RA. Understanding metamemory: neural correlates of the cognitive process and subjective level of confidence in recognition memory. Neuroimage. 2006;29(4):1150–60.
52. Henson RNA, Rugg MD, Shallice T, Dolan RJ. Confidence in recognition memory for words: dissociating right prefrontal roles in episodic retrieval. J Cogn Neurosci. 2000;12(6):913–23.

53. Wokke ME, Cleeremans A, Ridderinkhof KR. Sure I'm sure: prefrontal oscillations support metacognitive monitoring of decision making. J Neurosci. 2017;37(4):781–9.
54. Desender K, Van Opstal F, Hughes G, Van den Bussche E. The temporal dynamics of metacognition: Dissociating task-related activity from later metacognitive processes. Neuropsychologia. 2016;82:54–64.
55. Yeung N, Summerfield C. Metacognition in human decision-making: confidence and error monitoring. Philos Trans R Soc B Biol Sci. 2012;367(1594):1310–21.
56. Boldt A, Yeung N. Shared neural markers of decision confidence and error detection. J Neurosci. 2015;
57. Scheffers MK, Coles MGH. Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. J Exp Psychol Hum Percept Perform [Internet]. 2000;26(1):141. Available from: https://doi.org/10.1037/0096-1523.26.1.141
58. Jia W, Zhu H, Ni Y, Su J, Xu R, Jia H, et al. Disruptions of frontoparietal control network and default mode network linking the metacognitive deficits with clinical symptoms in schizophrenia. Hum Brain Mapp. 2020;41(6):1445–58.
59. Alkan E, Davies G, Greenwood K, Evans SLH. Brain structural correlates of metacognition in first-episode psychosis. Schizophr Bull. 2020;46(3):552–61.
60. Hester R, Nestor L, Garavan H. Impaired error awareness and anterior cingulate cortex hypoactivity in chronic cannabis users. Neuropsychopharmacology. 2009;34(11):2450–8.
61. Moeller SJ, Fleming SM, Gan G, Zilverstand A, Malaker P, Schneider KE, et al. Metacognitive impairment in active cocaine use disorder is associated with individual differences in brain structure. Eur Neuropsychopharmacol. 2016;26(4):653–62.
62. Haddara N, Rahnev D. The impact of feedback on perceptual decision making and metacognition: Reduction in bias but no change in sensitivity. 2020;
63. Desender K, Boldt A, Yeung N. Subjective confidence predicts information seeking in decision making. Psychol Sci. 2018;29(5):761–78.
64. Donoso M, Collins AGE, Koechlin E. Foundations of human reasoning in the prefrontal cortex. Science (80-). 2014;344(6191):1481–6.
65. Desender K, Donner TH, Verguts T. Dynamic expressions of confidence within an evidence accumulation framework. bioRxiv. 2020;

66. Jones RA. Self-fulfilling prophecies: Social, psychological, and physiological effects of expectancies. Lawrence Erlbaum Associates; 1977.
67. Madon S, Jussim L, Eccles J. In search of the powerful self-fulfilling prophecy. J Pers Soc Psychol. 1997;72(4):791.
68. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev. 1977;84(2):191.
69. Bandura A. Self-efficacy. In: V S Ramachaudran (Ed), Encyclopedia of human behavior. New York: Academic Press.; 1994. p. 71–81.
70. Wells A. Metacognitive therapy for anxiety and depression. Guilford press; 2011.
71. Orth U, Robins RW. Development of self-esteem across the lifespan. 2019;
72. Sowislo JF, Orth U. Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. Psychol Bull. 2013;139(1):213.
73. Orth U, Robins RW, Meier LL, Conger RD. Refining the vulnerability model of low self-esteem and depression: Disentangling the effects of genuine self-esteem and narcissism. J Pers Soc Psychol. 2016;110(1):133.
74. Beck AT. Cognitive models of depression. In: Clinical advances in cognitive psychotherapy: Theory and application. Springer Publishing Company; 2002. p. 29–61.
75. Korn CW, Sharot T, Walter H, Heekeren HR, Dolan RJ. Depression is related to an absence of optimistically biased belief updating about future life events. Psychol Med. 2014;44(3):579–92.
76. Disner SG, Beevers CG, Haigh EAP, Beck AT. Neural mechanisms of the cognitive model of depression. Nat Rev Neurosci. 2011;12(8):467–77.
77. Segal Z V. Appraisal of the self-schema construct in cognitive models of depression. Psychol Bull. 1988;103(2):147.
78. Pessiglione M, Le Bouc R, Vinckier F. When decisions talk: computational phenotyping of motivation disorders. Curr Opin Behav Sci. 2018;22:50–8.
79. Pessiglione M, Vinckier F, Bouret S, Daunizeau J, Le Bouc R. Why not try harder? Computational approach to motivation deficits in neuro-psychiatric diseases. Brain. 2018;141(3):629–50.
80. Rouault M, Dayan P, Fleming SM. Forming global estimates of self-performance from local confidence. Nat Commun. 2019;10(1):1–11.
81. Rouault M, Fleming SM. Formation of global self-beliefs in the human brain. PNAS. 2020;in press.

