Towards a common conceptual space for metacognition in perception and memory

Audrey Mazancieux, Michael Pereira, Nathan Faivre, Pascal Mamassian, Chris Moulin, Céline Souchay

Toward a common conceptual space for metacognition across and within domains

Audrey Mazancieux[1] *, Michael Pereira[2] *, Nathan Faivre[2], Pascal Mamassian[3], Chris J.A. Moulin[2], Céline Souchay[2]

[1] Cognitive Neuroimaging Unit, NeuroSpin center, Institute for Life Sciences Frédéric Joliot, Fundamental Research Division, Commissariat à l'Energie Atomique et aux énergies alternatives, INSERM, Université Paris-Sud, Université, Paris-Saclay, Gif-sur-Yvette, France
[2] Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France
[3] Laboratoire des Systèmes Perceptifs, Département d'Études Cognitives, École Normale Supérieure, PSL University, CNRS, Paris, France

* equal contribution

Corresponding author: Céline Souchay, celine.souchay@univ-grenoble-alpes.fr

Abstract

Evaluating, controlling and representing our cognitive states is paramount for efficient behavior. In this Review, we bridge landmark concepts of metacognitive abilities in the memory and perceptual domains to examine the different types of cognitive architectures that are at play when humans provide metacognitive judgments. Building upon this common conceptual framework, we review empirical evidence supporting or challenging domain-general metacognition. We also review commonalities across domains, focusing notably on the influence of decisional processes on metacognitive judgements. We emphasize the challenges of isolating metacognitive from cognitive processes and how this affects our interpretation of the domain-generality of metacognition, including in clinical conditions that are hypothesized to have metacognitive impairments. Finally, we also give an overview of the literature on what we classify as 'adecisional' metacognition: evaluations made outside the context of a decisional process. Based on current knowledge, we find no evidence for a strong form of domain-generality, but we outline how we may identify such an architecture in future research.

1. Opening section

Metacognitive evaluations are commonplace in daily life for expressing the commitment to complex or contested ideas and decisions, such as reporting being 70% sure that you remembered to start the dishwasher before leaving the house. . They are also important in perceptual decisions which take the form, for instance, of being sufficiently confident to step out to cross the road having judged the speed of an approaching car. Such reflections on the quality of perceptual or memory representations, as examples, have been studied in various fields.

Initially a developmental concept [1], metacognition has been investigated in the context of education [2], eyewitness memory [3], and in memory impairment [4]. Metacognition has also been used as a tool to study mechanisms underlying perceptual consciousness [5,6]. These last 15 years have witnessed an expansion of metacognitive topics: metacognitive processes have been evaluated in episodic and semantic memory, visual and auditory perception, reasoning, and motor function. Considering such a variety of domains as well as the numerous fields in which metacognition is studied (i.e., educational psychology, developmental psychology, neuropsychology, cognitive neuroscience, philosophy, economics), a need for a common conceptual space emerges. In parallel, a recent scientific goal has been to compare metacognition across modalities and domains [7–12], showing a growing interest in determining whether metacognition obeys domain-specific or domain-general rules [13].

Several reviews exist in both healthy [14,15] and pathological populations [16], but none have directly reviewed empirical evidence for metacognition across different cognitive domains, particularly metamemory (metacognition for memory processes) and metaperception (metacognition for perceptual processes). Since these two fields have developed separately, a comprehensive understanding of the cognitive architecture of metacognition as broadly construed is lacking.

In this review, the domain-generality of metacognition will be mapped out. Starting from the distinction between ‘metacognitive knowledge’ and ‘metacognitive experiences’ [1], we first propose possible architectures of metacognition across domains. Behavioral evidence for and against domain-generality is then explored. We review relevant models and critical properties of metacognition in the perceptual and memory domains in healthy and clinical populations with the goal of creating a conceptual space in which domain-generality can be discussed across fields.

2. Defining and measuring domain-general metacognition

From the outset, two levels of representation have been distinguished in metacognition. Flavell [1] proposed that these two levels were interrelated, one, 'metaknowledge' informing the other, 'metacognitive experience'. Metacognitive knowledge encompasses specific knowledge about one's own cognitive capabilities (e.g. 'I have more success in detecting my brothers face in a crowd when I meet him at the airport, than I do remembering to send him a birthday card'), knowledge about the impact of the task, and about strategy use. Metacognitive experience, however, is directly related to the task at hand and the decisional process to which it pertains. Koriat [17] refined this distinction as between information-based metacognition and experience-based metacognition. Although information-based metacognition involves inferential processes from explicit theories or beliefs, experience-based metacognition involves the experience of a cognitive process giving rise to a metacognitive feeling through the application of heuristics [18].

Experience-based metacognition can be subtended by several types of cognitive architectures with distinct levels of domain-generality. Starting from Nelson and Narens' influential view of metacognition [19], we propose four potential architectures (Figure 1) that link the meta-level and the object-level. For clarity, we define the meta-level as involving second-order representations and behaviours and the object level as involving first-order representations and behaviours, although a theoretical dissociation can be made between these two dimensions [20]. In a strongly domain-general architecture, one metacognitive module monitors and controls several independent cognitive domains (Figure 1a). At the algorithmic level, this meta-level module can instantiate second-order processes across domains that are independent of first-order processes. For instance, in an attempt to propose a multidomain approach to metacognition, Koriat's [21] self-consistency model proposes that monitoring is based on a process that samples different representations from a pool of representations. This sampling process can be the same across domains, even though representations are specific to each domain.

Another architecture represents domain-specific metacognitive modules that share processes in specific first-order contexts (e.g., evaluation of reaction time in decision-making tasks). This architecture would correspond to a 'weaker' domain-generality because it is limited to some specific aspects of metacognition (Figure 1b). Domain specificity is defined in cases where no process is shared between the first and second-order levels (Figure 1c). Note that there is likely a continuum of possible weak domain-generality architectures

starting from a process that is systematically used (strong domain-generality) to uniquely used (domain-specificity).

Crucially, because of the relationship between the two levels, variables that influence first-order processes in a domain-general way can also influence second-order processes, creating a spurious domain-generality (Figure 1d). This points to the importance of separating first-order and second-order processes [22]. For instance, working memory, attention, motivation, arousal, or mood (discussed in Section 3.3) can affect in parallel both cognitive and metacognitive measures across domains, giving the impression of a shared architecture, but where there is actually no metacognitive component. For example, if mood gives rise to better cognitive function and also influences metacognitive measures to the same extent, the aspects of first and second-order processes that share this influence of mood will appear related across domains, whilst there are actually no shared processes in the metacognitive architecture.

In this review, we focus on the computational (the behaviour we want to explain) and algorithmic levels (how this behaviour is computed) of Marr's three levels of analysis [23]. Note that notions of strong or weak domain-generality can also be examined at the implementation level (see [24] for a review of the neural correlates of metacognition). For instance, having the same neurons or regions involved in the computation of metacognition for several domains suggests a strong domain-general architecture [25]. It is possible, however, to have a strong domain-general process at the algorithmic level without having a strong domain-general implementation (e.g., several task-specific regions can perform similar computations).

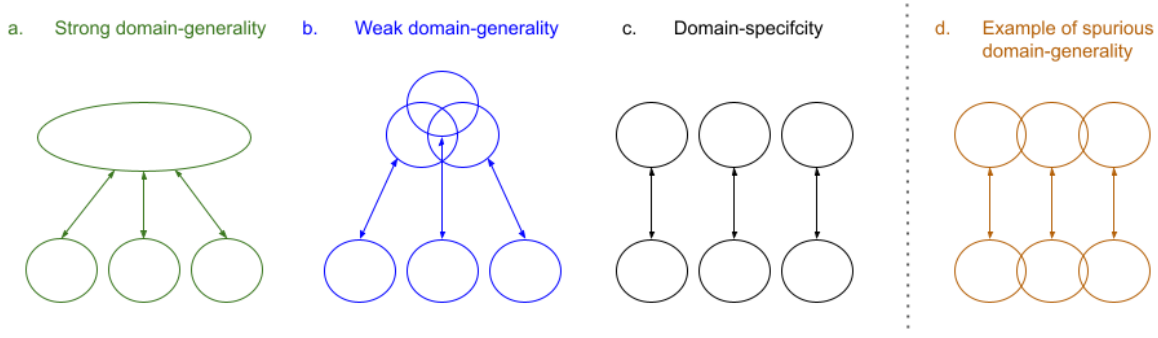


Figure 1. Theoretical architectures for domain-general metacognition showing configurations of first-order (below) and second-order (above) processes. *a) 'Strong' domain-generality where the same metacognitive mechanisms are involved in the monitoring and control of all cognitive domains. b) 'Weak' domain-generality where representations are specific to each domain but can share mechanisms in specific first-order contexts. c) Domain-specificity where a separate*

metacognitive level is associated with each cognitive domain. d) One example of spurious domain-generality where shared first-order mechanisms impact second-order mechanisms in a domain-general way.

To investigate the architecture of metacognition, we focus on direct measures (Table 1 see Box 1 for a description of each direct measure) whereby participants have instructions to self-reflect about a first-order task (such as the encoding of information, the recall of information or a decision) and report it. In indirect measures, other behaviours are used to infer metacognitive processes during a given cognitive process (e.g., the encoding of information or a decision); participants are not directly asked to introspect and evaluate their performance (see Box 2 for a description of indirect measures).

The commonly used distinction in the metamemory literature between prospective and retrospective judgements [26] does not help develop a shared conceptual space that functions across domains. For instance, a Judgment-of-Learning (JOL [27]) is classified as a prospective judgment, since it is a prediction of upcoming recall once a stimulus has been encoded. Nonetheless, the JOL is based on the experience during the encoding stage, and as such it pertains to a second-order judgment about a (past) process. Furthermore, as has been shown by delaying the point at which a JOL is made [28], these judgements arguably rely on a retrieval attempt. Likewise, the Feeling-Of-Knowing (FOK) judgment is a prediction of future recognition [29,30]. A retrieval attempt or evaluation of a presented cue is necessary to gauge the likelihood of future recognition. Indeed, partial knowledge available during recall has been shown to correlate with the accuracy of FOK judgements [31,32]. Describing FOKs as prospective because they are made for an upcoming test overlooks the fact that they result from a decision about the retrievability of information from memory. As we review below, this is subtly different from the judgements taken in metaperception designs [33] where participants first give their confidence level and then perform the decisional task (but see [34] for prospective metaperceptual judgements performed before stimulus onset). In short, the prospective-retrospective distinction limits the classification of metacognitive evaluations to metamemory since it is a conceptual framework which is only pertinent to encoding-storage-retrieval designs.

We propose to distinguish decisional metacognition as any type of evaluation made around the point at which a specific decision is made during a task qualified as first-order (or sub-task, such as encoding a word for later recall). This is a proposal, unlike the prospective and retrospective distinction, that can apply across domains and decision types. The critical issue is that in a decisional metacognitive evaluation, a first-order decision is taken which leaves a trace from which we can extract some information in order to make a second-order

evaluation. If no decision has been taken then we have no access to such decisional information on which to base our judgement, and the evaluation is of a different kind. We will see that referring to these as decisional judgments means that there are common methods and models that pertain to multiple cognitive domains. Of course, metacognition is not only limited to decisional judgements, and we offer some speculative comments on adecisional judgements in Section 4 (and it also features in Table 1). Decisional and adecisional metacognition are also different from Koriat's distinction as decisional judgements can be information-based (e.g. bias or starting point of evidence accumulation, see section 3.2) or experience-based (e.g. drift rate or sensitivity, see section 3.2). Likewise, adecisional judgements can be information-based (e.g., beliefs) or experience-based (e.g. emotional state).

