Human metacognition across domains: insights from individual differences and neuroimaging

Marion Rouault[1], Andrew McWilliams[1,2,3], Micah G. Allen[1], Stephen M. Fleming[1,4,*]

[1]*Wellcome Centre for Human Neuroimaging, University College London, London, UK*

[2]*Department of Child and Adolescent Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, 16 de Crespigny Park, London SE5 8AF, UK*

[3]*Great Ormond Street Hospital for Children NHS Trust, Great Ormond Street, London WC1N 3JH, UK*

[4]*Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK*

Number of figures: 6
Number of tables: 0

* Correspondence: stephen.fleming@ucl.ac.uk

ABSTRACT

Metacognition is the capacity to evaluate and control one's own cognitive processes. Metacognition operates over a range of cognitive domains, such as perception and memory, but the neurocognitive architecture supporting this ability remains controversial. Is metacognition enabled by a common, domain-general resource that is recruited to evaluate performance on a variety of tasks? Or is metacognition reliant on domain-specific modules? This article reviews recent literature on the domain-generality of human metacognition, drawing on evidence from individual differences and neuroimaging. A meta-analysis of behavioral studies found that perceptual metacognitive ability was correlated across different sensory modalities, but found no correlation between metacognition of perception and memory. However, evidence for domain-generality from behavioral data may suffer from a lack of power to identify correlations across model parameters indexing metacognitive efficiency. Neuroimaging data provide a complementary perspective on the domain-generality of metacognition, revealing co-existence of neural signatures that are common and distinct across tasks. We suggest that such an architecture may be appropriate for "tagging" generic feelings of confidence with domain-specific information, in turn forming the basis for priors about self-ability and modulation of higher-order behavioral control.

Whether a mental process is domain-general (shares resources across many situations or tasks) or domain-specific is a broad question that is pertinent to many areas of psychology. For instance, it has long been debated whether intelligence relies on a single underlying resource (a *g* factor) or on independent components (Chiappe & MacDonald, 2005; Kanazawa, 2004; Kievit et al., 2017). In cognitive neuroscience, Duncan and colleagues have proposed that a "multiple demand" system supports executive functions across many different tasks (Duncan, 2010). In this article we focus on recent research on the domain-generality of neurocognitive substrates supporting metacognition.

Metacognition is defined as cognition about cognition – the ability to reflect on, monitor and control another cognitive process (Dunlosky & Metcalfe, 2008; Nelson & Narens, 1990). In the laboratory, as we will see in more detail below, metacognition can be assessed by recording individuals' judgments of their performance on a particular task, such as their confidence in a decision or a judgment of whether learning will be successful (a "second-order" judgment). Because metacognition is by definition second-order to other cognitive processes, it may operate across multiple "domains" of cognition – for instance, one might engage in metacognition about percepts, about memories, about decisions, and so forth. Progress has been made on understanding the neural basis of metacognition (see Fleming & Dolan, 2012 for a review), which will be considered at more length below. However, it remains poorly understood as to whether metacognition relies on a domain-general resource that is "applied" to the task at hand, or whether different metacognitive processes are engaged when evaluating performance in different domains (Figure 1A). This article reviews and critically appraises progress on this issue.

Measures of metacognition

In order to assess the relationship between metacognition across domains, we require metrics of metacognitive ability that are robust and comparable across tasks. Here we focus on objective measurement of metacognition from behavioral tasks rather than self-report questionnaires. Some second-order judgments are less suitable for cross-domain comparison because they are inherently domain-specific. For instance,

judgments of learning (JOLs) refer directly to the learning process (Arbuckle & Cuddy, 1969), and are therefore not applicable when investigating metacognition of perception. For this reason comparisons of metacognition across domains have tended to focus on retrospective confidence judgments of performance - a judgment of confidence that a previous decision involving internal process X was correct, where X could refer to any cognitive process, such as perceptual discrimination or memory retrieval. Once such judgments have been collected over several trials of a task they can be compared to objective accuracy to build up a picture of an individual's metacognitive ability. In general, metacognition is said to be accurate when correct decisions are held with high confidence and incorrect decisions are held with lower confidence – in other words, metacognitive accuracy refers to the *correlation* between task performance and confidence. The various approaches for characterizing this correlation have been comprehensively reviewed elsewhere (Fleming & Lau, 2014).

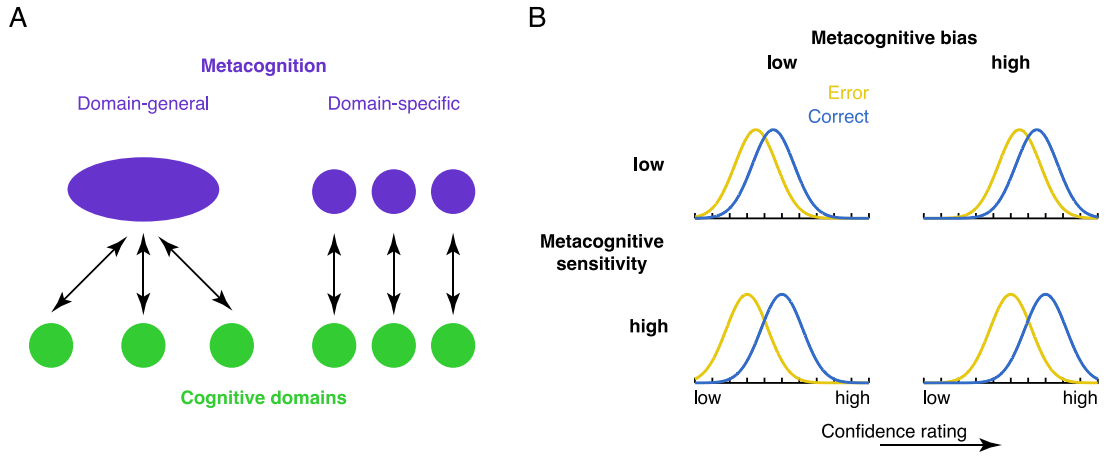

Figure 1. (A) It remains debated whether metacognition operates as a domain-general resource applied over cognitive domains (left) and/or whether metacognition itself relies on domain-specific components that operate over corresponding cognitive domains. (B) Metacognitive bias and metacognitive efficiency are two independent aspects of metacognition. Metacognitive bias corresponds to an overall tendency to rate confidence higher (right panels) or lower (left panels), irrespective of performance. Metacognitive sensitivity quantifies the extent to which correct and error trials can be discriminated (adapted from Fleming & Lau, 2014).

It is useful to distinguish two aspects of metacognitive judgments – their *sensitivity* and *bias*. These are illustrated in the cartoon in Figure 1B. Each panel shows example probability densities of confidence ratings conditional on correct and incorrect task performance. If these distributions are cleanly separated, this implies the subject is

able to recognize accurate from inaccurate performance using the confidence scale and we would describe them as having a high degree of metacognitive *sensitivity*. In contrast, metacognitive *bias* refers to an overall level of confidence averaging over performance. These aspects of metacognition are theoretically independent. For instance, someone who has low overall confidence (low bias) may still be sensitive on a trial-by-trial basis to fluctuations in performance (high sensitivity). By applying a modification of signal detection theory (SDT), known as "type 2" SDT, it is possible to quantify sensitivity and bias of ratings with respect to objective performance (Clarke, Birdsall, & Tanner, 1959; Fleming & Lau, 2014; Galvin, Podd, Drga, & Whitmore, 2003).

However, when making inferences about processes that are shared or distinct across domains, it is important to ensure that estimation of these components of metacognition is not confounded by first-order task performance. For instance, we might find that metacognitive sensitivity is highly correlated across two unrelated tasks, but this correlation would be less interesting if it were simply a consequence of first-order performance also being correlated between tasks. Indeed, several measures of metacognitive sensitivity (such as area under the type 2 Receiver Operating Characteristic curve (AUROC2) and confidence-accuracy correlations) are themselves affected by first-order performance (Galvin et al., 2003; Masson & Rotello, 2009) – the same individual will likely show greater metacognitive sensitivity on an easy task compared to a hard task. If performance is not matched or accounted for between conditions, erroneous conclusions may be drawn, for instance that a patient group has a deficit in metacognition when such a deficit is instead explained by a difference in first-order performance (Figure 2).

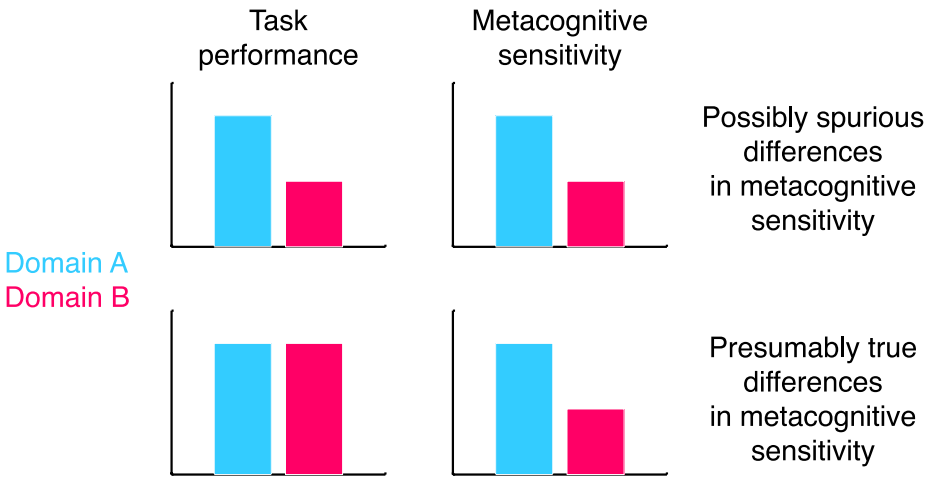

Figure 2. Top panel: Differences in task performance might produce spurious differences in metacognitive sensitivity between groups or task domains. Bottom panel: If task performance is matched between domains, differences in metacognitive sensitivity are likely to reflect true differences in metacognition.