82. Wittmann MK, Kolling N, Faber NS, Scholl J, Nelissen N, Rushworth MFS. Self-other mergence in the frontal cortex during cooperation and competition. Neuron. 2016;91(2):482–93.
83. Chavez RS, Heatherton TF. Multimodal frontostriatal connectivity underlies individual differences in self-esteem. Soc Cogn Affect Neurosci. 2015;10(3):364–70.
84. Lee ALF, de Gardelle V, Mamassian P. Global visual confidence. 2020;
85. Kiani R, Corthell L, Shadlen MN. Choice certainty is informed by both evidence and decision time. Neuron. 2014;84(6):1329–42.
86. Zacharopoulos G, Binetti N, Walsh V, Kanai R. The effect of self-efficacy on visual discrimination sensitivity. PLoS One. 2014;9(10):e109392.
87. Will G-J, Rutledge RB, Moutoussis M, Dolan RJ. Neural and computational processes underlying dynamic changes in self-esteem. Elife. 2017;6:e28098.
88. Ernst A, Moulin CJA, Souchay C, Mograbi DC, Morris R. Anosognosia and metacognition in Alzheimer's disease: Insights from experimental psychology. 2016;
89. Cosentino S, Metcalfe J, Cary MS, De Leon J, Karlawish J. Memory awareness influences everyday decision making capacity about medication management in Alzheimer's disease. Int J Alzheimer's Dis. 2011;2011.
90. Risko EF, Gilbert SJ. Cognitive offloading. Trends Cogn Sci. 2016;20(9):676–88.
91. Gilbert SJ, Bird A, Carpenter JM, Fleming SM, Sachdeva C, Tsai P-C. Optimal use of reminders: Metacognition, effort, and cognitive offloading. J Exp Psychol Gen. 2020;149(3):501.
92. McEvoy JP. The relationship between insight into psychosis and compliance with medications. Insight Psychos Aware Illn Schizophr Relat Disord. 2004;311:333.
93. Kampman O, Laippala P, Väänänen J, Koivisto E, Kiviniemi P, Kilkku N, et al. Indicators of medication compliance in first-episode psychosis. Psychiatry Res. 2002;110(1):39–48.
94. Cosentino S, Metcalfe J, Butterfield B, Stern Y. Objective metamemory testing captures awareness of deficit in Alzheimer's disease. Cortex. 2007;43(7):1004–19.
95. Chapman S, Colvin LE, Vuorre M, Cocchini G, Metcalfe J, Huey ED, et al. Cross domain self-monitoring in anosognosia for memory loss in Alzheimer's disease. Cortex. 2018;101:221–33.
96. Mazancieux A, Souchay C, Casez O, Moulin CJA. Metacognition and self-

awareness in Multiple Sclerosis. Cortex. 2019;111:238–55.