The most frequently used second-order decisional measure pertains to confidence. Retrospective confidence judgments refer to the level of confidence that a participant has in being correct on a given first-order decision. As they can be performed on any kind of first-order decision, they have been the main measure used to investigate domain-general metacognition (see Section 3.1).

Table 1. Measures of metacognition. *EOLs = Ease-of-learning; JOLs = Judgement-of-learning; FOKs = Feeling-of-knowing; RCJs = Retrospective confidence judgements. Star refers to measures that can be used across different domains (notably memory and perception). Note that all remaining measures pertain to the memory domain, thus there is no perceptual specific metacognitive measure.*

Direct measures		**Indirect measures**
Adecisional	**Decisional**	
Global predictions*	JOLs	Re-study choice
EOLs	FOKs	Study time allocation
	RCJs*	Opt-out paradigms*
	Confidence forced-choice*	Post-decision wagering*
		Optimal waiting time*

The relationship between metacognition and task performance is demonstrated in two different concepts: metacognitive bias and metacognitive sensitivity. Metacognitive bias refers to the magnitude of metacognitive evaluations relative to the level of performance,

mainly quantified by averaging ratings over trials. For instance, overconfidence is defined as "the tendency to give high [metacognitive] ratings, all else being equal" (p.5, [22]). Metacognitive sensitivity refers to the ability to discriminate between correct and incorrect responses when performance is neither at ceiling nor at chance levels. The optimal calculation of metacognitive sensitivity has been already widely discussed (see Box 3) and reviewed [22,35–38].

We now review evidence for and against domain-general metacognition for metacognitive bias and metacognitive sensitivity in decisional judgments. Using models in the metaperception and metamemory field, we describe potential domain-general processes with reference to our proposed metacognitive architectures (Figure 1).

3. Decisional judgments across perception and memory

3.1. Evidence for and against domain-generality of decisional judgements

In behavioural studies, there are different methods for assessing the generality of a metacognitive process. The first and most common class of methods consists in assessing the correlation between metacognitive bias or metacognitive sensitivity across domains, typically at the level of a group of participants. Domain-generality implies that evaluations of decisional judgements should be correlated across different domains, so that individuals with high levels of metaperformance in one domain also have high metaperformance in another domain. A second method to assess the domain-specificity of a process is to use functional independence. From this perspective, processes are supposed to be independent if one variable has an effect on one process and no effect or the opposite effect on another process.

According to correlation studies, metacognitive bias is consistently found to be stable across many domains including perception and episodic and semantic memory [7,9–12]. However, the pattern appears more complex for metacognitive sensitivity. A previous review [13] concluded that whilst domain-general metacognitive sensitivity can be identified for perception across different modalities, correlations across metamemory and metaperception are low, possibly due to variability in measures of metacognitive sensitivity that were used and the low sample size of the studies. Using more appropriate measures that isolate the confound between first- and second-order performance (e.g., metacognitive efficiency, see [22] and Box 3) as well

as more appropriate sample size to test hypotheses, several studies have found positive correlations across memory and perception tasks [9–12,39,40]. Note that these correlations are restricted to two-alternative forced-choice tasks (2AFC, i.e., discrimination tasks) and that they appear to be absent for yes/no tasks (i.e., detection tasks [9]), when the two have been compared. These kinds of results may point to different computations underlying confidence for detection and discrimination tasks [41].

Despite such between-subject correlations supporting domain-general metacognitive efficiency, a large variability is found in the magnitude of correlations. Most of the above-mentioned studies reporting between-domain correlations at the group level used hierarchical estimations of metacognitive efficiency [42]. These measures have the advantage of taking into account both within- and between-subject variability to directly estimate covariance in metacognitive efficiencies across domains but may also inflate correlation estimates [43] see [40,44] for recent evidence based on non-hierarchical estimates. Another main issue is that measures of metacognitive efficiency have low half-split reliability [45] which may limit the strength of the correlation across tasks. Besides the problem of hierarchical estimates, yet another issue is that correlations across domains may actually reflect the influence of domain-general metacognitive bias. Indeed, even though these two quantifications of metacognition are supposed to be independent, correlations between metacognitive efficiency and bias are found in practice [46]. In sum, most of the evidence supporting the domain-generality of confidence is relatively weak and based on between-subject correlations, and future work should investigate what factors may underpin the domain-generality of metacognition.

Beyond these between-subject correlations approach, more direct methods have been used to support the view that metacognition involves domain-general mechanisms, notably by showing that confidence is encoded with a ‘common currency’ across different tasks. Using the confidence forced-choice paradigm, de Gardelle and colleagues [47] showed that participants were able to compare confidence across visual and auditory decisions with the same precision as for the comparison of two trials within the same sensory modality. In the same vein, the computational models reproducing confidence estimates about audiovisual decisions use supramodal formats of confidence in which auditory and visual confidence signals are either integrated or compared to one another, also suggesting a common currency [8]. To our knowledge, research using this approach remains limited to comparisons between sensory modalities, and evidence supporting a common currency of confidence across cognitive domains including perception and memory is still lacking.

Another approach involves assessing the domain-generality of confidence in neuropsychological groups. Using three groups of participants (a control group, a group with lesions in the anterior prefrontal cortex (aPFC), and a group with temporal lobe lesions), Fleming and colleagues [48] showed a deficit in perceptual metacognitive efficiency (and not for memory) in the aPFC lesion group, while there was no difference in first-order performance or metacognitive bias. Sadeghi and colleagues [49] found the same pattern of results with substance-dependent individuals, supporting domain-specific impairments of metacognition. In contrast, metacognitive efficiency remained stable with age for both visual perception and recognition memory among healthy volunteers [12]. Similarly, no specific differences were found for transdiagnostic subclinical symptom dimensions between visual perception and semantic memory metacognitive efficiency [40]. Disruption of the precuneus has been shown to selectively alter metacognitive efficiency in a memory task but not a visual perception task [50,51]. In the memory domain, neuropsychological studies have consistently shown inaccurate episodic FOKs with preserved semantic FOKs in several populations (e.g., [52,53] see Section 3.4) although this pattern might be driven by first-order performance differences and therefore uninformative regarding metacognition per se. Overall, despite a growing interest in the neuropsychology of metacognition, the evidence supporting domain-general or domain-specific architecture is mixed, and systematic investigations of bias-free metacognitive performance indices across endophenotypes [54] and cognitive domains are needed.

As suggested in Section 2, the dissociation between first-order and second-order processes is critical for differentiating several architectures of metacognition. Models of decisional judgements have been developed which can account for processes driven by first-order or second-order evidence, and as such could point to domain-general mechanisms. We review these models in the next section.

3.2. Models of decisional judgements

Computational models attempt to explain how decisional judgments originate from sensory evidence. They mostly differ in the way that first-order evidence is reused (or not) by putative second-order metacognitive processes [55]. Most models are hierarchical in the sense that they assume a common source of evidence for first-order decisions and second-order judgments, but with possibly different readouts (Figure 2). Other models assume either that evidence for decisions or recall and metacognition are not identical albeit correlated [56–58] or that evidence for metacognition is processed separately from the first-order process [59,60].

Hierarchical models define confidence as a readout of the first-order decisional process, which is modeled either as a snapshot (based on signal detection theory or an extension thereof), or as a dynamical process that accounts for the way first-order decisions unfold over time. For perceptual decision-making, the latter models assume that noisy sensory evidence is accrued over time up to a decision boundary [61]. Similar models have been developed for recognition memory, assuming that the match between a test item and memory produces evidence that is also accrued over time [62,63]. Notably, these dynamical models can also be extended to be closer to a neural implementation using neuronal networks [64–67] that arguably rely on very similar readouts. Dynamical models can be further categorized into drift-diffusion models with one single accumulator encoding the relative evidence for a choice [68] and models with one accumulator per choice that can account for multiple choice decisions [69]. There are similarities between the static and dynamical models in the use of first-order evidence to make second-order judgments [14], but there are also differences especially when response times are considered to be important. The heterogeneity in hierarchical models of confidence formation lies mainly in the definition of the confidence readout, which varies largely depending on the underlying first-order model and whether it considers post-decisional evidence or not (Figure 2).

The simplest hierarchical models of confidence formation are solely based on the strength of first-order evidence and assume a readout of the distance between noisy sensory information and a decision boundary. These models are found both in recognition memory [70,71] and perception [72,73], with decisional judgments possibly degraded by an additional source of (metacognitive) noise [74–76]. Rather than simply mapping a readout of the level of evidence to a confidence scale, other confidence readouts have also been developed within a probabilistic framework. Examples include defining confidence as the probability of a choice being correct knowing the sensory evidence [77], as a log-probability ratio of two possible choices [78], or as the precision (inverse variance) of the underlying sensory distribution [79]. Recent models have formalized the idea that decisional judgements are based on both prior beliefs about memory (information-based metacognition) and processing experience during the memory process (experience-based metacognition) [18]. Such Bayesian inference models of confidence are found in both perception [80,81] and memory [82].