One elegant solution to the problem of controlling for performance confounds is the meta-*d'* framework developed by Maniscalco & Lau (2012). This approach posits a generative model of confidence data within a signal detection theory (SDT) framework. Fitting the model to data returns a parameter, meta-*d'*, that reflects the level of first-order performance (known as *d'*) that would have led to the observed confidence rating data under an ideal observer model. Meta-*d'* can then be compared to actual *d'* (for instance by computing the ratio meta-*d'/d'*) to give a measure of metacognitive *efficiency*, which quantifies the level of metacognitive sensitivity relative to first-order performance. By using metacognitive efficiency as our measure of metacognition we can meaningfully compare scores across individuals or task domains. Alternatively, if simpler measures of metacognitive sensitivity are employed, it is important to ensure that any potential confounds due to differences in first-order performance between conditions are examined and accounted for (Figure 2).

Domain-generality in metacognitive ability (1) – individual differences

A classical approach to studying domain-generality of mental processes is examining patterns of individual differences to ask whether variance is shared or distinct across tasks. Assuming that our metrics are reliable and free of confounds, cross-correlations between domains indicate shared constraints on a particular ability. For instance, if

we find that across individuals, faster choice response times are strongly predictive of IQ scores, we might conclude that greater processing speed contributes to both decision time and intelligence (Ratcliff, Thapar, & McKoon, 2010; Ritchie, Bates, Der, Starr, & Deary, 2013). Correlations of metacognitive measures with other stable individual differences (such as personality or mental health) may also reveal domain-general aspects of metacognition. In this section we review studies that have taken this approach to investigate domain-general and domain-specific aspects of metacognitive efficiency and bias, and provide a formal meta-analysis to quantify behavioral evidence for domain-generality.

In the perceptual domain, metacognitive sensitivity (measured as AUROC2) has been found to be correlated across individuals for contrast and orientation discrimination tasks (Song et al., 2011), despite perceptual thresholds (first-order performance) in each case being uncorrelated. Similar results were found when examining metacognitive efficiency (meta-*d'/d'*) correlations across visual, auditory and tactile modalities (Faivre, Filevich, Solovey, Kühn, & Blanke, 2018). Tactile metacognitive sensitivity (measured using AUROC2) was found to be uncorrelated with metacognitive sensitivity on cardiac and respiratory discrimination tasks, despite the latter correlating with each other (Garfinkel et al., 2016). Ais and colleagues found strong correlations between metacognitive bias (average confidence levels) across several perceptual tasks (auditory, luminance and contrast discrimination tasks and a "partial report" task which required identification of a letter in a briefly flashed array) (Ais, Zylberberg, Barttfeld, & Sigman, 2016). However, they found correlations in metacognitive sensitivity (AUROC2) only between auditory and luminance tasks. In addition this study identified similar confidence "profiles" for a given individual, indicating idiosyncratic and stable patterns of confidence ratings across tasks (Ais et al., 2016).

Although these studies explored inter-relationships between metacognition in different perceptual modalities, it could be argued that all such tasks belong to a broader perceptual domain, but are further in task space from other cognitive domains such as memory. Within the memory domain, metacognitive bias, but not metacognitive sensitivity (assessed by the degree of match between confidence and recall performance), was found to be correlated across face and word recall tasks

(Bornstein & Zickafoose, 1999; Sadeghi, Ekhtiari, Bahrami, & Ahmadabadi, 2017; Thompson & Mason, 1996; West & Stanovich, 1997) and across a variety of judgment-of-learning tasks (Kelemen, Frost, & Weaver, 2000) – albeit in these studies first-order performance was not matched across tasks. More recent studies have compared metacognition for perception and memory while also matching performance: in this case, metacognitive efficiency was correlated between perceptual and memory domains (McCurdy, Maniscalco, Metcalfe, Liu, de Lange & Lau, 2013; Palmer, David, & Fleming, 2014). Samaha and Postle also found evidence of domain-generality in metacognitive sensitivity for perceptual discrimination and visual working memory (measured using performance-confidence correlations and AUROC2), but whether this result reflects generalization beyond perception is unclear because perceptual resources may also be required for short-term memory of visual orientation (Samaha & Postle, 2017). In contrast, other studies found no correlation between metacognitive efficiency across memory and perceptual tasks (Baird, Cieslak, Smallwood, Grafton, & Schooler, 2015; Baird, Smallwood, Gorgolewski, & Margulies, 2013; Morales, Lau, & Fleming, 2017). These mixed findings may be due to differences in metacognition metrics (AUROC2 in Baird et al. vs. meta-*d'*/*d'* in McCurdy et al. and Morales et al.), and/or differences in task requirements (2AFC vs. Yes/No) (Ruby, Giles, & Lau, 2017), as we discuss further below. Finally, no correlation was found between metacognitive sensitivity (measured using AUROC2) on a visual discrimination task and a task involving mentalizing and reasoning (Valk, Bernhardt, Böckler, Kanske, & Singer, 2016) or between metacognitive sensitivity on perception, memory and error awareness tasks (Fitzgerald, Arvaneh, & Dockree, 2017).

The above studies can be clustered into two main groups: those that examine inter-correlations between metacognitive sensitivity in different perceptual discrimination tasks and those studying correlations between metacognition of recognition memory and perception. To assess evidence for domain-generality we conducted a meta-analysis of cross-domain correlation coefficients for these two study categories (Figure 3). From a literature search we identified studies that fell into one of these categories and employed signal-detection theoretic measures of metacognition (M-ratio or AUROC), revealing 12 manuscripts and 19 total independent experiments. Effect size (r), sample size (n), and types of cross-domain correlation (memory-

perception vs. perception-perception) were hand coded. In cases where multiple modalities were probed within a sample (e.g., audition, touch, and vision), the 'multi-modal' r-value was calculated as the average R across modalities. Meta-analytic effect sizes (cross-domain r) were calculated using a Fisher R-to-Z random effects model, implemented in the metafor R-package, version 3.3.2 (Viechtbauer, 2010). Specifically, we performed three meta-analyses – one of overall cross-domain correlations (n=19), one on cross-domain correlations within the perceptual modality (n=9), and another of memory-perception cross-domain correlations (n=10).

We found that across all studies, cross-domain correlations were significantly greater than zero (meta-analytic r = 0.27, 95% CI = [0.13, 0.41], $p < 0.001$) and exhibited significant heterogeneity across effect sizes (Q = 50.46, df = 18, $p < 0.001$, I^2 = 69.7%). This result was primarily driven by medium to strong cross-perceptual correlations (meta-analytic r = 0.55, 95% CI = [0.34, 0.76], $p < 0.001$). Restricting our analysis to perceptual effect sizes, we did not observe significant heterogeneity (Q = 14.49, df = 8, $p = 0.07$, I^2 = 45.4%). In contrast, cross-domain correlations between memory and perception-based tasks were not significant (r = 0.09, 95% CI = [-0.02, 0.21], $p = 0.10$), and did not show effect heterogeneity (Q = 14.76, df = 9, $p = 0.09$, i^2 = 38.3%). These results suggest that cross-task correlations in metacognitive ability are primarily obtained when examining tasks tapping into the same functional modality, i.e. perception (Figure 3).

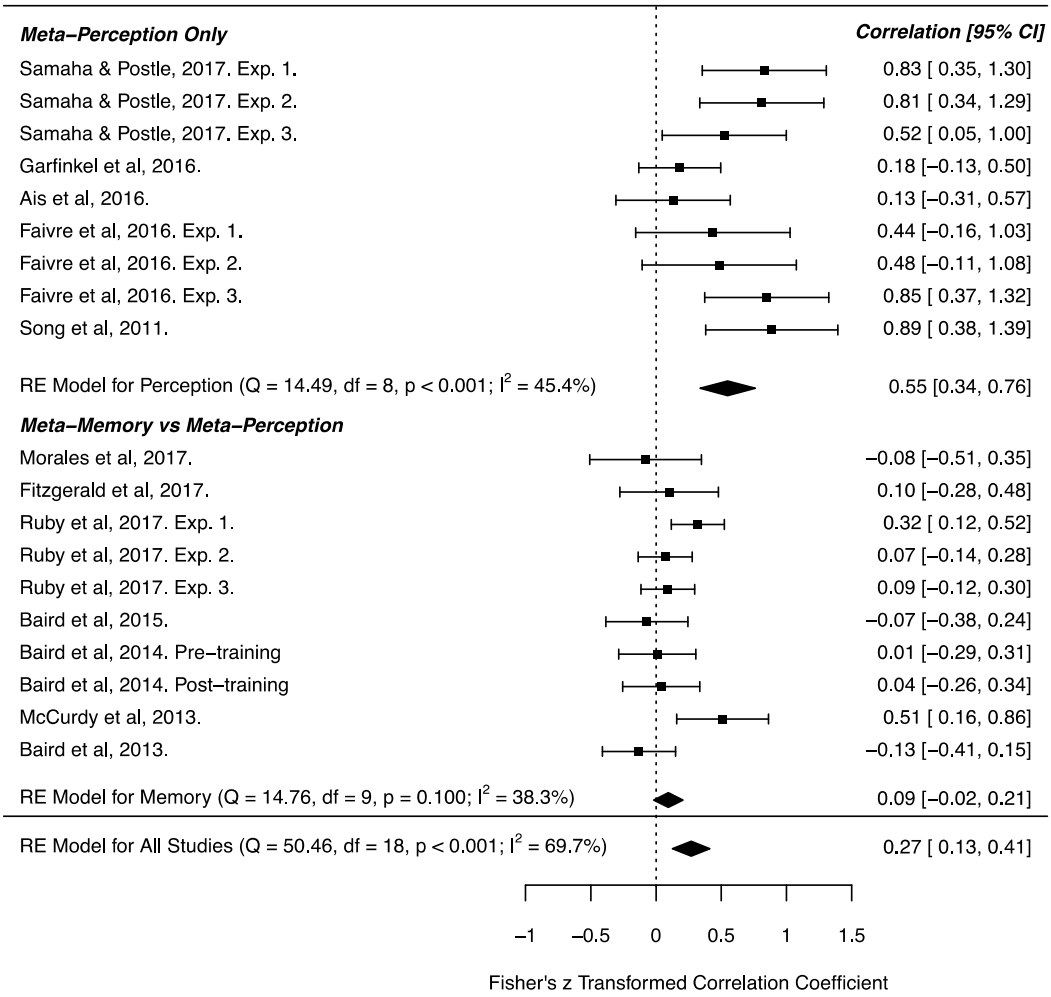

Figure 3. Forest plot for our three meta-analyses examining the meta-analytic strength of cross-domain correlations of metacognition. The first focused on studies in the perceptual domain (e.g. visual, tactile). The second examined the cross-domain correlation of metacognition in perceptual vs. memory-based tasks, and the third estimated the overall meta-analytic cross-domain correlation across all studies. The results show that metacognitive ability is primarily preserved across perceptual tasks, but does not generalize to memory-based tasks. The right column indicates the Fisher's z transformed correlation coefficient.