97. Souchay C, Isingrini M, Gil R. Metamemory monitoring and Parkinson's disease. J Clin Exp Neuropsychol. 2006;28(4):618–30.
98. Bhome R, McWilliams A, Huntley JD, Fleming SM, Howard RJ. Metacognition in functional cognitive disorder-a potential mechanism and treatment target. Cogn Neuropsychiatry. 2019;24(5):311–21.
99. Ball HA, McWhirter L, Ballard C, Bhome R, Blackburn DJ, Edwards MJ, et al. Functional cognitive disorder: dementia's blind spot. Brain. 2020;
100. Bègue I, Blakemore R, Klug J, Cojan Y, Galli S, Berney A, et al. Metacognition of visuomotor decisions in conversion disorder. Neuropsychologia. 2018;114:251–65.
101. Clare L, Marková IS, Roth I, Morris RG. Awareness in Alzheimer's disease and associated dementias: theoretical framework and clinical implications. Aging Ment Health. 2011;15(8):936–44.
102. Hyman SE. Can neuroscience be integrated into the DSM-V? Nat Rev Neurosci [Internet]. 2007;8(9):725–32. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed8&NEWS=N&AN=2007408844
103. Gillan CM, Fineberg NA, Robbins TW. A trans-diagnostic perspective on obsessive-compulsive disorder. Psychol Med. 2017;47(9):1528–48.
104. Fusar-Poli P, Solmi M, Brondino N, Davies C, Chae C, Politi P, et al. Transdiagnostic psychiatry: a systematic review. World Psychiatry. 2019;18(2):192–207.
105. Rouault M, Seow T, Gillan CM, Fleming SM. Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biol Psychiatry. 2018;84(6):443–51.
106. Gillan CM, Kosinski M, Whelan R, Phelps EA, Daw ND. Characterizing a psychiatric symptom dimension related to deficits in goal- directed control. Elife. 2016;5:e11305.
107. Hauser TU, Allen M, Rees G, Dolan RJ, Bullmore ET, Goodyer I, et al. Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. Sci Rep. 2017;7(1):6614.
108. Vaghi MM, Luyckx F, Sule A, Fineberg NA, Robbins TW, De Martino B. Compulsivity Reveals a Novel Dissociation between Action and Confidence.

Neuron. 2017;96(2):348-354.e4.

109. Fineberg NA, Hengartner MP, Bergbaum C, Gale T, Rössler W, Angst J. Lifetime comorbidity of obsessive-compulsive disorder and sub-threshold obsessive-compulsive symptomatology in the community: impact, prevalence, socio-demographic and clinical characteristics. Int J Psychiatry Clin Pract. 2013;17(3):188–96.
110. Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry. 2010;15(1):53.
111. Seow TXF, Gillan CM. Transdiagnostic Phenotyping Reveals a Host of Metacognitive Deficits Implicated in Compulsivity. Sci Rep [Internet]. 2020 Jan 1;10(1):2883. Available from: https://doi.org/10.1038/s41598-020-59646-4
112. Wise T, Dolan RJ. Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. Nat Commun. 2020;11(1):1–13.
113. Patzelt EH, Kool W, Millner AJ, Gershman SJ. The transdiagnostic structure of mental effort avoidance. Sci Rep. 2019;9(1):1–10.
114. Seow TXF, O'Connell R, Gillan CM. Model-based learning deficits in compulsivity are linked to faulty representations of task structure. bioRxiv [Internet]. 2020; Available from: https://doi.org/10.1101/2020.06.11.147447
115. Gillan CM, Kalanthroff E, Evans M, Weingarden HM, Jacoby RJ, Gershkovich M, et al. Comparison of the association between goal-directed planning and self-reported compulsivity vs obsessive-compulsive disorder diagnosis. JAMA Psychiatry. 2019 Oct 9;77(1):77–85.
116. Seow TXF, Benoit E, Dempsey C, Jennings M, Maxwell A, McDonough M, et al. A dimensional investigation of error-related negativity (ERN) and self-reported psychiatric symptoms. Int J Psychophysiol. 2020;158:340–8.
117. Olvet DM, Hajcak G. The error-related negativity (ERN) and psychopathology: Toward an endophenotype. Clin Psychol Rev [Internet]. 2008;28(8):1343–54. Available from: https://doi.org/10.1016/j.cpr.2008.07.003
118. Pasion R, Barbosa F. ERN as a transdiagnostic marker of the internalizing-externalizing spectrum: A dissociable meta-analytic effect. Neurosci Biobehav Rev. 2019;103:133–49.
119. Dalgleish T, Black M, Johnston D, Bevan A. Transdiagnostic approaches to

mental health problems: Current status and future directions. J Consult Clin Psychol. 2020;88(3):179.