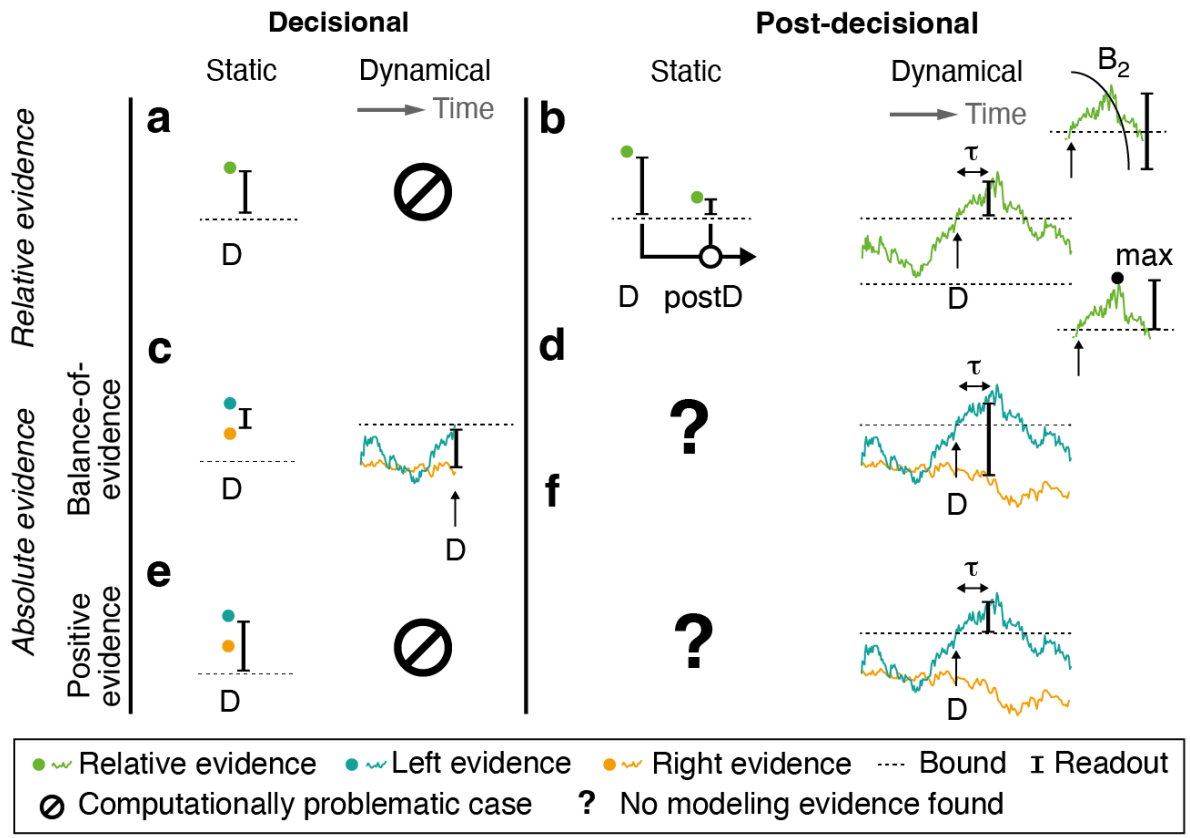


Figure 2. Confidence readouts for hierarchical models in a typical left-right discrimination task. *Confidence can be read out from evidence (vertical axis) at the time of the decision ("Decisional" column), or later ("Post-decisional"). "Static" models assume that only snapshots of evidence are available, and "Dynamical" ones capture how evidence evolves over time (horizontal axis). Perceptual evidence can be either relative (left minus right;* ***a–b****), or absolute, encoding evidence for left and right separately (****c–f****).* ***a.*** *Confidence is read out from the distance between (static) relative evidence (green full circle) and the decision bound (horizontal dashed line) at the time of the decision. D [72,73]. Note that this readout cannot be generalized to a dynamical model as accumulated relative evidence is constant at the time of the decision, unless a collapsing bound is assumed and confidence would then be equivalent to response time.* ***b.*** *The static model in (****a****) can integrate decisional (D) and post-decisional (postD) evidence either as a weighted sum, or through a second-order mechanism [57]. In dynamical models, relative evidence accumulated post-decisionally can be read out at a fixed timing τ [83,84], when it reaches a collapsing (second-order) boundary (B2; upper inset; [85]), or what maximal value it reaches (lower inset; [86,87]).* ***c.*** *Models of absolute evidence can read out confidence as the balance-of-evidence for left and right choices at the time of decision D ([88] for static models). In dynamical models, this is equivalent to reading out the state of the losing accumulator (i.e. right – incongruent with the choice; [62,78]).* ***d.*** *Dynamical models in (****c****) can easily be extended to integrate post-decisional evidence by delaying the readout by a fixed timing τ [89,90].* ***e.*** *Other models of confidence only read it out from the positive evidence (i.e. left – congruent with the choice; [55] for static models). Note that this readout cannot be generalized to a dynamical model as accumulated relative evidence congruent with the choice is constant at the time of the decision.* ***f.*** *Models in (****e****) can easily be extended to integrate post-decisional evidence by delaying the readout by a fixed timing τ ([91–93] for dynamic models).*

In sum, the aforementioned models describe how the strength of first-order evidence (possibly augmented by post-decisional evidence) relates to confidence. Although

commonalities have been found across domains in the way that this evidence is readout to build confidence it will be necessary to conduct cross-domain modelling studies to better isolate a common mechanism. As first-order evidence is arguably mainly domain-specific, the domain-generality of metacognition could either arise from a common source of metacognitive noise [8,47], or by (possibly domain-general) factors that differentially affect second-order metacognitive processes, which we review in the next subsection.

3.3. Dissociations between first-order and second-order processes

The aforementioned models assume some commonalities between cognitive and metacognitive processes. This partial overlap implies that although any factor affecting first-order processes will partially be reflected in decisional confidence judgements, this does not imply that it directly affects metacognition. For example, different types of visuo-spatial attention affect discrimination performance at the first-order level but not metacognitive sensitivity when controlling for first-order performance [94]. Similarly, visual information improves first-order performance in motor tasks but not metacognition *per se* [95]. Therefore, to isolate a second-order process, one needs to find differences in confidence that cannot be explained by differences in first-order processes, either by titrating task difficulty [96], or by comparing them to the ideal confidence observer [74]. Consequently, it is crucial to account for first-order performance to be able to test for the domain-generality of metacognition (Figure 1a-b), or else one risks being fooled by spurious domain-generality (Figure 1d) that is driven by factors that commonly affect first-order performance, such as attention.

One procedure to find dissociations at the second-order level is to manipulate the level of positive evidence for a perceptual decision while maintaining the same signal-to-noise ratio [97,98]. When one assumes that first-order performance is driven by the signal-to-noise ratio and second-order performance by the level of positive evidence, this procedure can be used to dissociate first and second-order processes behaviorally [91] and neurally [99]. Similarly, increasing the volatility of sensory evidence through time-varying noise leads to similar first-order performance but increased confidence ratings [100], suggesting that this volatility affects metacognition specifically. Various other factors have been found to affect second-order processes specifically. Metacognitive sensitivity is higher for perceptual decisions congruent with prior expectations [80] or for unexpected action outcomes [101] and participants make better use of priors at the metacognitive level than for first-order decisions [81]. Likewise, it is now clear that confidence is modulated by the motor actions that take place when providing a first-order response. Confidence, but not first-order performance, is modulated by subthreshold motor activity [102], the electrophysiological signature of motor preparation

preceding first-order responses [8], as well as sensorimotor activity leading to the first-order response [103–105]. Conversely, metacognitive performance is known to decrease when first-order responses occur after confidence ratings [106], when they are perturbed using transcranial magnetic stimulation [107] and sensorimotor conflicts [104], or when they are made by an external agent (although it is debated whether the same mechanisms are at play) [93,108,109].

Studies in the memory domain have used manipulations which differentially affect first-order performance and confidence, for instance, changing the magnitude of confidence judgements without changing memory performance [110]. So far, the same manipulations have not been studied across domains, and as such, metacognitive illusions within a domain do not provide direct evidence for domain generality, although it may point to spurious domain generality, in so far as identifying factors which are specific to second order judgements, which may or may or not be shared between domains.

These findings clearly show that some factors generate dissociations at the metacognitive level, independent from first-order performance. These dissociations are further supported by stimulation studies showing that metacognitive efficiency is specifically affected when disrupting brain activity in the occipital [111], premotor [107] or dorsolateral prefrontal cortices [112,113] but see [114]) in perceptual tasks, as well as when disrupting the dorsolateral prefrontal cortex for temporal working memory [115] or the precuneus for episodic memory [50]. It is unclear, however, how these different factors are integrated into decisional judgments. For example, reaction time affects confidence judgments [116,117] but this effect can be explained by a metacognitive process integrating reaction times into confidence ratings (e.g. Fleming 2018 Nat.Neuro.). In other computational modelling works, the same effect is solely explained by first-order processes such as the level of correlated noise between two accumulators [118]. Likewise, a recent study showed that response caution – or how participants balance response speed and accuracy – affects decisional judgments by changing the amount of post-decisional information available [119]. It is therefore unclear whether post-decisional evidence accumulation is a first- or second-order process. These considerations are important as they could imply spurious domain-generality; domain-specific first-order processes could inform metacognitive processes differently than for first-order decisions and common metacognitive measures would fail to diagnose such cases [119].

Another possibility is that people may have access to *cues* to specifically inform their decisional judgments beyond first-order processes [120]. In the metamemory field, people might not use the strength of their memory trace but infer it using different cues. Cues are

diagnostic when they are also pertinent to the memory retrieval itself, or non-diagnostic when they have no influence on the memory test per se. Mnemonic cues include the experience of internal indicators or signals that can be used to evaluate their level of memory performance. Several mnemonic cues are identified to influence a variety of decisional judgements such as the familiarity of the cue [121] or its fluency [122]. Fluency refers to the subjective experience of processing information easily [123]. Among different types of fluency, answer fluency (i.e., the ease with which information comes to mind [124]) has been shown to affect confidence judgments across multiple domains. Responses easily retrieved (both correct and incorrect) are judged with higher confidence in semantic memory tasks [125]. Within this framework, response times could also be used as a cue to inform confidence [126].

In sum, there is now strong evidence supporting a dissociation between the information available to first- and second-order processes. This has important implications for domain-generality, as some factors such as sensorimotor activity when reporting a decision might affect metacognition similarly for different domains and thus result in domain-generality. The fact that other factors are inherently specific to a domain such as sensory noise in metaperception [100] or familiarity in metamemory [121] could argue for a rather weak form of domain-generality. Alternatively, one could posit that a common metacognitive process adaptively weighs these different factors according to the task at hand, thereby still consistent with strong domain-generality. For instance, Shekhar and Rahnev [127] proposed that cross-task correlations for metacognitive sensitivity may be driven by common sources of metacognitive inefficiency (i.e., different types of noise) for these different tasks. As a result, cross-task correlations can be observed under a strong domain-general architecture where a common metacognitive module is altered by different types of metacognitive noise, but also in case first-order factors impact distinct metacognitive modules. Future work is required to understand the underlying mechanisms of these dissociations better. Even when accounting for first-order performance, it is not always straightforward to ascribe the influence of a factor to first- or second-order processes, which could lead to spurious domain-generality being overlooked.

3.4. Clinical dissociations and developmental studies

Dissociations between first- and second-order processes have also been observed in clinical populations. Patients with frontal lobe lesions are impaired on an episodic but not on a semantic memory FOK task, but only when their memory performance was impaired [128]. Interestingly, however, patients with frontal and temporal lobe lesions show preserved episodic FOK, JOL or confidence judgments, although their memory is impaired [129–131].

Likewise, decisional judgments have also been explored in Alzheimer's disease. Findings reveal that judgments such as JOLs and semantic FOKs are accurate [132–134] , as well as confidence judgments [135,136], again despite impaired memory performance. Episodic FOKs, however, were found to be inaccurate [137]. A dissociation between episodic and semantic FOK was also found in multiple sclerosis, with episodic FOK deficits more prevalent in the patients with low memory performance [52] but no FOK difference with controls when first-order performance was the same across groups [138]. FOK judgments have also been explored in Parkinson's disease and findings show that first-order impairment is associated with second-order deficits [139,140]. Recent work showed no impairment in metamemory or metaperception for functional cognitive disorder when controlling for first-order performance, but differences were found in global reports of subjective performance [141].

Decisional judgments have also been explored in psychiatric disorders such as bipolar disorders, most studies reporting limited correspondence between memory performance and metacognitive judgments [142] mainly resulting as underestimation of performance rather than a metacognitive sensitivity deficit per se. Underconfidence in both memory and perception was also observed in obsessive-compulsive disorder [16], but metacognitive sensitivity deficits were only found in a subclinical [143,144] and not in a clinical sample [145]. Finally, deficits in decisional judgments have been observed in patients with schizophrenia in vision [146,147], audition [148] and memory [149,150], but a recent meta-analysis found no evidence for a metacognitive deficit when isolating studies that controlled for first-order performances [151].

In sum, most research with clinical populations points to a specific metacognitive impairment that may be simply a consequence of a first-order deficit. The majority of studies, and particularly in metamemory, use measures of metacognitive sensitivity that do not control for differences in first-order performance (e.g. the gamma correlation). Ideally, a convincing datum would be a population with impaired second-order performance in the context of preserved first-order performance. The case of blindsight may correspond to such a dissociation between perception and metaperception, although the specific level at which vision is impaired remains debated [152,153]. In metamemory, a similar dissociation was reported by Wojcik and colleagues [53], who found inaccurate episodic FOK (but not semantic FOK) for children with Autism Spectrum Disorder (ASD) with no deficit in either recognition or recall. There are equivocal findings in this regard in the ASD literature, with studies showing impaired episodic FOK sensitivity in the context of impaired memory performance [154] and others showing no difference in first-order performance alongside impaired metacognitive sensitivity in confidence judgements [155]. A meta-analytic review [156] found a reduction in metacognitive accuracy across tasks, where performance was not 'universally

diminished'. It has been proposed that this specific impairment may result from an inability to cast the self in the past, related to autonoetic consciousness [157]. This points to dysfunction related to self-representation. Future research should consider the access to metacognitive information in cases such as autism spectrum disorder where there is apparently not a first-order deficit.