Such results may initially appear to support a conclusion that memory and perceptual metacognition rely on largely separate, domain-specific processes. However one potential caveat is that some studies have compared yes-no (Y/N) tasks with 2-alternative forced choice tasks (2AFC). In 2AFC tasks, a pair of stimuli is presented, for instance reporting which of two intervals contains a brighter stimulus. In Y/N tasks, a single stimulus is presented which must be classified as a target or lure. Metacognition for Y/N and 2AFC tasks may appear different not because of a true difference between domains, but because of a difference in the processes that generate confidence ratings in the two cases (Ruby et al., 2017). Specifically, previous studies

have documented that meta-*d'* following "no" responses in a Y/N task is substantially lower than meta-*d'* following equivalent "yes" responses (Meuwese, van Loon, Lamme, & Fahrenfort, 2014; Kanai et al., 2010), potentially obscuring a latent domain-general component (Ruby et al., 2017). In addition, an absence of correlation may result from a lack of statistical power rather than a true null effect, particularly given the small sample sizes often employed in previous studies. Future studies could profitably employ Bayesian statistics to directly assess evidence in favor of the null hypothesis when examining cross-domain correlations.

Another candidate explanation for discrepant findings when studying across-domain correlations in metacognitive efficiency corresponds to the variety of metrics employed to assess metacognitive sensitivity (correlations, AUROC2, meta-*d'*), and the reliability of within-subject measures of metacognition. Metacognitive sensitivity is itself a measure of association between two variables (performance and confidence) that requires several trials to be estimated with sufficient stability and therefore there is inevitable uncertainty in the estimation of within-domain parameters (Fleming, 2017). This within-subject uncertainty is rarely taken into account in analyses of individual differences (although see Samaha & Postle, 2017), which typically rely on point estimates such as AUROC2 or maximum likelihood estimates of meta-*d'*. Recently we have developed a Bayesian framework (HMeta-d) for estimating meta-*d'* both at the level of individual subjects and groups of subjects (Fleming, 2017). One advantage of this framework for analyses of domain-generality is that it can be extended to estimate correlation coefficients between domains. Unlike classic point-estimate approaches, this ensures that uncertainty in individual metacognitive efficiency estimates appropriately propagates through to uncertainty around the cross-domain correlation coefficient.

This effect can be appreciated in simulations plotted in Figure 4 (code available at https://github.com/metacoglab/RouaultDomainReview). Here we generated confidence rating data from N=100 simulated subjects across two "domains". The group metacognitive efficiency was set to 0.8 in both domains, and individual subject meta-*d'/d'* values sampled from a bivariate Gaussian distribution with a true correlation in metacognitive efficiency between domains of 0.5. We sampled confidence rating counts for known meta-*d'/d'* values using the metad_sim function

from the HMeta-d toolbox (https://github.com/metacoglab/HMeta-d), keeping confidence rating criteria fixed across domains and subjects. The number of trials per subject differed between simulations (50 vs. 400). The model outputs a posterior belief distribution over the across-domain correlation coefficient. It can be seen that as the number of trials per subject increases (i.e. the certainty associated with individual meta-*d'* estimates goes up, lower panel in Figure 4), we can be more certain about the presence of a domain-general correlation (narrower posterior density). We recommend applying such multi-level models when analyzing individual difference correlations to ensure this parameter uncertainty is appropriately taken into account.

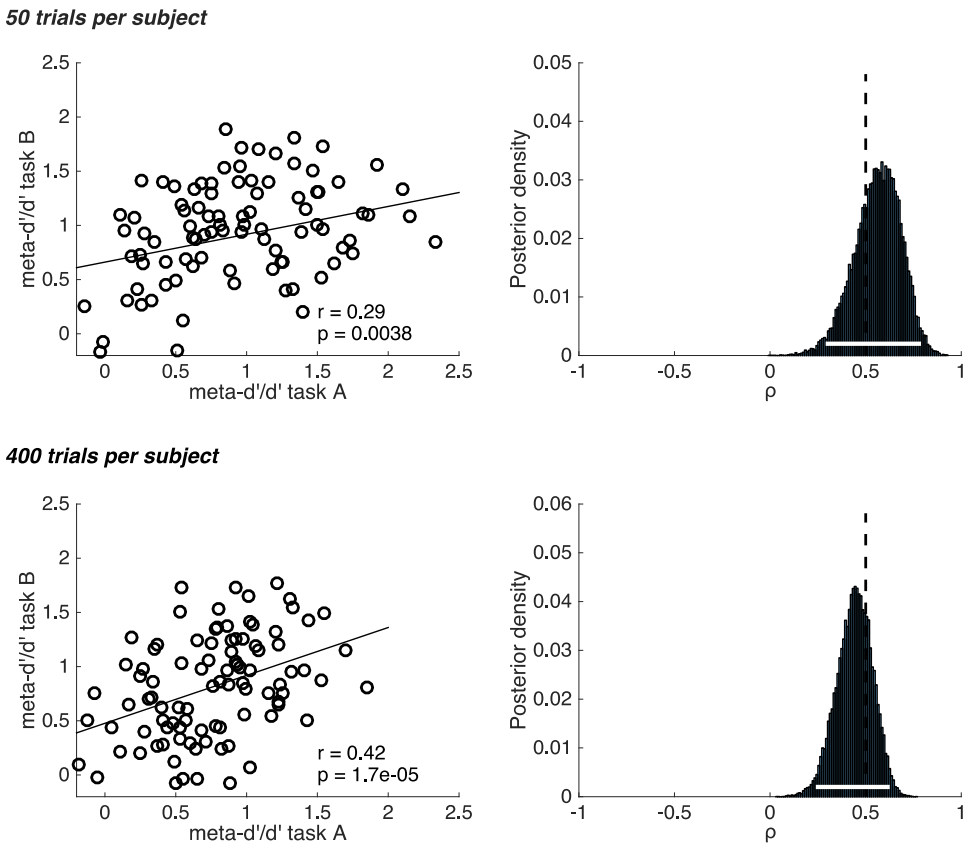

Figure 4. Simulations of hierarchical meta-*d'* model (HMeta-d) estimation of the covariance between metacognitive efficiencies for 100 simulated subjects with an average meta-*d'/d'* ratio of 0.8. Upper panels correspond to 50 trials per subject, lower panels to 400 trials per subject. The "ground truth" correlation coefficient in both cases was 0.5, and in both cases we recovered a significant correlation between point estimates obtained using single-subject maximum likelihood. Notably, the posterior over the correlation coefficient is narrower around the true value (shown by the dotted vertical line) when there are more trials per subject (lower row), reflecting increased certainty in subject-level meta-*d'* parameter estimation.

Other studies have investigated relationships between metacognition and other aspects of personality and executive function which, if found, would lend support to a domain-general metacognitive contribution to metacognition. For instance, factors such as general intelligence or motivation to engage with a task could affect metacognition over multiple domains. However, if such general factors are significantly altered, it is unlikely that metacognitive processes would be selectively affected while also leaving first-order performance spared; for instance, an altered ability to follow task instructions following a prefrontal cortex lesion is likely to affect both task performance and metacognitive evaluation. The integrity of these "global" factors can thus be seen as a necessary but not sufficient condition for enabling metacognition. Notably however, despite both relying on aspects of higher cognition, we have found that over several datasets perceptual metacognitive efficiency is not related to measures of fluid intelligence (Fleming, Huijgen, & Dolan, 2012; Palmer et al., 2014), even when such correlations were examined in a large-scale dataset of ~1000 individuals (Rouault, Seow, Gillan and Fleming, 2018). Such independence may be due to fluid intelligence relying on posterolateral frontal and parietal "multiple demand" regions (Woolgar et al., 2010), whereas metacognition has been linked to anterior prefrontal regions, as considered in more detail below.

In summary, analyses of individual differences in metacognitive efficiency indicate the presence of domain-general contributions to confidence judgments across distinct perceptual discrimination tasks. Such variation in metacognition is isolated from variation in first-order performance. However it remains unclear whether a shared resource supports metacognitive efficiency across more distant domains, such as recognition memory and perceptual discrimination. One important consideration in conducting such cross-domain correlation analyses is to ensure that uncertainty in estimation of metacognition *within* a particular domain is appropriately propagated to the analysis of *between*-domain correlations, which is now possible within hierarchical Bayesian frameworks. Taken together these findings also raise the issue of how to define a separation between domains, and whether the notion of domain should instead be considered as existing along a continuum or gradient (see "Computational processes" section below). Furthermore we should remain mindful that the architecture of metacognition (and therefore any shared variance between different tasks) may well be organized along different lines than the cognitive

processes being monitored, and which are typically compared in the laboratory (e.g. perception, memory).

Domain-generality in metacognitive ability (2) – neuropsychology

The study of individual differences identifies shared variance in behavioral performance across a large number of healthy individuals. In contrast, neuropsychology seeks to identify dissociations between abilities induced by patterns of brain damage. Classic studies by Shimamura and colleagues revealed that metamemory abilities (such as feeling-of-knowing or judgments of learning) are selectively impaired following frontal lesions (Janowsky, Shimamura, & Squire, 1989; Shimamura & Squire, 1986). Metamemory evaluation has itself been divided into distinct judgment types (see Chua et al., 2014, for a review). A key distinction is that judgments can be either prospective, occurring prior to memory retrieval, or retrospective. Prospective judgments include feeling of knowing (FOK), the likelihood of recognizing an item that currently cannot be recalled, and judgment of learning (JOL), a belief during learning about the success of subsequent recall. More recent studies indicated such lesion deficits may not apply to all forms of metamemory judgment. For instance, two independent studies found that damage to the medial prefrontal cortex was associated with decreased prospective feeling-of-knowing accuracy but intact retrospective confidence judgments (Schnyer et al., 2004) and judgments of learning (Modirrousta & Fellows, 2008). The reverse dissociation was reported by Pannu and Kaszniak, who found that deficits in retrospective confidence judgments were associated with lateral frontal lesions (Pannu & Kaszniak, 2005).