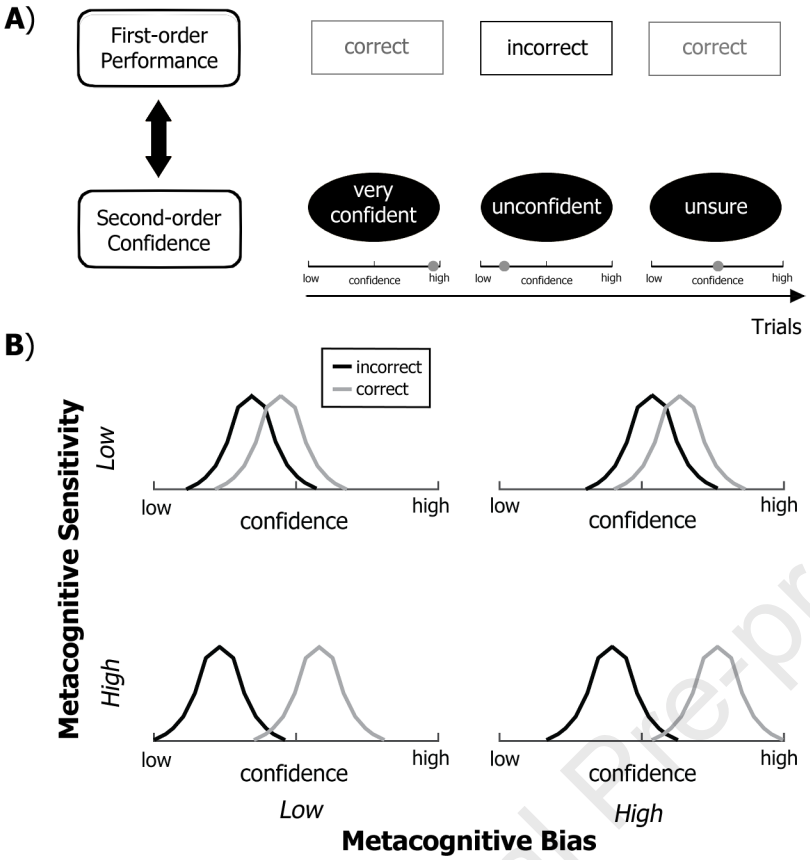
120. Watts AL, Poore HE, Waldman ID. Riskier tests of the validity of the bifactor model of psychopathology. Clin Psychol Sci. 2019;7(6):1285–303.
121. Bornovalova MA, Choate AM, Fatimah H, Petersen KJ, Wiernik BM. Appropriate use of bifactor analysis in psychopathology research: appreciating benefits and limitations. Biol Psychiatry. 2020;
122. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. Biol Psychiatry. 2016;80(7):552–61.
123. Gillan CM, Seow TXF. Carving out new transdiagnostic dimensions for research in mental health. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020;5(10):932–4.
124. Sharp PB, Dolan RJ, Eldar E. Disrupted state transition learning as a computational marker of compulsivity and anxious arousal. 2020;
125. Evensen J, Røssberg JI, Barder H, Haahr U, ten Velden Hegelstad W, Joa I, et al. Apathy in first episode psychosis patients: a ten year longitudinal follow-up study. Schizophr Res. 2012;136(1–3):19–24.
126. Normann N, van Emmerik AAP, Morina N. The efficacy of metacognitive therapy for anxiety and depression: A meta-analytic review. Depress Anxiety. 2014;31(5):402–11.
127. Callesen P, Reeves D, Heal C, Wells A. Metacognitive therapy versus cognitive behaviour therapy in adults with major depression: a parallel single-blind randomised trial. Sci Rep. 2020;10(1):1–10.
128. Clark DA. Focus on “cognition” in cognitive behavior therapy for OCD: Is it really necessary? Cognitive Behaviour Therapy. 2005.
129. Fisher PL, Wells A. Experimental modification of beliefs in obsessive-compulsive disorder: A test of the metacognitive model. Behav Res Ther. 2005;43(6):821–9.
130. Moritz S, Burlon M, Woodward TS. Metacognitive training for schizophrenic patients. Hamburg. Germany: VanHam Campus Verlag; 2005.
131. Moritz S, Woodward TS. Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. Curr Opin Psychiatry. 2007;20(6):619–25.
132. Aghotor J, Pfueller U, Moritz S, Weisbrod M, Roesch-Ely D. Metacognitive training for patients with schizophrenia (MCT): feasibility and preliminary evidence for its