Global judgements, with their simplicity and low attentional demands, are particularly suited to patient studies. In samples too small to carry out between-subject correlations, these studies operationalise accuracy as the unsigned difference in predictions and performance. In Alzheimer's disease, when comparisons between control groups and patients' groups have been made, patients overestimate their performance for word lists [158,159] or flashbulb memories [160] and visuospatial tasks [161]. Critically, however, accuracy improves when predictions are made after participants have experienced the task [158,162,163]. These findings, according to the Cognitive Awareness Model [164], a model of anosognosia (see Box 5) suggest that patients fail to transfer their online awareness to a lasting belief about memory performance. Due to their memory problems Alzheimer's patients do not consolidate the information concerning their cognitive performance in their metaknowledge. A similar interpretation can be found in the motor domain, where anosognosia of hemiplegia has been postulated to be explained by a deficit in monitoring processes allowing abnormal perceptions to become abnormal beliefs [165].

Finally, metacognition has also been explored in development. The potential of domain transferability of metacognition has been of particular interest in education in relation to school achievement. It is mostly suggested that the development of metacognition begins as domain-specific and then generalizes across domains as children mature [166–168]. For example, Vo et al. [169] demonstrated that children as young as 5 years were metacognitively accurate on two nonverbal discrimination tasks but that metacognition in one domain (emotion discrimination) was unrelated to metacognition in another domain (numerical discrimination) giving support for a domain-dependant metacognition. Similar findings were more recently reported by Bellon, Fias and Bert De Smedt [170] between two distinct academic domains: arithmetic and spelling. Future longitudinal studies could examine how metacognition develops with age considering the relationship between self-concept defined as domain-specific self-perceived competence [171] and metacognitive monitoring [172].

The results of studies in special populations point to patterns of dissociations across tasks and domains which on face value point to a lack of domain-generality, and in some cases, such as with the episodic FOK suggest a domain specificity architecture. However, where

metacognitive measures and meta-analyses have taken into consideration first-order performance, the results are less compelling: the pattern of second-order deficits is somewhat a function of the first-order deficits in these groups. Of note, developmental studies and patient studies use paradigms that are more suitable for these specific populations. These measures can be seen as mainly underpinned by information-based metacognition and can be described as adecisional judgements (see Section 2). Such judgements are discussed in more detail in the next section.

4. Beliefs and adecisional judgements

So far we have focused on metacognition that is at play in decisional judgements. However, it is clear that metacognitive processes also extend to more generalised beliefs and evaluations, and extend across items within a task, and across tasks with domains. Here we review works that are based on such measures of metacognition that belong to beliefs, expertise and more to information-based metacognition. These adecisional judgements are based on metaknowledge rather than experiences because participants have not experienced the task at the moment of the judgment (see Table 1). This is notably the case with global predictions, especially when an evaluation is made before performing a task. In a typical paradigm, participants judge how many items from an entire study list they will subsequently recall [162,173,174]. However, these judgements have also been extended to other first-order tasks such as short-term memory [175], processing speed and verbal fluency [138] or perceptual tasks [144].

Most studies have used between-subject correlations to examine population-level accuracy of global predictions in different groups [176]. If global predictions are made before and after a study phase, the between-subject correlation between prediction and performance was found to be higher after studying the items, indicating an effect of monitoring [173,176]. However, it has been shown that the modal value of individual-level predictions is tethered to the midpoint of the scale (e.g. predicting recall of 6 items for a list of 12 [177]), suggesting that to a large extent these judgements are based on generalized beliefs and rules of thumb. Of course, such values may represent a Gaussian distribution around a central point, but the extent to which these estimations are rule-based and not distribution-based comes from the fact that the same participants predict recall of 5 items from a 10-item list but 10 items from a 20-item list [178].

Where item-by-item and global predictions have been carried out in the same task, mean values of item-by-item judgements (i.e., metacognitive bias) have mostly correlated with the global predictions [173]. Similarly, to examine between-subject patterns of accuracy, some authors have used the mean value of item-by-item judgements, with the finding that, for instance, mean retrospective confidence correlates with their performance but seemingly more so for general knowledge than episodic memory [179,180]. Some researchers have considered global metacognition as a 'self-rated ability' scale (e.g. 0–9% of people would be worse than me), where participants rate their performance in comparison with their peers [181]. This yields similar findings with correlations between self-rated ability and mean levels of retrospective confidence on general knowledge (sport) questions but not for an episodic task (face recognition) [181]. In applied fields, there was much enthusiasm for the idea that metacognitive accuracy (and self-rated ability) in one domain might help interpret evaluations in another domain, such as the comparison between global evaluations of general knowledge and eyewitness testimony, but recent interest in this approach has been lacking, probably because early research did not find robust relationships across domains.

The above examples of global evaluations consider applied questions such as differences in age, educational settings and neuropsychology, and perhaps unsurprisingly, global metacognition, defined as constructs such as self-efficacy and self-beliefs, has also received attention as a theoretical entity, since it 'may be more closely related to daily functioning…' (p.436 [182]). As such, the idea is that in daily life the individual draws on a set of beliefs and evaluations which may, for example, influence their mental health [182] or learning [174], as examples. The issue of how self-beliefs - which may span several trials of a task, or even several tasks within the day - are built from local metacognitive evaluations has received interest. Rouault, Dayan and Fleming [183] described global metacognition as self-performance estimates and proposed that global metacognition is 'aggregated' over time from local confidence, emphasizing the role of external feedback (see also [184]). In return, global estimates before task completion may also influence metacognitive bias in local confidence as both are correlated [12]. Recent models of self-esteem [185] have indeed proposed a two-way relationship between multiple levels of beliefs.

Global metacognition draws upon generalised beliefs and rules of thumb as well as experience-based judgements, and taken at face value, this issue should help us address the notion of domain-generality. On a theoretical level, the notion of a generalized belief about function based on an agglomeration of local judgements seems like it should operate at a superordinate level in metacognition. However, there are too few studies to date to support this view. The field would benefit more from experiments that test the idea that local

judgements for one task are extrapolated up into a form of metacognitive awareness at a global level for a different kind of task. Here, however, it seems reasonable to propose that metacognitive bias, and indeed self-efficacy may produce a spurious domain-generality. That is, if a participant finds one task easy, this may spill over into their belief about other tasks, or as argued above, a general arousal level (for example) may produce a spurious form of domain-generality. As we argue for decisional judgements, the extent to which there is transfer from one global evaluation of a task to another may rest upon how related the two first-order tasks are.

In sum, if the generalized global metacognitive evaluation is appropriate and not simply a factor which influences both first and second-order performance, we would conclude that it is domain-general. Moreover, if there is a cue or second-order output from one domain that can be systematically transferred to a global metacognition across tasks, that would argue for a strong form of domain-generality.

For adecisional global evaluations, the idea that a predisposition or a rule of thumb guides judgements would imply a form of domain specificity based on the knowledge of the task: a prediction based on recalling about half the items from a list of words, for example. However, for statements such as our example above, of being better at detecting your brother at the airport than remembering his birthday, it is hard to see how such a cross-domain comparison could be made without some form of domain-general comparison common currency, but this is again something that needs more empirical investigation. A further point of interest is to consider the process of aggregation which in our conceptualisation leads to accumulation of decisional metacognitive evaluations into a global evaluation which does not pertain to any one decision (and hence is 'adecisional'). Whether we can extrapolate across different tasks in different domains to form one generalised belief or sense of expertise which is aggregated across different domains is an interesting question.

5. Concluding section

The domain-generality of metacognition is of importance to understand its architecture and to establish how metacognition extends across different types of processes in real-world decision-making and in different pathologies. If metacognition were domain-general, it would mean that training of metacognition on one domain might mitigate difficulties in metacognition in another cognitive domain.

In order to create a common conceptual space in which to consider metacognitive processes, we offer a number of possible configurations of the metacognitive architecture and emphasize the division into decisional and adecisional forms. Our conceptualisation of decisional metacognition is critical for constraining the types of cues and experiences that are defined as metacognitive, and means that we can better consider perceptual and memory tasks in parallel. Adecisional metacognition groups together beliefs, knowledge about functioning and strategic regulation in a way which is not strictly second-order: the evaluation is not derivative of any process in hand, but an estimation of task factors. Thus, the extent to which metacognition is domain-general in the real world rests upon the balance of decisional and adecisional factors, and this will change according to context and task demands, meaning that isolating specific domain-general metacognitive processes requires narrowing the conceptual space.

Because adecisional metacognition considers processes that could be dissociated across domains, perhaps the clearest evidence in this review for domain-generality comes from decisional metacognition (Section 3). We report cross-domain correlations for both metacognitive bias and metacognitive efficiency for such decisional processes, even though the magnitude of the latter seems low. Neuropsychological studies, in contrast, posit that domain-specific processes are at play in decisional metacognition, pointing toward a weak-domain-generality (Figure 1b). When metacognitive measures do not control for first-order performance their correlation across tasks may reflect shared first-order components across domains (e.g., task difficulty), leading to spurious domain-generality. However, even when accounting for first-order performance, it is not always straightforward to ascribe the influence of a factor to first- or second-order processes.

What therefore should we assess if we wish to establish a strong domain-generality of metacognition? Thus far, with a few exceptions, the main approach has been to index second-order processes by using retrospective confidence judgements: cross-task correlations yield between-subject correlations for confidence judgements. This is an appropriate first step in accessing domain-generality, but conclusions from this approach are limited by the fact that first-order and adecisional factors can produce spurious correlations across tasks with extremely similar structures and task demands. The use of between-subject correlations can confirm domain-general patterns, but is not suitable for hypothesis testing about the architecture of metacognition. Our conclusion is that if a strong version of a domain-general metacognition is to be found, it will be in mechanisms which apply across tasks of diverse types and structures, such as when reaction times are estimated to build

confidence. Because reaction times are present in any decision-making across a variety of tasks, finding that confidence is estimated from reaction times is evidence for domain-generality (although again first-order strategies influencing reaction time can also result in spurious domain-generality, [119]). Described like this, a domain-general architecture would involve a 'final common pathway' for all types of decisional metacognitive evaluations, and the possibility of a common process computed from processing times is an exciting development for the field.

To go beyond correlations between domains, we need to tackle more complex experimental paradigms. A first step would be to consider richer tasks that include multiple experimental conditions. If participants behave similarly not just across domains but also for all the conditions, this would be converging evidence for domain-generality. For instance, [186] showed that the sense of confidence is shaped by the behavioural goal that participants are set to achieve similarly in a visual perception and value-based decisions. Another approach is what we can call metacognitive transfer, in analogy to Bayesian transfer that has been proposed as a test of Bayesian decision theory to be a good model of human visual perception [187]. Testing metacognitive transfer would consist in identifying some key metacognitive features in two domains (say, domain 'A' for perception and 'B' for memory). We argue that if this metacognitive feature generalizes across domains, then this is good evidence in favor of domain-generality. For instance, the metacognitive feature can be the change in efficiency in evaluating global confidence judgments for two different set sizes, labeled '1' and '2'. If there is domain-generality, measuring metacognitive performance in one modality (i.e. for combinations 'A-1' and 'A-2') should allow us to predict metacognitive performance for the other modality (for combinations 'B-1' and 'B-2'). In these last two examples (richer tasks and metacognitive transfer), it is clear that disposing of better metacognitive models that work across domains would be highly beneficial.

Because neuropsychological data have contributed to the studies reviewed above, a final concrete suggestion for future research in special populations would be two-fold: firstly to test metacognition systematically across domains and not just for the impaired first order function. Secondly, in a related manner, for groups with a metacognitive deficit we should ensure that metacognition is tested on tasks which do not show a first order deficit (either through manipulating difficulty or by extending the scope of the study to other domains). Ideally, if metacognitive proficiency relies on a final common pathway, it should be possible to find a patient or a group of patients with impaired metacognitive access across all domains.