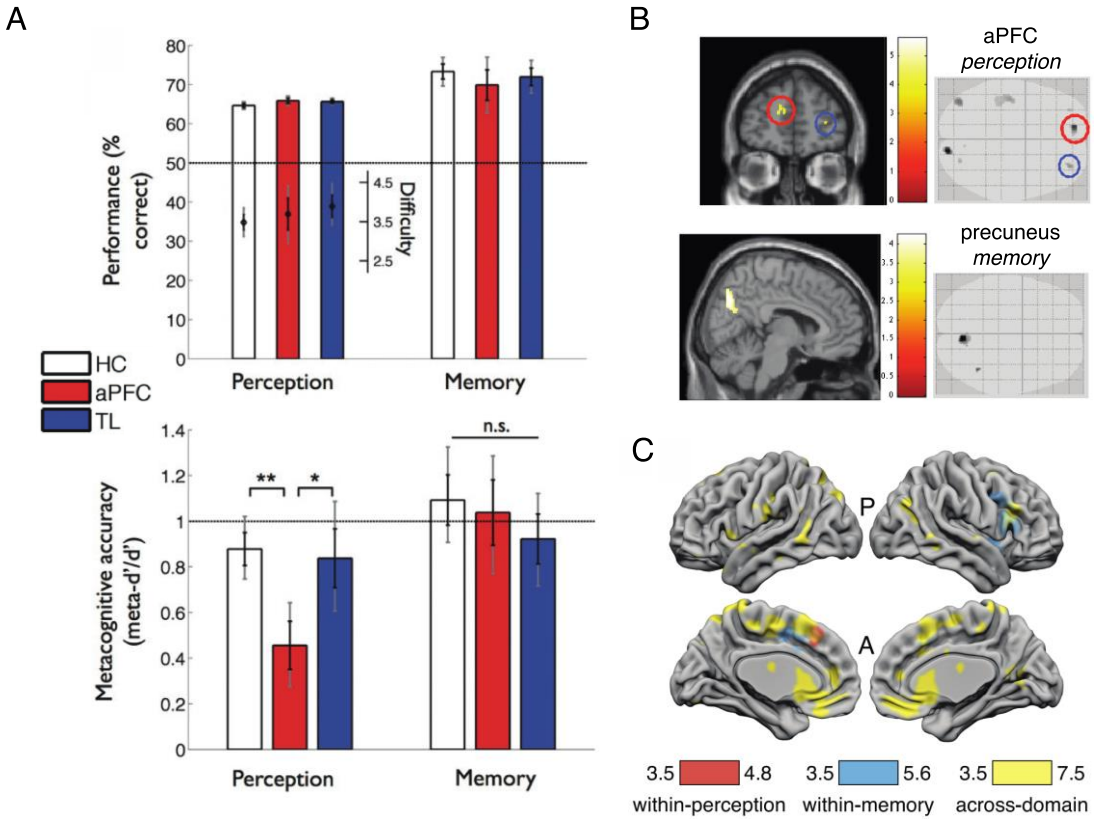

Figure 5. Different methodologies for quantifying brain structure and function shed light on the underpinnings of metacognition across domains. (A) Human subjects with anterior PFC lesions (aPFC) were found to have reduced metacognitive efficiency on a perceptual but not a memory task (lower panel), compared to temporal lobe lesion patients (TL) and healthy controls (HC), despite matched performance and task difficulty (upper panel; reproduced from Fleming et al., 2014). (B) Individual differences in metacognitive efficiency for perception were found to correlate with aPFC gray matter volume, whereas individual differences in metacognitive efficiency for memory were found to correlate with medial parietal cortex (precuneus) gray matter volume. Structural variation in each of these regions was in turn positively correlated across participants, translating into a behavioral correlation of metacognitive efficiencies between domains (reproduced from McCurdy et al., 2013). (C) Multivariate analyses of human neuroimaging data revealed widespread classification of confidence level in dACC/pre-SMA, vmPFC and striatum that generalized across domains (yellow). In contrast, domain-specific patterns of confidence-related activity were identified in right lateral aPFC (ROI analysis not shown; reproduced from Morales et al., 2018).

Fewer studies have taken a neuropsychological approach to ask whether metacognitive deficits are shared across multiple task domains. Fleming et al. studied three groups of subjects matched for age and IQ – a healthy control group, a group with anterior prefrontal cortex (aPFC) lesions, and a group with temporal lobe lesions (Fleming, Ryu, Golfinos, & Blackmon, 2014). Each participant completed both a

recognition memory task with word stimuli and a perceptual discrimination task about the relative density of two dot patches. In both tasks retrospective confidence ratings were elicited on a trial-by-trial basis, allowing assessment of meta-*d'* for each subject/task domain. A selective deficit in metacognitive efficiency (meta-*d'/d'*) for perceptual discrimination was observed in the aPFC group (Figure 5A) despite equivalent first-order performance and metacognitive bias. Such a result is consistent with a contribution of aPFC to metacognition of perceptual decision-making (Allen et al., 2017; Fleming, Weil, Nagy, Dolan, & Rees, 2010; Yokoyama et al., 2010), but suggests other brain regions may be sufficient to support intact metacognition of recognition memory (Baird et al., 2013; McCurdy et al., 2013). Notably, all patients were tested in a post-acute phase of their lesion, so it is possible that an early domain-general deficit may have been observed sooner after surgery. More generally, neuroplasticity and reorganization following lesions make it difficult to draw strong conclusions about the typical functional anatomy of metacognition from lesion studies alone (Lemaitre, Herbet, Duffau, & Lafargue, 2017).

Domain-generality in metacognitive ability (3) – neuroimaging

Behavioral and neuropsychological data can inform on whether a mental process relies on a shared resource, but provide less insight into the mechanisms that underpin this resource. As noted above, a common resource may be involved but not be detected due to domain-specific unreliability in the measurement of metacognition. Conversely, a domain-general pattern may be driven by a third factor that affects domain-specific processes in equal measure, such as stress (Reyes, Silva, Jaramillo, Rehbein, & Sackur, 2015) or fatigue (Maniscalco, McCurdy, Odegaard, & Lau, 2017).

Several recent studies have focused on the neural basis of human metacognition, which have been reviewed at length elsewhere (Fleming & Dolan, 2012). Briefly, in concordance with the neuropsychological literature highlighted above, anatomical (Allen et al., 2017; Fleming et al., 2010; McCurdy et al., 2013) and functional (Baird et al., 2013; Cortese, Amano, Koizumi, Kawato, & Lau, 2016; De Martino, Fleming, Garrett, & Dolan, 2013; Fleck, 2006; Fleming et al., 2012; Hilgenstock, Weiss, & Witte, 2014; Yokoyama et al., 2010) neuroimaging data indicate that a frontoparietal

network contributes to metacognitive estimates of task performance across a range of tasks. Within this network, electrophysiological studies in humans have focused on the role of the posterior medial frontal cortex (pMFC) and associated error-related negativity in performance monitoring (Gehring et al., 1993; Dehaene et al., 1994). More recently, anterior PFC has also been implicated in supporting explicit metacognitive judgments, leading Fleming and Dolan to propose that connectivity between interoceptive cortices (cingulate and insula) and anterior PFC may underpin the fidelity of explicit metacognition (Fleming & Dolan, 2012). Non-human electrophysiological work has also identified a key role for frontoparietal areas in confidence formation. In particular, recordings in monkey lateral intraparietal cortex (LIP) indicate that variability in LIP firing rates is predictive of both decisions and decision confidence (Kiani and Shadlen, 2009; Hanks et al., 2011). Furthermore, activity in rat orbitofrontal cortex carries signals related to decision confidence in a perceptual discrimination task (Kepecs et al., 2008). However, the distinct computational roles of these regions, and whether such neural substrates of confidence are shared or distinct across tasks remains unclear. In what follows we selectively focus on studies which directly compare neural correlates of metacognition across task domains using neuroimaging techniques in humans.

An intriguing example of a domain-general pattern in behavior that could be explained by domain-specific neural resources was reported by McCurdy et al. (2013). In this study, the same participants carried out 2AFC perceptual discrimination (Gabor contrast discrimination) and recognition memory judgments together with confidence ratings. As noted above, this study obtained behavioral evidence for a domain-general correlation between metacognitive abilities across the two domains. However, each participant also underwent a structural MRI to enable analysis of individual variation in grey matter volume across the cortex. It was found that metacognitive efficiency (meta-*d'/d'*) for perception correlated with grey matter volume in the anterior prefrontal cortex (Figure 5B; see also Allen et al., 2017; Fleming et al., 2010), whereas meta-*d'/d'* on the memory task correlated with grey matter volume in the precuneus. Using structural equation modeling, the best model of the data was one in which the structure of two domain-specific regions was correlated across individuals, thereby explaining a domain-general finding in behavior via the coupling of two domain-specific resources. This example shows how

neuroimaging data can shed additional light on the cognitive architectures that determine individual differences in metacognition.

Other studies combining analyses of individual differences with structural and diffusion imaging measurements have also provided evidence for the involvement of distinct neural structures in metacognition across domains. Metacognitive accuracy on a visual task was shown to correlate with white matter microstructure underlying the ACC, whereas metacognitive accuracy for a memory task correlated with white matter underlying the inferior parietal lobule (IPL) (Baird et al., 2015). In analysis of resting state fMRI data, connectivity between ACC and anterior PFC was related to more accurate perceptual metacognition judgments, whereas increased connectivity between precuneus, IPL and anterior PFC predicted better metamemory (Baird et al., 2013). Furthermore, cortical thickness mapping revealed domain-specific substrates structurally related to metacognition of perception (right medial PFC) vs. mentalizing (bilateral PFC, temporo-parietal cortex, posterior medial parietal cortex) (Valk et al., 2016).