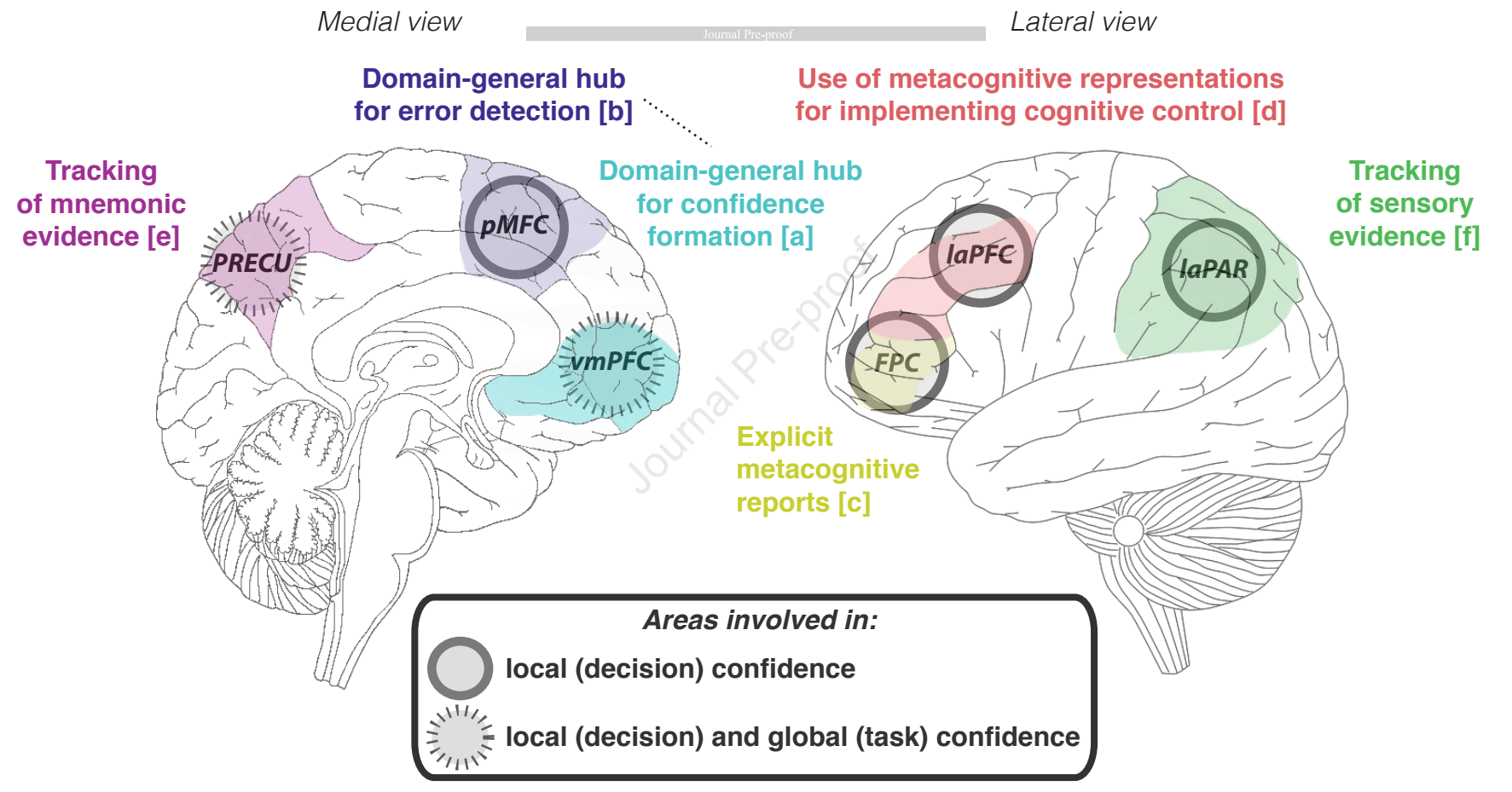
efficacy. J Behav Ther Exp Psychiatry. 2010;41(3):207–11.