In conclusion, our common conceptual space for metacognition proposes a decisional process based on the outputs of first-order processing, a view of metacognitive function that can be common across tasks and domains. Thus far, the evidence for a domain-general metacognition favours a weak version of domain-generality, with some processes shared between tasks across different domains. Whilst it is necessary to constrain the problem space to decisional metacognition, in contrast, it is now necessary to test the architectures proposed in this review across a set of more diverse metacognitive tasks and also at mechanistic, implementational and neural levels.

Author contributions

All authors discussed the original idea and common conceptual framework. Joint first authors, Audrey Mazancieux and Michael Pereira took the lead in writing the manuscript. All authors provided feedback and contributed to the final version.

Competing interests

The authors declare no conflicts of interests.

Box 1: Direct measures of metacognition

Global judgements of metacognition are perhaps the easiest to conceptualise, and involve making estimations of performance at a task level, e.g. how many words will be correctly recalled from a list [162], how many words from a particular category were generated in a given period [138] or even a prediction of percentage grade achieved in a university exam [188]. A critical factor in these methods is the point at which the prediction is made; typically before and after completing the task. This means that predictions of future performance can be made for example having completed the university course and before the exam, but also once the exam is completed. Moreover, an initial prediction before the task starts (or before the university course has been taught) acts as a reference point for interpreting the global prediction made after experiencing the task, with any shift between an initial and informed prediction being based on the capacity to monitor the task and with the initial - pre-task prediction being based on expectancies, beliefs and metacognitive knowledge [176]. Global judgements do not refer to any one decision, and as such reveal processes used in generalised evaluations of function, rather than pinpointing metacognitive mechanisms and processes.

Ease of Learning (EOL) Judgments [189] are predictions about what will be easy or difficult to learn and pertain to items that have not yet been learned. Participants are therefore not asked to learn the items for an upcoming memory test when making their judgment. On the presentation of each item, the common question asked to the participants is: “How likely is it that you will learn this word for the test?”. After the EOL judgments participants are asked to study the items and recall them. Despite the importance of this initial assessment of how difficult a material is and its potential impact on learning, most studies suggest that EOL judgments poorly or moderately predict the actual learnability of to-be-learned material [190–192] but increase when items vary enough in difficulty [193].

Judgement of Learning (JOL) takes the same form as an EOL judgement, with the exception that it is made once the item has been processed in the encoding phase of a memory task, i.e. after having attempted to learn the item. Typically, the JOL is made for cue-target word pairs, with the metacognitive judgement being made for the likelihood of retrieving a target when prompted by the cue word. It is, like the EOL, a judgment of the likelihood of subsequent recall and has likewise been used to explore the cues which are used to make metacognitive judgements of memory, but equally factors which influence recall (for meta-analytic reviews see [194]). Of note, JOLs can be made immediately after the encoding of the item, or after a delay (either in a second phase after the initial block of encoding of word

pairs, or more typically, after several intervening cue-target pairs). The delayed JOL effect is a robust phenomenon whereby metacognitive sensitivity is higher for judgements made after a delay than immediately [195]. Another well-established JOL phenomenon is the font-size effect, whereby JOL magnitude is increased by font-size: words written in a large font at encoding are judged more likely to be recalled than words in a smaller font, even though this factor influences recall to a lesser extent (for an explanatory meta-analysis see [196]).

Feeling-of-knowing (FOK) judgments (FOK, [30]) are predictions about the likelihood of subsequent recognition of currently non-recalled information [29,197]. In a FOK experiment participants are presented either with new information to learn such as word pairs (episodic memory task) or are presented with general knowledge questions such as 'what is the capital of France?' (semantic memory task). When presented with the question or the first word of the pair participants are asked to recall the corresponding information. If they cannot recall the information, the FOK judgment consists in asking participants to predict whether they will be able to recognize the missing information if presented to them later. Thus, FOK judgements are predictions about material that participants failed to retrieve and, although not perfect, these judgements have been found to be relatively accurate in young adults [29,198,199].

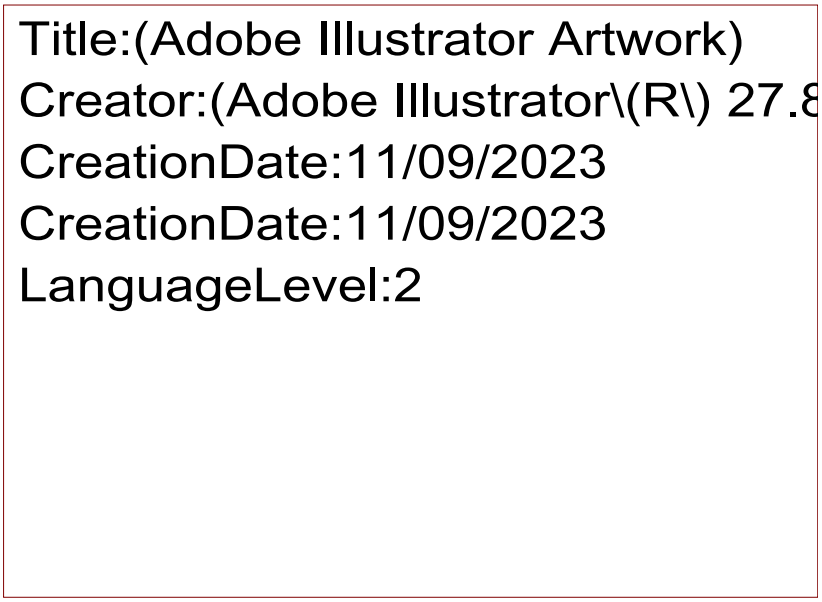
Retrospective confidence judgements (RCJs) are the main measure of metacognition used in the field. They refer to the level of confidence that a participant has in a given answer using a multiple-point scale. They have been extensively used in decision-making (Grimaldi et al., 2015) notably to investigate cross-domain comparisons [9,10,39] but also in other tasks such as statistical learning [200].

The *confidence forced-choice paradigm* [201] requires participants to choose which of two decisions made about two different stimuli is more likely to be correct. By varying the difficulty levels within pairs of stimuli, one can estimate a psychometric function for chosen vs. declined decisions. The difference in slopes between these two curves serves as a proxy for metacognitive performance, irrespective of confidence bias. By asking participants to choose between decisions that pertain to different cognitive domains, this method is instrumental to characterize the domain-generality of metacognition [202].

Box 2: Indirect of measures metacognition

In indirect measures of metacognition, participants are not directly asked for a self-evaluation but evaluate other behaviours that are used to infer metacognition. For instance in *post-decision wagering*, participants have to place bets on the correctness of their decisions. Early versions of this paradigm were used as an objective measure of subjective visibility [203], assuming that participants would place higher bets following seen vs. unseen stimuli. Concerns were raised that this measure was affected by loss aversion [204] and pertained more to metacognitive access than to subjective visibility [205]. As a result, more recent developments are now used to incentivise optimal metacognitive judgements irrespective of loss aversion [206,207]. *Opt-out paradigms* consist in allowing participants to opt out of a decision if their putative confidence in the choice is low, providing a proxy for low confidence. This procedure has the advantage of being applicable to non-verbal species such as non-human primates [78], rodents [73] or preverbal infants [208]. In some versions of the paradigm, an opt-out response option is provided [78] while in others, a delay is imposed between the response and the reward, during which participants can opt out and restart a new trial without reward [73,208]. In the latter case, the waiting time can be used as a continuous proxy to confidence. Evidence for domain-specific metacognition has been recently found using this task with non-human primates [209]. The main criticism is that participants could achieve such behavior by simple reinforcement learning without relying on a second-order monitoring mechanism [210], see [211,212] for a more general overview of animal metacognition. In memory, specific indirect measures have been developed. When two tests using the same material are performed, one can either measure the time a participant would re-study an item (*study-time allocation*, [213]) or the decision of re-studying an item or not (*re-study choice*, [214]). In this context, it has been shown that participants allocate more time to re-study an item that they did not previously recall compared to recalled items, leading to the idea that they have accurate knowledge about previous failures [192].

Box 3: How to quantify metacognition



Metacognitive sensitivity relates to the ability of an individual to adjust a decisional judgment (Table 1; typically confidence) to the performance of the first-order task. Therefore, measures of metacognition are not directly informative of metacognitive sensitivity. Confidence gaps compare average confidence judgments after correct versus incorrect first-order responses but overlook the variance of the two distributions. One workaround is to assess the relationship between confidence judgments and first-order performance, using Pearson's or Goodman–Kruskall gamma correlations [215]. Unfortunately, these methods cannot isolate metacognitive sensitivity from metacognitive *bias* [22,37]. This issue can be avoided by computing the area under the receiving operating characteristic (AUROC) curve. A two-dimensional curve is constructed by computing the percentage of correct responses classified as such (true-positive rate; vertical axis) and the percentage of incorrect responses classified as correct (false-negative rate; horizontal axis) by setting different thresholds on the confidence. The area under this curve ranges from 0.5 (chance level) and 1.0 when there is a threshold that perfectly classifies correct and incorrect responses. AUROC curves can dissociate metacognitive sensitivity from bias. Metacognitive *efficiency* refers to the metacognitive sensitivity given the information available to the first-order decision. To compare metacognitive efficiency using AROCs, task performance should be titrated across conditions and participants. Another possibility that does not require titration is to use a model-based approach and compare first-order sensitivity (d') and an estimated (meta-)d' given an "ideal observer" model of confidence (i.e. without metacognitive noise) [74]. Most recent studies use the ratio between meta-d' and d' termed the M-ratio [42]. Finally, new approaches are being proposed that attempt to fit metacognitive noise with a generative

model of confidence judgments [216], including in situations where the model includes other parameters that can be easily confounded with metacognitive noise [58].

Box 4: Motor metacognition

Although most studies on metacognition have focused on the perceptual and memory domains, the motor domain has seen a recent surge of interest. An early attempt to characterize motor metacognition found that participants largely misjudged the effects of distorting visual feedback of their hand trajectories although they appropriately them [217], suggesting that we have a limited access to the details of our movements as the goal is achieved [218,219]. Similar metacognitive inefficiencies are found when participants are asked to manually track a dynamic noisy visual stimulus [220]. Other studies showed that participants judiciously adjust their confidence about detecting distortions, which seems to contradict the notion of a limited access to motor performance discussed above [221]. This contradiction may be resolved by considering that participants optimally calibrate their confidence, even when they fail to report distorsions, based on a summary statistic of the visual feedback [222]. In other words, one can automatically correct distortions that remain undetected, and still have a subjective feeling for performance informing confidence. Although this heuristic may be useful for automatically monitoring motor actions, one can ask if we can explicitly monitor low-level movement parameters when prompted to do so. When throwing a virtual ball, participants can monitor their performance based both on the position of the arm as well as the resulting trajectory of the ball thrown [95]. It is also important to keep in mind that participants may be better at monitoring some aspects of their motor actions (e.g. the movement duration) than others (e.g. when they initiated the movement; [223]). Finally, studying motor metacognition is also of interest regarding the distinction between the monitoring of internal (e.g., mnesic) and external (e.g., sensory) signals, as both efferent and reafferent signals might serve as objects for meta-representations. However, there was only an effect of confidence bias for active versus passive finger movement when distinguishing the contribution of efferent and reafferent signals on confidence, suggesting that efferent signals increase confidence but do not improve the quality of metacognitive monitoring [224]. In the same vein, actions paired to visual stimuli were found to lead to higher confidence ratings, but to leave metacognitive performance unchanged [105]. These results suggest that confidence ratings do not improve based on efferent information.

Box 5: Anosognosia

Metacognition has long been studied in neurological disorders for motor, sensory, and cognitive deficits, through the lens of self-awareness. For example, motor awareness has been explored since Babinski's proposal of the term anosognosia [225] initially suggested to describe patients unaware of the existence of paralysis. The picture is however complex with some patients failing to acknowledge one deficit but recognize another (e.g. upper but not lower limb paralysis, [226]), or patients failing to adapt their behavior according to their knowledge (e.g. admitting their deficits but yet attempting to walk, or denying deficits whilst remaining in bed; [227,228]).