In perceptual decision-making tasks, classical univariate analyses of fMRI BOLD signal reveal a negative parametric relationship between confidence reports and activity in posterior medial frontal cortex (pMFC, encompassing dorsal anterior cingulate cortex and pre-supplementary motor area) (Fleck, Daselaar, Dobbins, & Cabeza, 2006; Heereman, Walter, & Heekeren, 2015; Morales et al., 2018), which are also observed in the memory domain (episodic retrieval) (Fleck et al., 2006). In pMFC and vmPFC, multivariate fMRI analyses revealed that it was possible to predict confidence in a memory task from patterns decoded in a perceptual task matched for stimulus and task requirements, and vice-versa (Figure 5C), suggesting that confidence covaries with task-independent neural representations (Morales et al., 2018). In contrast, right lateral aPFC instead showed significant decoding effects within- but not across-domain. This domain-specific neural representation of confidence suggests that lateral aPFC may “tag” metacognitive representations with task-specific information, which could be particularly relevant for future meta-level control decisions such as which task to engage in next.

We note that some observations of domain-generality in neural data may be due not only to confidence but other correlated variables, for instance decision time, which has been proposed as a relevant input for a confidence computation (Kiani, Corthell, & Shadlen, 2014), and expected value. Some studies of confidence explicitly modeled reaction times in their fMRI analysis (Fleck et al., 2006; Gherman et al., 2017; Lebreton et al., 2015), whereas others did not (Heereman et al., 2015; Morales et al., 2018), and it is not straightforward to determine whether decision time should be treated as a confound or a relevant variable of interest for studies of the neural basis of metacognition. Notably, regions often implicated in encoding expected value such as ventral striatum and vmPFC (Clithero and Rangel, 2013) are also often found to scale with confidence in perception, value and memory domains (De Martino et al., 2013; Gherman & Philiastides, 2017; Lebreton, Abitbol, Daunizeau, & Pessiglione, 2015; Morales et al., 2018), suggesting that being confident is valuable, and/or that when highly confident, subjects expect imminent reward. The fact that a majority of studies have not dissociated confidence from implicit expected value potentially explains these pervasive, domain-general activations. Future studies are required to directly investigate a putative commonality of confidence and value representations, and to separate the component inputs to confidence formation (Bang & Fleming, 2018).

Previous studies have suggested that involvement of the precuneus (medial parietal cortex) is specific to metamemory judgments. Indeed individual metacognitive efficiency in a memory task, but not in a perceptual task, was found to correlate with gray matter volume in the precuneus (McCurdy et al., 2013), and resting-state functional connectivity revealed that better metamemory was associated with increased connectivity between medial aPFC and precuneus (Baird et al., 2013). In addition, TMS application over precuneus impaired metacognitive efficiency for memory but not perception, both measured as meta-*d'* - *d'* (Ye et al., 2018), and univariate fMRI activation in precuneus was selectively increased during metacognitive judgments of memory, but not perception (Morales et al., 2018). However, a relationship between precuneus grey matter volume and metacognitive efficiency has also been detected in a perceptual decision-making task, albeit at an uncorrected whole-brain threshold (Fleming et al., 2010). Moreover, the same region was found to correlate negatively with confidence level in a visual motion-

discrimination task (Heereman et al., 2015). Lastly, using multivariate decoding of fMRI activity, Morales et al. found that classification of high vs. low confidence trials in precuneus generalized across memory and perception domains (Morales et al., 2018). Together these results suggest that precuneus involvement may not be specific to metamemory. Interestingly, recent results suggest that the vividness of episodic memory was tracked by precuneus activity over and above memory precision and retrieval success (Richter, Cooper, Bays, & Simons, 2016). To the extent to which similar appraisals of vividness feed into the formation of perceptual confidence, this may explain the domain-general nature of findings in this region. An interesting alternative possibility is that precuneus is engaged when subjects leverage prior beliefs about self-ability to compute confidence in perception, hence needing to retrieve global beliefs about past experience from memory (see also Figure 6).

In summary, neuroimaging studies indicate a more nuanced picture than studies of behavior or individual differences, which have tended to argue for either domain-specific or domain-general aspects of metacognition. It is possible to reconcile these perspectives by demonstrating that both domain-specific and domain-general signals co-exist in the human brain, and that there may exist a gradient in which some tasks (such as different types of perceptual judgment) are more likely to rely on shared circuitry for metacognitive evaluation than others. In the next section we attempt to formalize these ideas through the lens of computational modeling.

Computational processes supporting metacognition across domains

Models of confidence formation

The simplest first-order models of confidence formation (such as signal detection theory) assume that the internal states supporting decisions and confidence estimates are identical. Such frameworks predict that any covariation between metacognitive efficiency across domains should be accompanied by covariation in lower-level performance, and struggle to accommodate the evidence reviewed above that confidence can be selectively altered or impaired independently of task performance (see also Cortese et al., 2016; Lak et al., 2014; Rounis, Maniscalco, Rothwell,

Passingham, & Lau, 2010). An extension to first-order models of confidence introduces post-decisional processing, thus explaining additional variability in confidence estimates without altering the fidelity of first-order performance (Navajas et al., 2017; Pleskac & Busemeyer, 2010; van den Berg et al., 2016). A somewhat more elaborate but flexible model is a "second-order" computation of confidence (Fleming & Daw, 2017). In the second-order framework, the efficacy of actions is monitored based on higher-order knowledge of the reliability of the decision-making system.

A second-order account provides a natural perspective on findings of domain-generality in metacognition. Similar circuits for second-order inference may be engaged across different domains to the extent to which their lower-level states and actions are similar (the "inputs" and "outputs"; see Figure 6). To take a concrete example, suppose that we are comparing metacognitive efficiency for a visual and auditory discrimination task, both requiring a right-handed button press to indicate the first-order judgment. While in each task, "state" estimation may depend on distinct (visual and auditory) neural circuitry; actions are supported by a common output (the left motor cortex). This commonality in response mapping may be sufficient to induce commonalities in second-order inference, leading to the observations of domain-general confidence signals in neuroimaging data. Such a pattern was observed by Faivre and colleagues, who suggest, "the supramodality of metacognition relies on supramodal confidence estimates and decisional signals that are shared across sensory modalities" (Faivre et al., 2018).

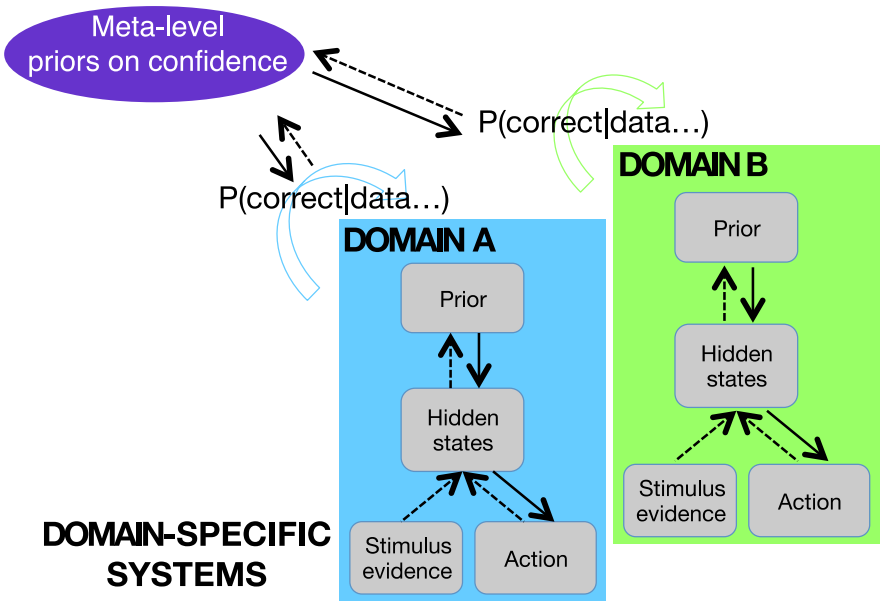

Figure 6. Theoretical framework for metacognition, grounded in models of sensory systems. The two boxes represent domains-specific computations solving two different tasks such as visual and auditory discrimination. Decision-making proceeds in a domain-specific fashion following the principles of Bayesian inference, while a metacognitive layer computes confidence (= P(correct|data)) in each task. Metacognitive inference is itself under the control of priors that may be updated based on previous experience.

However, current computational models of confidence are relatively narrow in scope, focusing on experimentally controllable states and actions. To the extent that other internal variables covary with expected success, these may also become relevant "inputs" for subjective confidence. For instance, response times can provide a proxy for decision time, and be subsequently available to the agent for computing confidence (Benjamin, Bjork, & Schwartz, 1998; Kiani et al., 2014). However for response times to be informative, one needs to have an estimation of the expected level of performance on the task, for instance in the form of a running average over decision accuracy in a given experimental condition. If expected performance is variable within a task, or if it varies significantly across tasks, response times might be less useful as proxies for inferring decision accuracy and hence confidence. For example, in situations where noise in the stimulus induces a deviation from the expected performance level, the expected mapping between confidence and decision time may break down (Rahnev, Maniscalco, Luber, Lau & Lisanby, 2011; Fetsch, Kiani, Newsome & Shadlen, 2014; Zylberberg, Fetsch & Shadlen, 2016; Peters et al., 2017). Similarly, interoceptive states may provide additional proxies when evaluating

self-performance (Allen et al., 2016; Chua & Bliss-Moreau, 2016). This perspective on the formation of confidence in decision-making converges with established "inferential" or cue-based models of confidence formation in the metamemory literature, which suggest that cues such as accessibility (the degree of partial knowledge about the target) contribute to confidence estimates (Koriat and Levy-Sadot, 2001). Similarly, familiarity with the stimuli (De Martino et al., 2013), and volatility/variability in stimulus evidence (Zylberberg et al., 2016; Meyniel, Schlunegger & Dehaene, 2015) also inform confidence. To the extent that some cues, such as response time or fluency, are useful across many different tasks, they may provide domain-general inputs to confidence and metacognition (Alter and Oppenheimer, 2009; Boldt, de Gardelle & Yeung, 2017).

How does confidence guide behavior?

If confidence constitutes a proxy for the probability of success in a task (Peirce and Jastrow, 1884; Pouget, Drugowitsch, & Kepecs, 2016), it may act as a "common currency" signal for estimating and comparing the relative likelihood of success between different tasks (de Gardelle & Mamassian, 2014). Such a common currency would be particularly useful when deciding on which tasks or goals to pursue in the future, especially when external feedback is unavailable. For instance when choosing a career, it would be advantageous to internally evaluate and compare our performance or skill in different potential jobs. The existence of a common currency for confidence is supported by studies showing that subjects are able to compare confidence across visual and auditory tasks with the same precision as when comparing two trials within the same task (de Gardelle & Mamassian, 2014; de Gardelle, Le Corre, & Mamassian, 2016).