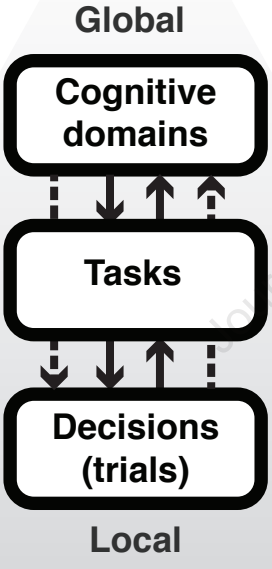
133. Eichner C, Berna F. Acceptance and efficacy of metacognitive training (MCT) on positive symptoms and delusions in patients with schizophrenia: a meta-analysis taking into account important moderators. Schizophr Bull. 2016;42(4):952–62.
134. Miegel F, Demiralay C, Sure A, Moritz S, Hottenrott B, Cludius B, et al. The Metacognitive Training for obsessive-compulsive disorder: A pilot study. Curr Psychol. 2020;1–11.
135. Winter L, Alam M, Heissler HE, Saryyeva A, Milakara D, Jin X, et al. Neurobiological mechanisms of metacognitive therapy–an experimental paradigm. Front Psychol. 2019;10:660.
136. Carpenter J, Sherman MT, Seth AK, Fleming SM, Lau H, Kievit RA, et al. Domain-general enhancements of metacognitive ability through adaptive training. J Exp Psychol Gen. 2019;148(1):51–64.
137. Engeler NC, Gilbert SJ. The effect of metacognitive training on confidence and strategic reminder setting. PLoS One. 2020;15(10):e0240858.
138. Hall MG, Dux PE. Training attenuates the influence of sensory uncertainty on confidence estimation. Attention, Perception, Psychophys. 2020;1–11.
139. Cortese A, Amano K, Koizumi A, Kawato M, Lau H. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nat Commun. 2016;7(1):1–18.
140. Wells A, Fisher P, Myers S, Wheatley J, Patel T, Brewin CR. Metacognitive therapy in treatment-resistant depression: A platform trial. Behav Res Ther. 2012;50(6):367–73.
141. Browning M, Kingslake J, Dourish CT, Goodwin GM, Harmer CJ, Dawson GR. Predicting treatment response to antidepressant medication using early changes in emotional processing. Eur Neuropsychopharmacol. 2019;29(1):66–75.
142. Barlow DH, Farchione TJ, Sauer-Zavala S, Latin HM, Ellard KK, Bullis JR, et al. Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide. Oxford University Press; 2017.
143. Boettcher H, Correa J, Cassiello-Robbins C, Ametaj A, Rosellini AJ, Brown TA, et al. Dimensional Assessment of Emotional Disorder Outcomes in Transdiagnostic Treatment: A Clinical Case Study. Cogn Behav Pract. 2020;27(4):442–53.
144. Sakiris N, Berle D. A systematic review and meta-analysis of the unified protocol as a transdiagnostic emotion regulation based intervention. Clin Psychol Rev.

2019;101751.

145. Lebreton M, Abitbol R, Daunizeau J, Pessiglione M. Automatic integration of confidence in the brain valuation signal. Nat Neurosci. 2015;18(8):1159–67.
146. Taylor SF, Stern ER, Gehring WJ. Neural systems for error monitoring: recent findings and theoretical perspectives. Neurosci. 2007;13(2):160–72.
147. Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S. The role of the medial frontal cortex in cognitive control. Science (80-). 2004;306(5695):443–7.
148. Ullsperger M, Danielmeier C, Jocham G. Neurophysiology of performance monitoring and adaptive behavior. Physiol Rev. 2014;94(1):35–79.
149. Ye Q, Zou F, Lau H, Hu Y, Kwok SC. Causal evidence for mnemonic metacognition in human precuneus. J Neurosci. 2018;38(28):6379–87.
150. Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. Science (80-). 2009;324(5928):759–64.







«I generally have good memory»

«I did pretty well on this exam»

«I got this question correct»

A hierarchy of metacognitive evaluation