In the sensory domain, despite being amongst one the rarest neurological conditions, the Anton-Babinski Syndrome (ABS) merits to be cited here. Indeed, patients with ABS, also called visual anosognosia present with binocular visual loss with denial of this blindness and a relatively well-preserved cognition. Since Babinski's initial work, hundreds of scientific papers have been published [229] and the concept has also been used to describe unawareness of cognitive functions or lack of cognitive insight in neurological and psychiatric disorders such as frontal lobe lesions [128], Alzheimer's disease (see [230] for a review) Parkinson's disease [231,232], or schizophrenia [233]. Those studies have revealed a multifaceted view of anosognosia with many examples of patients showing that unawareness can affect different cognitive domains. For example, patients with frontal lobe lesions often demonstrate what is called 'utilization behavior' [234]. In this peculiar situation, patients will engage in a stereotypical action in the sight of an object, despite not being explicitly asked to (e.g., starting to use a stapler put on a desk). According to Blakemore and al. [218] this behavior would be explained by a lack of awareness of goals and intentions. Domain-generality or domain-specificity of anosognosia has been explored in several models across domains. Agnew and Morris [235] and later Morris and Hannesdottir [236] proposed a model of anosognosia for memory disorders: the *Cognitive Awareness Model* (CAM). Similarly, to the DICE model [237], this model posits the existence of a separate awareness system, the *Metacognitive Awareness System*, which provides conscious awareness of ability or error. In a more recent version of this model, Clare et al. [238] proposed that domain-specific monitoring processes are situated at a lower level and refer to the Cognitive Comparator mechanisms (CCMs). The CMMs' role would be to compare recent errors to previous experiences in each domain, leading to a more global self-representation of one's own abilities. Neuropsychological studies and models, therefore, seem to predict a general awareness system or central supervisory system, which once disconnected would lead to anosognosia across domains. On the other hand, an impairment of the comparator (CMM) would lead to specific anosognosia. Interestingly, in the motor domain, Blakemore, Wolpert & Firth [218] also predicted the existence of a comparator system in their Comparator Model of motor control

in which it is predicted that for each movement, an individual implicitly monitors their intentions and predicted outcome in relation to sensory and perceptual feedback about the actual outcome. This comparison forges the detection of a discordance that would occur in the context of a movement error. In the anosognosia literature, unawareness of hemiplegia would therefore be explained by a discrepancy in monitoring between one's intentions (i.e., motor plan) and one's actual motor performance [226,239,240]. To conclude, neuropsychological studies posit that domain-specific processes are at play in decisional metacognition, pointing toward a weak domain-generality.

References

1. Flavell, J. H. Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *Am. Psychol.* **34**, 906–911 (1979).
2. Fleur, D. S., Bredeweg, B. & van den Bos, W. Metacognition: ideas and insights from neuro- and educational sciences. *NPJ Sci Learn* **6**, 13 (2021).
3. Evans, J. R. & Fisher, R. P. Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Appl. Cogn. Psychol.* **25**, 501–508 (2011).
4. Pannu, J. K. & Kaszniak, A. W. Metamemory experiments in neurological populations: a review. *Neuropsychol. Rev.* **15**, 105–130 (2005).
5. Rosenthal, D. M. Consciousness, content, and metacognitive judgments. *Conscious. Cogn.* **9**, 203–214 (2000).
6. Brown, R., Lau, H. & LeDoux, J. E. Understanding the Higher-Order Approach to Consciousness. *Trends Cogn. Sci.* **23**, 754–768 (2019).
7. Ais, J., Zylberberg, A., Barttfeld, P. & Sigman, M. Individual consistency in the accuracy and distribution of confidence judgments. *Cognition* **146**, 377–386 (2016).
8. Faivre, N., Filevich, E., Solovey, G., Kühn, S. & Blanke, O. Behavioral, Modeling, and Electrophysiological Evidence for Supramodality in Human Metacognition. *J. Neurosci.* **38**, 263–277 (2018).
9. Lee, A. L. F., Ruby, E., Giles, N. & Lau, H. Cross-Domain Association in Metacognitive Efficiency Depends on First-Order Task Types. *Front. Psychol.* **9**, 2464 (2018).
10. Mazancieux, A., Fleming, S. M., Souchay, C. & Moulin, C. J. A. Is there a G factor for metacognition? Correlations in retrospective metacognitive sensitivity across tasks. *J. Exp. Psychol. Gen.* **149**, 1788–1799 (2020).
11. Mazancieux, A., Dinze, C., Souchay, C. & Moulin, C. J. A. Metacognitive domain specificity in feeling-of-knowing but not retrospective confidence. *Neurosci Conscious* **2020**, niaa001 (2020).

12. McWilliams, A., Bibbey, H., Steinbeis, N., David, A. S. & Fleming, S. M. Age-related decreases in global metacognition are independent of local metacognition and task performance. *PsyArXiv* (2022) doi:10.31234/osf.io/nmhxv.
13. Rouault, M., McWilliams, A., Allen, M. G. & Fleming, S. M. Human metacognition across domains: insights from individual differences and neuroimaging. *Personality Neuroscience* **1**, (2018).
14. Mamassian, P. Visual Confidence. *Annu Rev Vis Sci* **2**, 459–481 (2016).
15. Grimaldi, P., Lau, H. & Basso, M. A. There are things that we know that we know, and there are things that we do not know we do not know: Confidence in decision-making. *Neurosci. Biobehav. Rev.* **55**, 88–97 (2015).
16. Hoven, M. *et al.* Abnormalities of confidence in psychiatry: an overview and future perspectives. *Transl. Psychiatry* **9**, 268 (2019).
17. Koriat, A. The feeling of knowing: some metatheoretical implications for consciousness and control. *Conscious. Cogn.* **9**, 149–171 (2000).
18. Koriat, A. Metacognition and consciousness. in *The Cambridge handbook of consciousness , (pp* (ed. Zelazo, P. D.) vol. 981 289–325 (Cambridge University Press, xiv, 2007).
19. Nelson, T. O. & Narens, L. Why investigate metacognition. *Metacognition: Knowing about knowing* **13**, 1–25 (1994).
20. Fleming, S. M., Dolan, R. J. & Frith, C. D. Metacognition: computation, biology and function. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 1280–1286 (2012).
21. Koriat, A. The self-consistency model of subjective confidence. *Psychol. Rev.* **119**, 80–113 (2012).
22. Fleming, S. M. & Lau, H. C. How to measure metacognition. *Front. Hum. Neurosci.* **8**, 443 (2014).
23. Marr, D. Vision: A computational approach (san fr. Preprint at (1982).
24. Vaccaro, A. G. & Fleming, S. M. Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain Neurosci Adv* **2**,

2398212818810591 (2018).

25. Fu, Z. *et al.* The geometry of domain-general performance monitoring in the human medial frontal cortex. *Science* **376**, eabm9922 (2022).
26. Dunlosky, J. & Tauber, S. U. K. The Oxford handbook of metamemory. (2016).
27. Arbuckle, T. Y. & Cuddy, L. L. Discrimination of item strength at time of presentation. *J. Exp. Psychol.* **81**, 126–131 (1969).
28. Rhodes, M. G. & Tauber, S. K. The influence of delaying judgments of learning on metacognitive accuracy: a meta-analytic review. *Psychol. Bull.* **137**, 131–148 (2011).
29. Schacter, D. L. Feeling of knowing in episodic memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **9**, 39 (1983).
30. Hart, J. T. Memory and the feeling-of-knowing experience. *J. Educ. Psychol.* **56**, 208–216 (1965).
31. Koriat, A. How do we know that we know? The accessibility model of the feeling of knowing. *Psychol. Rev.* **100**, 609–639 (1993).
32. Norman, E., Blakstad, O., Johnsen, Ø., Martinsen, S. K. & Price, M. C. The Relationship between Feelings-of-Knowing and Partial Knowledge for General Knowledge Questions. *Front. Psychol.* **7**, 996 (2016).
33. Fleming, S. M., Massoni, S., Gajdos, T. & Vergnaud, J.-C. Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neurosci Conscious* **2016**, niw018 (2016).
34. Mei, N., Rankine, S., Olafsson, E. & Soto, D. Similar history biases for distinct prospective decisions of self-performance. *Sci. Rep.* **10**, 5854 (2020).
35. Barrett, A. B., Dienes, Z. & Seth, A. K. Measures of metacognition on signal-detection theoretic models. *Psychol. Methods* **18**, 535–552 (2013).
36. Benjamin, A. S. & Diaz, M. Measurement of relative metamnemonic accuracy. *Handbook of memory and metamemory* 73–94 (2008).
37. Masson, M. E. J. & Rotello, C. M. Sources of bias in the Goodman–Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes.

J. Exp. Psychol. Learn. Mem. Cogn. **35**, 509–527 (2009).

38. Sherman, M., Barrett, A. B. & Kanai, R. Inferences about consciousness using subjective reports of confidence. in *Behavioral Methods in Consciousness Research* 87–106 (Oxford University Press, 2015).
39. Morales, J., Lau, H. & Fleming, S. M. Domain-General and Domain-Specific Patterns of Activity Supporting Metacognition in Human Prefrontal Cortex. *J. Neurosci.* **38**, 3534–3546 (2018).
40. Benwell, C. S. Y., Mohr, G., Wallberg, J., Kouadio, A. & Ince, R. A. A. Psychiatrically relevant signatures of domain-general decision-making and metacognition in the general population. *npj Mental Health Research* **1**, 1–17 (2022).
41. Mazor, M., Friston, K. J. & Fleming, S. M. Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *Elife* **9**, (2020).
42. Fleming, S. M. HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neurosci Conscious* **2017**, nix007 (2017).
43. Paulewicz, B. & Blaut, A. The bhsdtr package: a general-purpose method of Bayesian inference for signal detection theory models. *Behav. Res. Methods* **52**, 2122–2141 (2020).
44. Hu, X., Yang, C. & Luo, L. Are the contributions of processing experience and prior beliefs to confidence ratings domain-general or domain-specific? *J. Exp. Psychol. Gen.* (2022) doi:10.1037/xge0001257.
45. Guggenmos, M. Measuring metacognitive performance: type 1 performance dependence and test-retest reliability. *Neurosci Conscious* **2021**, niab040 (2021).
46. Xue, K., Shekhar, M. & Rahnev, D. Examining the robustness of the relationship between metacognitive efficiency and metacognitive bias. *Conscious. Cogn.* **95**, 103196 (2021).
47. de Gardelle, V., Le Corre, F. & Mamassian, P. Confidence as a Common Currency between Vision and Audition. *PLoS One* **11**, e0147901 (2016).
48. Fleming, S. M., Ryu, J., Golfinos, J. G. & Blackmon, K. E. Domain-specific impairment in

metacognitive accuracy following anterior prefrontal lesions. *Brain* **137**, 2811–2822 (2014).