The notion that confidence may be compared between task domains to facilitate flexible decision making is consistent with the existence of both domain-general and domain-specific signatures of confidence in neuroimaging data reviewed above. A midline network with hubs in pMFC and vmPFC may provide a nexus for monitoring arbitrary task-sets. In contrast, task-specific confidence representations in lateral aPFC may allow the hierarchical control of decision-making in situations in which subjects need to regularly switch between tasks or strategies on the basis of their

reliability (Donoso, Collins, & Koechlin, 2014; Morales et al., 2018). However we note that the ability to compare confidence between domains does not itself imply that the representation of confidence at the neural level is domain-general, and more investigation is needed into neural signals that support such cross-task comparisons.

Another intriguing possibility is that the existence of midline hubs for confidence formation leads to domain-general representations of confidence which in turn act as priors on confidence in other domains. In behavioral studies, a confidence "leak" has been identified between a color and a symbol discrimination task, where confidence in one task influences confidence in the other regardless of actual performance. Notably, the ability to resist such leakage was positively correlated with lateral aPFC gray matter volume across subjects (Rahnev, Koizumi, McCurdy, D'Esposito, & Lau, 2015). It is important however to distinguish the notion of *confidence leak* (the temporal autocorrelation of confidence ratings when several tasks are interleaved) from *metacognitive bias* (the overall tendency to rate confidence higher or lower irrespective of performance). A larger confidence leak across trials is not necessarily linked to a higher or a lower metacognitive bias, but confidence leak may represent a mediating factor in explaining why metacognitive bias often generalizes across tasks (Baird et al., 2013; Ais et al., 2016). In addition, if confidence leak is large, confidence may become more loosely coupled to current performance in tasks in which autocorrelations in stimuli are absent, which could in turn decrease metacognitive efficiency. However, it remains to be explored whether confidence leak extends across task domains beyond perception. It could be that some domains are more susceptible to leak than others.

Ultimately, to understand the structure of metacognition across task domains we should aim to understand what functions metacognition provides to the system (the computational level in Marr's taxonomy). The formation of accurate beliefs about performance is useful for learning, cognitive control and for social interaction (such as when communicating confidence to others) (Bang et al., 2017; Donoso et al., 2014). But when is it advantageous to share this computation across modalities or inputs? We can think of two possible reasons. Firstly, as stated above, if tasks are composed of arbitrary state-action mappings then it may be more computationally efficient to infer performance in a global, task-independent frame of reference, for use

in control of future behavior (Donoso et al., 2014). Secondly, inferences about performance in one domain may be useful as priors for performance in other domains to the extent they share instrumental characteristics. For instance, if I infer that I am very skilled at skiing, I might infer that I would also be good at similar sports such as ice-skating, but, on this basis, it would not be wise to also think I will be able to remember my to-do list. In other words, it may be useful to generalize confidence level across tasks according to their distance in task space. Conversely, overgeneralization might be maladaptive, and the extent of metacognitive generalization may itself constitute a stable individual difference. For example, people with depression tend to generalize more strongly from poor performance in one domain to other domains, in turn reinforcing a lower level of self-esteem and poorer self-efficacy that cuts across various areas of life (Bandura, 1977; Elliott et al., 1996; Stephan et al., 2016).

Implications of domain-specific alterations in metacognition for clinical populations

The study of metacognition provides an experimental window into our subjective estimates of our internal states. The explanatory potential of metacognition for mechanisms of pathogenesis and maintenance of mental illness is therefore considerable, and metacognitive deficits might be usefully measured in the clinic to guide assessment and management (Wells et al., 2012). Dissecting computational mechanisms supporting metacognitive evaluation could permit development of behavioral and neural interventions to modulate and restore more accurate self-evaluation (Hauser et al., 2017; Moro, Scandola, Bulgarelli, Avesani, & Fotopoulou, 2015; Nair, Palmer, Aleman, & David, 2014; Paulus, Huys, & Maia, 2016).

Domain-general beliefs about self-abilities are systematically lowered in depressed and anxious patients, and form a promising target for therapy (Bandura, 1977; Wells et al., 2012). One recent theory specifies a central role for metacognition in the computational etiology of such beliefs (Stephan et al., 2016). Briefly, symptoms of fatigue and depression are understood as sequential responses to pervasive “dyshomeostasis” – chronically enhanced surprise about internal bodily signals. This dyshomeostasis is monitored by a domain-general metacognitive layer that

downgrades beliefs about the brain's capacity to regulate performance on a range of tasks (self-efficacy beliefs). Recently we have systematically investigated the relation between subclinical psychiatric symptoms and metacognitive bias and efficiency in a large general population sample, finding dissociable relationships between psychopathology and metacognition in the absence of any links to first-order performance on a perceptual decision-making task (Rouault et al., 2018). A symptom dimension related to anxiety and depression was associated with lower metacognitive bias (lower confidence level) and heightened metacognitive efficiency, whereas a dimension characterizing compulsive behavior and intrusive thoughts was associated with higher metacognitive bias and lower metacognitive efficiency. Metacognitive bias has also been linked to trait optimism (Ais et al., 2016), which is intriguing as in this study the questionnaire was administered long after the experiment, suggesting a stable confidence level that transcends testing sessions.

In contrast, domain-specific deficits in metacognition of perception may play a role in the formation of hallucinations in psychosis (Klein, Altinyazar, & Metz, 2013; Moritz et al., 2014). Inaccurate metacognition for memory ability might explain symptoms of functional memory loss, a problem seen commonly in memory clinics (Stone et al., 2015) and account for why people with Alzheimer's disease often do not acknowledge their memory deficits (as evaluated with anosognosia questionnaires (Orfei et al., 2010); or objective tests of metamemory performance (Cosentino, Metcalfe, Butterfield, & Stern, 2007)). It remains to be explored how metacognitive efficiency as studied in laboratory tasks relates to real-world metacognition, but a few studies hint at such a link. For instance, participants with a higher level of metacognitive efficiency (perceptual meta-*d'/d'*) were perceived by their informants (e.g. relatives) to have fewer problems with attentional control in everyday life (Fitzgerald et al., 2017). Another study in older participants found that experimentally-measured metacognition (error awareness in a Go/NoGo task) correlated with error monitoring deficits in everyday life (Harty, O'Connell, Hester & Robertson, 2013).

Some illnesses might appear to co-occur with a generalized metacognitive impairment that underpins multiple problems or leads to severe deficits in daily functioning. In some situations, this metacognitive deficit is regarded as central to the mental

disturbance, such as a lack of insight common in some mental illnesses (David, 1990). For instance, patients with neurological disease and impaired knowledge of their disease (a symptom known as *anosognosia*) might be viewed as having a *prima facie* disturbance in metacognitive efficiency, although such a hypothesis remains to be directly tested. Strikingly, full insight often does not return despite overwhelming evidence of the neurological deficit (Cocchini, Beschin, Fotopoulou, & Sala, 2010; Fotopoulou et al., 2008). In the example of anosognosia for hemiplegia, a deficit in metacognition generalized across cognitive domains yet specific to the body part in question could go some way to explain this.

Initial findings of specific deficits in metacognition in neuropsychiatric conditions have tended to focus on specific domains of processing (such as memory in Alzheimer's disease) and it remains unknown whether these deficits generalize to other domains. Approaches to tackling this question are synergistic with transdiagnostic perspectives of psychopathology emerging in neuroscience (Barch, 2017; Rouault et al., 2018). Greater knowledge of the relationship between patterns of metacognitive deficits and individual neuropsychiatric profiles may eventually allow development of personalized therapeutic approaches and suggest pathways to train resilience to mental illness (Moro et al., 2015; Paulus et al., 2016).

Conclusions and future directions

Here we have reviewed the recent literature comparing the neurocognitive architecture of metacognition across domains. We have distinguished between the constructs of metacognitive efficiency (one's sensitivity to fluctuations in task performance) and metacognitive bias (one's overall confidence level, irrespective of performance). We have considered the importance of taking into account variations in task performance when measuring metacognition to allow meaningful comparisons across domains. In particular, it is critical in future studies match task and stimulus characteristics for across-domain comparison (e.g. Y/N vs. 2AFC). Finally, the use of a hierarchical Bayesian framework allows uncertainty in metacognitive efficiency parameters to be taken into account when examining correlations across domains (Fleming, 2017).

Our review of neuroimaging studies indicates a more nuanced picture of domain-generality than studies of behavior or individual differences, which have tended to argue for either domain-specific or domain-general aspects of metacognition, but not both. Recent studies suggest that both domain-specific and domain-general signals co-exist in the human brain, and that there may exist a gradient in which some tasks (such as different types of perceptual judgment) are more likely to rely on shared circuitry for metacognitive evaluation than others. Finally, we have highlighted the utility of computational models in providing a framework for understanding how confidence is formed across different tasks, and why it might be useful to maintain confidence in a common currency when switching between tasks. We suggest that the formation of confidence in one domain may provide useful priors on confidence formation in other domains. Notably the extent of such generalization itself could represent an individual difference, with extreme over-generalization possibly contributing to pervasive low self-efficacy often seen in depression and anxiety disorders.

Acknowledgements

We thank Robert Jagiello for contributing research to this article. Andrew McWilliams is a National Institute of Health Research-funded academic clinical fellow. The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust 203147/Z/16/Z.

Financial Disclosure

None of the other authors declare any conflict of interest.