49. Sadeghi, S., Ekhtiari, H., Bahrami, B. & Ahmadabadi, M. N. Metacognitive Deficiency in a Perceptual but Not a Memory Task in Methadone Maintenance Patients. *Sci. Rep.* **7**, 7052 (2017).
50. Ye, Q., Zou, F., Lau, H., Hu, Y. & Kwok, S. C. Causal Evidence for Mnemonic Metacognition in Human Precuneus. *J. Neurosci.* **38**, 6379–6387 (2018).
51. Ye, Q. *et al.* Individual susceptibility to TMS affirms the precuneal role in meta-memory upon recollection. *Brain Struct. Funct.* **224**, 2407–2419 (2019).
52. Beatty, W. W. & Monson, N. Metamemory in multiple sclerosis. *J. Clin. Exp. Neuropsychol.* **13**, 309–327 (1991).
53. Wojcik, D. Z., Moulin, C. J. A. & Souchay, C. Metamemory in children with autism: Exploring 'feeling-of-knowing' in episodic and semantic memory. *Neuropsychology* **27**, 19–27 (2013).
54. Gottesman, I. I. & Gould, T. D. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am. J. Psychiatry* **160**, 636–645 (2003).
55. Maniscalco, B., Peters, M. A. K. & Lau, H. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Atten. Percept. Psychophys.* **78**, 923–937 (2016).
56. Jang, Y., Wallsten, T. S. & Huber, D. E. A stochastic detection and retrieval model for the study of metacognition. *Psychol. Rev.* **119**, 186–200 (2012).
57. Fleming, S. M. & Daw, N. D. Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychol. Rev.* **124**, 91–114 (2017).
58. Mamassian, P. & de Gardelle, V. Modeling perceptual confidence and the confidence forced-choice paradigm. *Psychol. Rev.* **129**, 976–998 (2022).
59. Balsdon, T., Wyart, V. & Mamassian, P. Confidence controls perceptual evidence accumulation. *Nat. Commun.* **11**, 1753 (2020).
60. Balsdon, T., Mamassian, P. & Wyart, V. Separable neural signatures of confidence

during perceptual decisions. *Elife* **10**, (2021).

61. Shadlen, M. N. & Kiani, R. Decision Making as a Window on Cognition. *Neuron* **80**, 791–806 (2013).
62. Van Zandt, T. ROC curves and confidence judgments in recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **26**, 582–600 (2000).
63. Ratcliff, R. & Starns, J. J. Modeling confidence and response time in recognition memory. *Psychol. Rev.* **116**, 59–83 (2009).
64. Insabato, A., Pannunzi, M., Rolls, E. T. & Deco, G. Confidence-related decision making. *J. Neurophysiol.* **104**, 539–547 (2010).
65. Wei, Z. & Wang, X.-J. Confidence estimation as a stochastic process in a neurodynamical system of decision making. *J. Neurophysiol.* **114**, 99–113 (2015).
66. Maniscalco, B. *et al.* Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. *PLoS Comput. Biol.* **17**, e1008779 (2021).
67. Atiya, N. A. A., Huys, Q. J. M., Dolan, R. J. & Fleming, S. M. Explaining distortions in metacognition with an attractor network model of decision uncertainty. *PLoS Comput. Biol.* **17**, e1009201 (2021).
68. Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion Decision Model: Current Issues and History. *Trends Cogn. Sci.* **20**, 260–281 (2016).
69. Churchland, A. K., Kiani, R. & Shadlen, M. N. Decision-making with multiple alternatives. *Nat. Neurosci.* **11**, 693–702 (2008).
70. Yonelinas, A. P. The nature of recollection and familiarity: A review of 30 years of research. *J. Mem. Lang.* **46**, 441–517 (2002).
71. Wixted, J. T. Dual-process theory and signal-detection theory of recognition memory. *Psychol. Rev.* **114**, 152–176 (2007).
72. Treisman, M. & Faulkner, A. The setting and maintenance of criteria representing levels of confidence. *J. Exp. Psychol. Hum. Percept. Perform.* **10**, 119–139 (1984).
73. Kepecs, A., Uchida, N., Zariwala, H. A. & Mainen, Z. F. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231 (2008).

74. Maniscalco, B. & Lau, H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* **21**, 422–430 (2012).
75. Maniscalco, B. & Lau, H. The signal processing architecture underlying subjective reports of sensory awareness. *Neurosci Conscious* **2016**, (2016).
76. Shekhar, M. & Rahnev, D. The nature of metacognitive inefficiency in perceptual decision making. *Psychol. Rev.* (2021).
77. Pouget, A., Drugowitsch, J. & Kepecs, A. Confidence and certainty: distinct probabilistic quantities for different goals. *Nat. Neurosci.* **19**, 366–374 (2016).
78. Kiani, R. & Shadlen, M. N. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* **324**, 759–764 (2009).
79. Meyniel, F., Sigman, M. & Mainen, Z. F. Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron* **88**, 78–92 (2015).
80. Sherman, M. T., Seth, A. K., Barrett, A. B. & Kanai, R. Prior expectations facilitate metacognition for perceptual decision. *Conscious. Cogn.* **35**, 53–65 (2015).
81. Constant, M., Salomon, R. & Filevich, E. Judgments of agency are affected by sensory noise without recruiting metacognitive processing. *Elife* **11**, (2022).
82. Hu, X. *et al.* A Bayesian inference model for metamemory. *Psychol. Rev.* **128**, 824–855 (2021).
83. Pleskac, T. J. & Busemeyer, J. R. Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychol. Rev.* **117**, 864–901 (2010).
84. Desender, K., Donner, T. H. & Verguts, T. Dynamic expressions of confidence within an evidence accumulation framework. *Cognition* **207**, 104522 (2021).
85. Moran, R., Teodorescu, A. R. & Usher, M. Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cogn. Psychol.* **78**, 99–147 (2015).
86. Pereira, M. *et al.* Evidence accumulation relates to perceptual consciousness and monitoring. *Nat. Commun.* **12**, 3261 (2021).

87. Pereira, M., Perrin, D. & Faivre, N. A leaky evidence accumulation process for perceptual experience. *Trends Cogn. Sci.* **26**, 451–461 (2022).
88. King, J.-R. & Dehaene, S. A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130204 (2014).
89. Vickers, D. Uncertainty, Choice, and the Marginal Efficiencies. *J. Post Keynes. Econ.* **2**, 240–254 (1979).
90. van den Berg, R. *et al.* A common mechanism underlies changes of mind about decisions and confidence. *Elife* **5**, e12192 (2016).
91. Zylberberg, A., Barttfeld, P. & Sigman, M. The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.* **6**, (2012).
92. Rahnev, D., Nee, D. E., Riddle, J., Larson, A. S. & D'Esposito, M. Causal evidence for frontal cortex organization for perceptual decision making. *Proceedings of the National Academy of Sciences of the United States of America* vol. 113 6059–6064 (2016).
93. Pereira, M. *et al.* Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8382–8390 (2020).
94. Landry, M., Da Silva Castanheira, J., Sackur, J. & Raz, A. Investigating how the modularity of visuospatial attention shapes conscious perception using type I and type II signal detection theory. *J. Exp. Psychol. Hum. Percept. Perform.* **47**, 402–422 (2021).
95. Arbuzova, P. *et al.* Measuring metacognition of direct and indirect parameters of voluntary movement. *J. Exp. Psychol. Gen.* **150**, 2208–2229 (2021).
96. Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J. & Rees, G. Relating introspective accuracy to individual differences in brain structure. *Science* **329**, 1541–1543 (2010).
97. Koizumi, A., Maniscalco, B. & Lau, H. Does perceptual confidence facilitate cognitive control? *Atten. Percept. Psychophys.* **77**, 1295–1306 (2015).
98. Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J. & Postle, B. R. Dissociating Perceptual Confidence from Discrimination Accuracy Reveals No Influence of

Metacognitive Awareness on Working Memory. *Front. Psychol.* **7**, 851 (2016).

99. Peters, M. A. K. *et al.* Perceptual confidence neglects decision-incongruent evidence in the brain. *Nature human behaviour* **1**, (2017).
100. Zylberberg, A., Fetsch, C. R. & Shadlen, M. N. The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *Elife* **5**, (2016).
101. Yon, D., Zainzinger, V., de Lange, F. P., Eimer, M. & Press, C. Action biases perceptual decisions toward expected outcomes. *J. Exp. Psychol. Gen.* **150**, 1225–1236 (2021).
102. Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G. & Davranche, K. Revealing subthreshold motor contributions to perceptual confidence. *Neurosci Conscious* **2019**, niz001 (2019).
103. Dotan, D., Pinheiro-Chagas, P., Al Roumi, F. & Dehaene, S. Track It to Crack It: Dissecting Processing Stages with Finger Tracking. *Trends Cogn. Sci.* **23**, 1058–1070 (2019).
104. Faivre, N. *et al.* Sensorimotor conflicts alter metacognitive and action monitoring. *Cortex* **124**, 224–234 (2020).
105. Filevich, E., Koß, C. & Faivre, N. Response-Related Signals Increase Confidence But Not Metacognitive Performance. *eNeuro* **7**, (2020).
106. Siedlecka, M., Paulewicz, B. & Wierzchoń, M. But I Was So Sure! Metacognitive Judgments Are Less Accurate Given Prospectively than Retrospectively. *Front. Psychol.* **7**, 218 (2016).
107. Fleming, S. M. *et al.* Action-specific disruption of perceptual confidence. *Psychol. Sci.* **26**, 89–98 (2015).
108. Patel, D., Fleming, S. M. & Kilner, J. M. Inferring subjective states through the observation of actions. *Proceedings of the Royal Society of London B: Biological Sciences* **279**, 4853–4860 (2012).
109. Vuillaume, L., Martin, J.-R., Sackur, J. & Cleeremans, A. Comparing self- and hetero-metacognition in the absence of verbal communication. *PLoS One* **15**, e0231530 (2020).

110. Hanczakowski, M., Butowska, E., Philip Beaman, C., Jones, D. M. & Zawadzka, K. The dissociations of confidence from accuracy in forced-choice recognition judgments. *J. Mem. Lang.* **117**, 104189 (2021).
111. Rahnev, D. A., Maniscalco, B., Luber, B., Lau, H. & Lisanby, S. H. Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J. Neurophysiol.* **107**, 1556–1563 (2012).
112. Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E. & Lau, H. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci.* **1**, 165–175 (2010).
113. Chiang, T.-C., Lu, R.-B., Hsieh, S., Chang, Y.-H. & Yang, Y.-K. Stimulation in the dorsolateral prefrontal cortex changes subjective evaluation of percepts. *PLoS One* **9**, e106943 (2014).
114. Bor, D., Schwartzman, D. J., Barrett, A. B. & Seth, A. K. Theta-burst transcranial magnetic stimulation to the prefrontal or parietal cortex does not impair metacognitive visual awareness. *PLoS One* **12**, e0171793 (2017).
115. Gogulski, J., Zetter, R., Nyrhinen, M., Pertovaara, A. & Carlson, S. Neural Substrate for Metacognitive Accuracy of Tactile Working Memory. *Cereb. Cortex* **27**, 5343–5352 (2017).
116. Baranski, J. V. & Petrusic, W. M. Probing the locus of confidence judgments: experiments on the time to determine confidence. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 929–945 (1998).
117. Rahnev, D. *et al.* The Confidence Database. *Nat Hum Behav* **4**, 317–325 (2020).
118. Kiani, R., Corthell, L. & Shadlen, M. N. Choice Certainty Is Informed by Both Evidence and Decision Time. *Neuron* **84**, 1329–1342 (2014).
119. Desender, K., Vermeylen, L. & Verguts, T. Dynamic influences on static measures of metacognition. *Nat. Commun.* **13**, 4208 (2022).
120. Koriat, A. Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *J. Exp. Psychol. Gen.* **126**, 349–370 (1997).