References

Ais, J., Zylberberg, A., Barttfeld, P., & Sigman, M. (2016). Individual consistency in the accuracy and distribution of confidence judgments. *Cognition*, *146*(C), 377–386. http://doi.org/10.1016/j.cognition.2015.10.006

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *eLife*, *5*, e18103. http://doi.org/doi:10.7554/eLife.18103

Allen, M., Glen, J. C., Müllensiefen, D., Schwarzkopf, D. S., Fardo, F., Frank, D., et al. (2017). Metacognitive ability correlates with hippocampal and prefrontal microstructure. *NeuroImage*, *149*, 415–423. https://doi.org/10.1016/j.neuroimage.2017.02.008

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and social psychology review*, *13*(3), 219-235. http://doi.org/DOI:10.1177/1088868309341564
Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, *81*(1), 126. http://dx.doi.org/10.1037/h0027455
Baird, B., Cieslak, M., Smallwood, J., Grafton, S. T., & Schooler, J. W. (2015). Regional white matter variation associated with domain-specific metacognitive accuracy. *Journal of Cognitive Neuroscience*. http://doi.org/doi:10.1162/jocn_a_00741
Baird, B., Mrazek, M. D., Phillips, D. T., & Schooler, J. W. (2014). Domain-specific enhancement of metacognitive ability following meditation training. *Journal of Experimental Psychology: General*, *143*(5), 1972. http://doi.org/doi:10.1037/a0036882
Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *Journal of Neuroscience*, *33*(42), 16657–16665. https://doi.org/10.1523/JNEUROSCI.0786-13.2013
Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191. http://doi.org/DOI:10.1016/0146-6402(78)90002-4
Bang, D., Aitchison, L., Moran, R., Castanon, S. H., Rafiee, B., Mahmoodi, A., et al. (2017). Confidence matching in group decision-making. *Nature Human Behaviour*, *1*, 0117. http://doi.org/doi:10.1038/s41562-017-0117
Bang, D. & Fleming, S. M. Dissociating decision confidence from evidence reliability in the human brain. (2018). http://doi.org/doi:10.1101/251330
Barch, D. M. (2017). The neural correlates of transdiagnostic dimensions of psychopathology. https://doi.org/10.1176/appi.ajp.2017.17030289
Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*(1), 55. http://doi.org/10.1037/0096-3445.127.1.55
Boldt, A., de Gardelle, V., & Yeung, N. (2017). The Impact of Evidence Reliability on Sensitivity and Bias in Decision Confidence. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(8), 1520–1531. http://doi.org/10.1037/xhp0000404
Bornstein, B. H., & Zickafoose, D. J. (1999). " I know I know it, I know I saw it": The stability of the confidence–accuracy relationship across domains. *Journal of Experimental Psychology: Applied*, *5*(1), 76. doi:10.1037/1076-898X.5.1.76
Chiappe, D., & MacDonald, K. (2005). The evolution of domain-general mechanisms in intelligence and learning. *The Journal of General Psychology*, *132*(1), 5–40. http://doi.org/10.3200/GENP.132.1.5-40
Chua, E. F., Pergolizzi, D., & Weintraub, R. R. (2014). The cognitive neuroscience of metamemory monitoring: understanding metamemory processes, subjective levels expressed, and metacognitive accuracy. In *The cognitive neuroscience of metacognition* (pp. 267-291). Springer, Berlin, Heidelberg.
Chua, E. F., & Bliss-Moreau, E. (2016). Knowing your heart and your mind: The relationships between metamemory and interoception. *Consciousness and Cognition*, *45*(C), 146–158. http://doi.org/10.1016/j.concog.2016.08.015
Clarke, F. R., Birdsall, T. G., & Tanner, W. P., Jr. (1959). Two types of ROC curves and definitions of parameters. *The Journal of the Acoustical Society of America*, *31*(5), 629–630. https://doi.org/10.1121/1.1907764
Clithero, J. A., & Rangel, A. (2013). Informatic parcellation of the network involved in the computation of subjective value. *Social cognitive and affective neuroscience*, *9*(9), 1289-1302. https://doi.org/10.1093/scan/nst106
Cocchini, G., Beschin, N., Fotopoulou, A., & Sala, Della, S. (2010). Explicit and implicit anosognosia or upper limb motor impairment. *Neuropsychologia*, *48*(5), 1489–1494. http://doi.org/doi:10.1016/j.neuropsychologia.2010.01.019
Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual

performance. *Nature Communications*, *7*. http://doi.org/DOI:10.1038/ncomms13669
Cosentino, S., Metcalfe, J., Butterfield, B., & Stern, Y. (2007). Objective metamemory testing captures awareness of deficit in Alzheimer's disease. *Cortex*, *43*(7), 1004–1019. https://doi.org/10.1016/S0010-9452(08)70697-X
David, A. S. (1990). Insight and psychosis. *The British Journal of Psychiatry*, *156*(6), 798–808. https://doi.org/10.1192/bjp.156.6.798
de Gardelle, V., & Mamassian, P. (2014). Does confidence use a common currency across two visual tasks? *Psychological Science*, *25*(6), 1286–1288. http://doi.org/10.1177/0956797614528956
de Gardelle, V., Le Corre, F., & Mamassian, P. (2016). Confidence as a common currency between vision and audition. *PLoS ONE*, *11*(1), e0147901–11. http://doi.org/10.1371/journal.pone.0147901
Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a neural system for error detection and compensation. *Psychological Science*, *5*(5), 303-305. https://doi.org/10.1111/j.1467-9280.1994.tb00630.x
De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110. http://doi.org/10.1038/nn.3279
Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Foundations of human reasoning in the prefrontal cortex. *Science*, *344*(6191), 1481–1486. http://doi.org/10.1126/science.1252254
Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172–179. http://doi.org/DOI:10.1016/j.tics.2010.01.004
Dunlosky, J., & Metcalfe, J. (2008). Metacognition. Sage Publications.
Elliott, R., Sahakian, B. J., McKay, A. P., Herrod, J. J., Robbins, T. W., & Paykel, E. S. (1996). Neuropsychological impairments in unipolar depression: the influence of perceived failure on subsequent performance. *Psychological Medicine*, *26*(5), 975–989. https://doi.org/10.1017/S0033291700035303
Faivre, N., Filevich, E., Solovey, G., Kühn, S., Blanke, O. (2018). Behavioural, modeling, and electrophysiological evidence for supramodality in human metacognition. *Journal of Neuroscience*, *38*, 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017
Fetsch, C. R., Kiani, R., Newsome, W. T., & Shadlen, M. N. (2014). Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron*, *83*, 797-804. https://doi.org/10.1016/j.neuron.2014.07.011
Fitzgerald, L. M., Arvaneh, M., & Dockree, P. M. (2017). Domain-specific and domain-general processes underlying metacognitive judgments. *Consciousness and Cognition*, *49*, 264–277. http://doi.org/doi:10.1016/j.concog.2017.01.011
Fleck, M. S., Daselaar, S. M., Dobbins, I. G., & Cabeza, R. (2006). Role of prefrontal and anterior cingulate regions in decision-making processes shared by memory and nonmemory tasks. *Cerebral Cortex*, *16*(11), 1623–1630. http://doi.org/doi:10.1093/cercor/bhj097
Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1), 1–14. http://doi.org/10.1093/nc/nix007
Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. http://doi.org/10.1037/rev0000045
Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1338–1349. http://doi.org/10.1098/rstb.2011.0417
Fleming, S. M., & Lau, H. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*(443), 1–9. http://doi.org/10.3389/fnhum.2014.00443/abstract
Fleming, S. M., Huijgen, J., & Dolan, R. J. (2012). Prefrontal contributions to metacognition in perceptual decision making. *Journal of Neuroscience*, *32*(18), 6117–6125. http://doi.org/10.1523/JNEUROSCI.6489-11.2012

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, *137*(10), 2811–2822. http://doi.org/10.1093/brain/awu221

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science*, *329*(5998), 1541–1543. http://doi.org/doi:10.1126/science.1191883

Fotopoulou, A., Tsakiris, M., Haggard, P., Vagopoulou, A., Rudd, A., & Kopelman, M. (2008). The role of motor intention in motor awareness: an experimental study on anosognosia for hemiplegia. *Brain*, *131*(12), 3432–3442. http://doi.org/doi:10.1093/brain/awn225

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin Review*, *10*(4), 843–876. https://doi.org/10.3758/BF03196546

Garfinkel, S. N., Manassei, M. F., Hamilton-Fletcher, G., In den Bosch, Y., Critchley, H. D., & Engels, M. (2016). Interoceptive dimensions across cardiac and respiratory axes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1708), 20160014–10. http://doi.org/10.1098/rstb.2016.0014

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological science*, *4*(6), 385-390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Gherman, S., & Philiastides, M. G. (2017). Human VMPFC encodes early signatures of confidence in perceptual decisions. *bioRxiv*, 224337. https://doi.org/10.1101/224337

Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E., & Shadlen, M. N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *The Journal of Neuroscience*, *31*(17), 6339–6352. http://doi.org/10.1523/JNEUROSCI.5613-10.2011

Harty, S., OConnell, R. G., Hester, R., & Robertson, I. H. (2013). Older adults have diminished awareness of errors in the laboratory and daily life. *Psychology and Aging*, *28*(4), 1032. http://dx.doi.org/10.1037/a0033567

Hauser, T. U., Allen, M., Purg, N., Moutoussis, M., Rees, G., & Dolan, R. J. (2017). Noradrenaline blockade specifically enhances metacognitive performance. *eLife*, *6*, e24901. https://doi.org/10.7554/eLife.24901.001

Heereman, J., Walter, H., & Heekeren, H. R. (2015). A task-independent neural representation of subjective certainty in visual perception. *Frontiers in Human Neuroscience*, *9*(e96511), 195–12. http://doi.org/10.3389/fnhum.2015.00551

Hilgenstock, R., Weiss, T., & Witte, O. W. (2014). You'd better think twice: post-decision perceptual confidence. *NeuroImage*, *99*, 323–331. http://doi.org/10.1016/j.neuroimage.2014.05.049

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *109*(4), 679. http://doi.org/DOI:10.1037//0033-295X.109.4.679

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology*, *17*(1), 3–11. doi:10.3758/BF03337811

Kanai, R., Walsh, V., & Tseng, C. H. (2010). Subjective discriminability of invisibility: a framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and cognition*, *19*(4), 1045-1057. http://doi.org/doi:10.1016/j.concog.2010.06.003

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, *111*(2), 512–523. http://doi.org/10.1037/0033-295X.111.2.512

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, *28*(1), 92–107. http://doi.org/10.3758/BF03211579

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227. http://doi.org/doi:10.1038/nature07200

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science*, *324*(5928), 759-764. http://doi.org/doi:10.1126/science.1169405

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342. http://doi.org/10.1016/j.neuron.2014.12.015

Kievit, R. A., Lindenberger, U., Goodyer, I. M., Jones, P. B., Fonagy, P., Bullmore, E. T., et al. (2017). Mutualistic coupling between vocabulary and reasoning supports cognitive development during late adolescence and early adulthood. *Psychological Science*, *28*(10), 1419–1431. https://doi.org/10.1177/0956797617710785

Klein, S. B., Altinyazar, V., & Metz, M. A. (2013). Facets of self in schizophrenia: The reliability and accuracy of trait self-knowledge. *Clinical Psychological Science*, *1*(3), 276–289. http://doi.org/10.1177/2167702612474263

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 34. DOI: 10.1037//0278-7393.27.1.34

Lak, A., Costa, G. M., Romberg, E., Koulakov, A. A., Mainen, Z. F., & Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron*, *84*(1), 190–201. http://doi.org/10.1016/j.neuron.2014.08.039

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167. http://doi.org/10.1038/nn.4064

Lemaitre, A.-L., Herbet, G., Duffau, H., & Lafargue, G. (2017). Preserved metacognitive ability despite unilateral or bilateral anterior prefrontal resection. *Brain and Cognition*, *120*, 48-57. http://doi.org/10.1016/j.bandc.2017.10.004

Maniscalco, B., McCurdy, L. Y., Odegaard, B., & Lau, H. (2017). Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. *Journal of Neuroscience*, *37*(5), 1213–1224. http://doi.org/10.1523/JNEUROSCI.2271-13.2016.

Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 509–527. http://doi.org/10.1037/a0014876

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *Journal of Neuroscience*, *33*(5), 1897–1906. http://doi.org/10.1523/JNEUROSCI.1890-12.2013

Meuwese, J. D., van Loon, A. M., Lamme, V. A., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics*, *76*(4), 1057–1068. http://doi.org/10.3758/s13414-014-0643-1

Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The sense of confidence during probabilistic learning: A normative account. *PLoS computational biology*, *11*(6), e1004305. https://doi.org/10.1371/journal.pcbi.1004305

Modirrousta, M., & Fellows, L. K. (2008). Medial prefrontal cortex plays a critical and selective role in “feeling of knowing”meta-memory judgments. *Neuropsychologia*, *46*(12), 2958–2965. http://doi.org/10.1016/j.neuropsychologia.2008.06.011

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-general and domain-specific patterns of activity support metacognition in human prefrontal cortex. 2360-17. *Journal of Neuroscience, in press* https://doi.org/10.1523/JNEUROSCI.2360-17.2018

Moritz, S., Ramdani, N., Klass, H., Andreou, C., Jungclaussen, D., Eifler, S., et al. (2014). Overconfidence in incorrect perceptual judgments in patients with schizophrenia. *Schizophrenia Research: Cognition*, *1*(4), 165–170. http://doi.org/10.1016/j.scog.2014.09.003

Moro, V., Scandola, M., Bulgarelli, C., Avesani, R., & Fotopoulou, A. (2015). Error-based training and emergent awareness in anosognosia for hemiplegia. *Neuropsychological Rehabilitation*, *25*(4), 593–616. http://doi.org/10.1080/09602011.2014.951659

Nair, A., Palmer, E. C., Aleman, A., & David, A. S. (2014). Relationship between cognition, clinical and cognitive insight in psychotic disorders: a review and meta-analysis. *Schizophrenia Research*, *152*(1), 191–200. https://doi.org/10.1016/j/schres.2013.11.033

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P. E., & Bahrami, B. (2017). The idiosyncratic nature of confidence. *Nature Human Behaviour*, *1*(11), 810–818. http://doi.org/10.1038/s41562-017-0215-1

Nelson, T. O., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. *The Psychology of Learning and Motivation*, *26*, 125–173. http://doi.org/10.1016/S0079-7421(08)60053-5

Orfei, M. D., Blundo, C., Celia, E., Casini, A. R., Caltagirone, C., Spalletta, G., & Varsi, A. E. (2010). Anosognosia in mild cognitive impairment and mild Alzheimer's disease: frequency and neuropsychological correlates. *The American Journal of Geriatric Psychiatry*, *18*(12), 1133-1140. http://doi/org/10.1097/JGP.0b013e3181dd1c50

Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, *28*(C), 151–160. http://doi.org/10.1016/j.concog.2014.06.007

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: A review. *Neuropsychology Review*, *15*(3), 105–130. http://doi.org/10.1007/s11065-005-7091-6

Paulus, M. P., Huys, Q. J., & Maia, T. V. (2016). A Roadmap for the development of applied computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. http://dx.doi.org/10.1016/j.bpsc.2016.05.001

Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation.

Peters, M. A., Fesi, J., Amendi, N., Knotts, J. D., Lau, H., & Ro, T. (2017). Transcranial magnetic stimulation to visual cortex induces suboptimal introspection. *Cortex*, *93*, 119-132. https://doi.org/10.1016/j.cortex.2017.05.017

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864. http://doi.org/ 10.1037/a0019737

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374. http://doi.org/10.1038/nn.4240

Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H., & Lisanby, S. H. (2011). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *Journal of neurophysiology*, *107*(6), 1556-1563. https://doi.org/10.1152/jn.00985.2011

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*(11), 1664–1680. http://doi.org/10.1177/0956797615595037

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*(3), 127–157. http://doi.org/10.1016/j.cogpsych.2009.09.001

Reyes, G., Silva, J. R., Jaramillo, K., Rehbein, L., & Sackur, J. (2015). Self-knowledge dim-out: stress impairs metacognitive accuracy. *PLoS ONE*, *10*(8), e0132320. https://doi.org/10.1371/journal.pone.0132320

Richter, F. R., Cooper, R. A., Bays, P. M., & Simons, J. S. (2016). Distinct neural mechanisms underlie the success, precision, and vividness of episodic memory. *eLife*, *5*, e18260. http://doi.org/10.7554/eLife.18260

Ritchie, S. J., Bates, T. C., Der, G., Starr, J. M., & Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging*, *28*(2), 515–521. http://doi.org/10.1037/a0030820

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. *Biological Psychiatry*, in press. https://doi.org/10.1016/j.biopsych.2017.12.017
Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175. http://doi.org/10.1080/17588921003632529
Ruby, E., Giles, N., & Lau, H. (2017). Finding domain-general metacognitive mechanisms requires using appropriate tasks. *bioRxiv*, 211805. https://doi.org/10.1101/211805
Sadeghi, S., Ekhtiari, H., Bahrami, B., & Ahmadabadi, M. N. (2017). Metacognitive deficiency in a perceptual but not a memory task in methadone maintenance patients. *Scientific Reports*, *7*(1), 7052. http://doi.org/10.1038/s41598-017-06707-w
Samaha, J., & Postle, B. R. (2017). Correlated individual differences suggest a common mechanism underlying metacognition in visual perception and visual short-term memory. In *Proc. R. Soc. B* (Vol. 284, No. 1867, p. 20172035). The Royal Society. http://doi.org/10.1098/rspb.2017.2035
Schnyer, D. M., Verfaellie, M., Alexander, M. P., LaFleche, G., Nicholls, L., & Kaszniak, A. W. (2004). A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: Evidence from patients with lesions to frontal cortex. *Neuropsychologia*, *42*(7), 957–966. http://doi.org/10.1016/j.neuropsychologia.2003.11.020
Shimamura, A. P., & Squire, L. R. (1986). Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3).
Song, C., Kanai, R., Fleming, S. M., Weil, R. S., Schwarzkopf, D. S., & Rees, G. (2011). Relating inter-individual differences in metacognitive performance on different perceptual tasks. *Consciousness and Cognition*, *20*(4), 1787–1792. http://doi.org/10.1016/j.concog.2010.12.011
Stephan, K. E., Manjaly, Z. M., Mathys, C. D., Weber, L. A. E., Paliwal, S., Gard, T., … Petzschner, F. H. (2016). Allostatic Self-efficacy: A Metacognitive Theory of Dyshomeostasis-Induced Fatigue and Depression. *Frontiers in Human Neuroscience*, *10*, 550. http://doi.org/10.3389/fnhum.2016.00550
Stone, J., Pal, S., Blackburn, D., Reuber, M., Thekkumpurath, P., & Carson, A. (2015). Functional (psychogenic) cognitive disorders: a perspective from the neurology clinic. *Journal of Alzheimer's Disease*, *48*(s1), S5–S17. http://doi.org/10.3233/JAD-150430
Thompson, W. B., & Mason, S. E. (1996). Instability of individual differences in the association between confidence judgments and memory performance. *Memory & Cognition*, *24*(2), 226–234. https://doi.org/10.3758/BF03200883
Valk, S. L., Bernhardt, B. C., Böckler, A., Kanske, P., & Singer, T. (2016). Substrates of metacognition on perception and metacognition on higher-order cognition relate to different subsystems of the mentalizing network. *Human Brain Mapping*, *37*(10), 3388–3399. http://doi.org/10.1002/hbm.23247
van den Berg, R., Anandalingham, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *eLife*, 1–36. http://doi.org/10.7554/eLife.12192.001
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J Stat Softw*, *36*(3), 1–48. http://doi.org/10.18637/jss.v036.i03
Wells, A., Fisher, P., Myers, S., Wheatley, J., Patel, T., & Brewin, C. R. (2012). Metacognitive therapy in treatment-resistant depression: A platform trial. *Behaviour Research and Therapy*, *50*(6), 367–373. http://doi.org/10.1016/j.brat.2012.02.004
West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin & Review*, *4*(3), 387–392. https://doi.org/10.3758/BF03210798
Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., et al. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and

parietal cortex. *Proceedings of the National Academy of Sciences*, *107*(33), 14899–14902. https://doi.org/10.1073/pnas.1007928107
Ye, Q., Zou, F., Lau, H., Hu, Y. & Kwok, S. C. Causal evidence for mnemonic metacognition in human precuneus. *bioRxiv* 280750 (2018). http://doi.org/10.1101/280750
Yokoyama, O., Miura, N., Watanabe, J., Takemoto, A., Uchida, S., Sugiura, M., et al. (2010). Right frontopolar cortex activity correlates with reliability of retrospective rating of confidence in short-term recognition memory performance. *Neuroscience Research*, *68*(3), 199–206. http://doi.org/10.1016/j.neures.2010.07.2041
Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *eLife*, *5*, e17688. http://doi.org/10.7554/eLife.17688