121. Metcalfe, J., Schwartz, B. L. & Joaquim, S. G. The cue-familiarity heuristic in metacognition. *J. Exp. Psychol. Learn. Mem. Cogn.* **19**, 851–861 (1993).
122. Alter, A. L. & Oppenheimer, D. M. Uniting the tribes of fluency to form a metacognitive nation. *Pers. Soc. Psychol. Rev.* **13**, 219–235 (2009).
123. Oppenheimer, D. M. The secret life of fluency. *Trends Cogn. Sci.* **12**, 237–241 (2008).
124. Benjamin, A. S. & Bjork, R. A. Retrieval fluency as a metacognitive index. *Implicit memory and metacognition* (2014) doi:10.4324/9781315806136-19/retrieval-fluency-metacognitive-index-aaron-benjamin-robert-bjork.
125. Kelley, C. M. & Lindsay, D. S. Remembering Mistaken for Knowing: Ease of Retrieval as a Basis for Confidence in Answers to General Knowledge Questions. *J. Mem. Lang.* **32**, 1–24 (1993).
126. Fleming, S. M., van der Putten, E. J. & Daw, N. D. Neural mediators of changes of mind about perceptual decisions. *Nat. Neurosci.* **21**, 617–624 (2018).
127. Shekhar, M. & Rahnev, D. Sources of Metacognitive Inefficiency. *Trends Cogn. Sci.* **25**, 12–23 (2021).
128. Janowsky, J. S., Shimamura, A. P. & Squire, L. R. Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology* **17**, 3–11 (1989).
129. Shimamura, A. P. & Squire, L. R. Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. *J. Exp. Psychol. Learn. Mem. Cogn.* **12**, 452–460 (1986).
130. Schnyer, D. M. *et al.* A role for right medial prefrontal cortex in accurate feeling-of-knowing judgments: evidence from patients with lesions to frontal cortex. *Neuropsychologia* **42**, 957–966 (2004).
131. Pinon, K., Allain, P., Kefi, M. Z., Dubas, F. & Le Gall, D. Monitoring processes and metamemory experience in patients with dysexecutive syndrome. *Brain Cogn.* **57**, 185–188 (2005).
132. Bäckman, L. & Lipinska, B. Monitoring of general knowledge: evidence for preservation

in early Alzheimer's disease. *Neuropsychologia* **31**, 335–345 (1993).

133. Lipinska, B. & Bäckman, L. Feeling-of-knowing in fact retrieval: further evidence for preservation in early Alzheimer's disease. *J. Int. Neuropsychol. Soc.* **2**, 350–358 (1996).

134. Moulin, C. J., Perfect, T. J. & Jones, R. W. Evidence for intact memory monitoring in Alzheimer's disease: metamemory sensitivity at encoding. *Neuropsychologia* **38**, 1242–1250 (2000).

135. Moulin, C. J. A., James, N., Perfect, T. J. & Jones, R. W. Knowing What You Cannot Recognise: Further Evidence for Intact Metacognition in Alzheimer's Disease. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* **10**, 74–82 (2003).

136. Pappas, B. A. *et al.* Alzheimer's disease and feeling-of-knowing for knowledge and episodic memory. *J. Gerontol.* **47**, P159–64 (1992).

137. Souchay, C., Isingrini, M., Pillon, B. & Gil, R. Metamemory accuracy in Alzheimer's disease and frontotemporal lobe dementia. *Neurocase* **9**, 482–492 (2003).

138. Mazancieux, A., Moulin, C. J. A., Casez, O. & Souchay, C. A Multidimensional Assessment of Metacognition Across Domains in Multiple Sclerosis. *J. Int. Neuropsychol. Soc.* **27**, 124–135 (2021).

139. Souchay, C., Isingrini, M. & Gil, R. Metamemory monitoring and Parkinson's disease. *J. Clin. Exp. Neuropsychol.* **28**, 618–630 (2006).

140. Souchay, C. & Smith, S. J. Subjective states associated with retrieval failures in Parkinson's disease. *Conscious. Cogn.* **22**, 795–805 (2013).

141. Bhome, R. *et al.* Metacognition in functional cognitive disorder. *Brain Commun* **4**, fcac041 (2022).

142. Demant, K. M., Vinberg, M., Kessing, L. V. & Miskowiak, K. W. Assessment of subjective and objective cognitive function in bipolar disorder: Correlations, predictors and the relation to psychosocial function. *Psychiatry Res.* **229**, 565–571 (2015).

143. Hauser, T. U., Allen, M., NSPN Consortium, Rees, G. & Dolan, R. J. Publisher Correction: Metacognitive impairments extend perceptual decision making weaknesses in compulsivity. *Sci. Rep.* **8**, 6046 (2018).

144. Rouault, M., Seow, T., Gillan, C. M. & Fleming, S. M. Psychiatric Symptom Dimensions Are Associated With Dissociable Shifts in Metacognition but Not Task Performance. *Biol. Psychiatry* **84**, 443–451 (2018).

145. Hoven, M. *et al.* Metacognition and the effect of incentive motivation in two compulsive disorders: Gambling disorder and obsessive-compulsive disorder. *Psychiatry Clin. Neurosci.* **76**, 437–449 (2022).

146. Moritz, S. *et al.* Sowing the seeds of doubt: a narrative review on metacognitive training in schizophrenia. *Clin. Psychol. Rev.* **34**, 358–366 (2014).

147. Dietrichkeit, M., Grzella, K., Nagel, M. & Moritz, S. Using virtual reality to explore differences in memory biases and cognitive insight in people with psychosis and healthy controls. *Psychiatry Res.* **285**, 112787 (2020).

148. Gawęda, Ł. & Moritz, S. The role of expectancies and emotional load in false auditory perceptions among patients with schizophrenia spectrum disorders. *Eur. Arch. Psychiatry Clin. Neurosci.* **271**, 713–722 (2021).

149. Moritz, S., Woodward, T. S. & Rodriguez-Raecke, R. Patients with schizophrenia do not produce more false memories than controls but are more confident in them. *Psychol. Med.* **36**, 659–667 (2006).

150. Berna, F., Zou, F., Danion, J.-M. & Kwok, S. C. Overconfidence in false autobiographical memories in patients with schizophrenia. *Psychiatry research* vol. 279 374–375 (2019).

151. Rouy, M. *et al.* Systematic review and meta-analysis of metacognitive abilities in individuals with schizophrenia spectrum disorders. *Neurosci. Biobehav. Rev.* **126**, 329–337 (2021).

152. Ko, Y. & Lau, H. A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 1401–1411 (2012).

153. Balsdon, T. & Azzopardi, P. Absolute and relative blindsight. *Conscious. Cogn.* **32**, 79–91 (2015).

154. Grainger, C., Williams, D. M. & Lind, S. E. Metacognition, metamemory, and mindreading in high-functioning adults with autism spectrum disorder. *J. Abnorm. Psychol.* **123**, 650–659 (2014).
155. Grainger, C., Williams, D. M. & Lind, S. E. Metacognitive monitoring and control processes in children with autism spectrum disorder: Diminished judgement of confidence accuracy. *Conscious. Cogn.* **42**, 65–74 (2016).
156. Carpenter, K. L. & Williams, D. M. A meta-analysis and critical review of metacognitive accuracy in autism. *Autism* 13623613221106004 (2022).
157. Souchay, C., Guillery-Girard, B., Pauly-Takacs, K., Wojcik, D. Z. & Eustache, F. Subjective experience of episodic memory and metacognition: a neurodevelopmental approach. *Front. Behav. Neurosci.* **7**, 212 (2013).
158. Ansell, E. L. & Bucks, R. S. Mnemonic anosognosia in Alzheimer's disease: a test of Agnew and Morris (1998). *Neuropsychologia* **44**, 1095–1102 (2006).
159. Moulin, C. J. A. *et al.* Retrieval-induced forgetting in Alzheimer's disease. *Neuropsychologia* **40**, 862–867 (2002).
160. Budson, A. E. *et al.* Memory and emotions for the september 11, 2001, terrorist attacks in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. *Neuropsychology* **18**, 315–327 (2004).
161. Barrett, A. M., Eslinger, P. J., Ballentine, N. H. & Heilman, K. M. Unawareness of cognitive deficit (cognitive anosognosia) in probable AD and control subjects. *Neurology* **64**, 693–699 (2005).
162. Moulin, C. J. A., Perfect, T. J. & Jones, R. W. Global predictions of memory in Alzheimer?S disease: Evidence for preserved metamemory monitoring. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* **7**, 230–244 (2000).
163. Moulin, C. & de La Rochefoucauld, F. D. 9 Sense and sensitivity: metacognition in. *Applied metacognition* 197 (2002).
164. Morris, R. G. & Mograbi, D. C. Anosognosia, autobiographical memory and self knowledge in Alzheimer's disease. *Cortex* **49**, 1553–1565 (2013).

165. Davies, M., Davies, A. A. & Coltheart, M. Anosognosia and the Two-factor Theory of Delusions. *Mind Lang.* **20**, 209–236 (2005).
166. Lyons, K. E. & Ghetti, S. Metacognitive Development in Early Childhood: New Questions about Old Assumptions. in *Trends and Prospects in Metacognition Research* (eds. Efklides, A. & Misailidi, P.) 259–278 (Springer US, 2010).
167. Veenman, M. V. J. & Spaans, M. A. Relation between intellectual and metacognitive skills: Age and task differences. *Learn. Individ. Differ.* **15**, 159–176 (2005).
168. Geurten, M., Meulemans, T. & Lemaire, P. From domain-specific to domain-general? The developmental path of metacognition for strategy selection. *Cogn. Dev.* **48**, 62–81 (2018).
169. Vo, V. A., Li, R., Kornell, N., Pouget, A. & Cantlon, J. F. Young children bet on their numerical skills: metacognition in the numerical domain. *Psychol. Sci.* **25**, 1712–1721 (2014).
170. Bellon, E., Fias, W. & De Smedt, B. Metacognition across domains: Is the association between arithmetic and metacognitive monitoring domain-specific? *PLoS One* **15**, e0229932 (2020).
171. Shavelson, R. J. & Hubner, J. J. Self-concept: Validation of construct interpretations. *Rev. Educ. Res.* (1976).
172. Dapp, L. C. & Roebers, C. M. Metacognition and self-concept: Elaborating on a construct relation in first-grade children. *PLoS One* **16**, e0250845 (2021).
173. Connor, L. T., Dunlosky, J. & Hertzog, C. Age-related differences in absolute but not relative metamemory accuracy. *Psychol. Aging* **12**, 50–71 (1997).
174. Händel, M., de Bruin, A. B. H. & Dresel, M. Individual differences in local and global metacognitive judgments. *Metacognition and Learning* **15**, 51–75 (2020).
175. Bertrand, J. M. *et al.* In the here and now: Short term memory predictions are preserved in Alzheimer's disease. *Cortex* **119**, 158–164 (2019).
176. Silva, A. R., Pinho, M. S., Macedo, L., Souchay, C. & Moulin, C. Mnemonic anosognosia in Alzheimer's disease is caused by a failure to transfer online evaluations of

performance: Evidence from memory training programs. *J. Clin. Exp. Neuropsychol.* **39**, 419–433 (2017).

177. Hertzog, C., Saylor, L. L., Fleece, A. M. & Dixon, R. A. Metamemory and aging: Relations between predicted, actual and perceived memory task performance. *Neuropsychol. Dev. Cogn. B Aging Neuropsychol. Cogn.* **1**, 203–237 (1994).
178. Moulin, C. Sense and sensitivity: Metacognition in Alzheimer's disease. *Applied metacognition* 197–223 (2002).
179. Luna, K. & Martín-Luengo, B. Confidence-Accuracy Calibration with General Knowledge and Eyewitness Memory Cued Recall Questions. *Appl. Cogn. Psychol.* **26**, 289–295 (2012).
180. Perfect, T. J. & Hollins, T. S. Predictive feeling of knowing judgements and postdictive confidence judgements in eyewitness memory and general knowledge. *Appl. Cogn. Psychol.* **10**, 371–382 (1996).
181. Perfect, T. J. The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Appl. Cogn. Psychol.* **18**, 157–168 (2004).
182. Seow, T. X. F., Rouault, M., Gillan, C. M. & Fleming, S. M. How Local and Global Metacognition Shape Mental Health. *Biol. Psychiatry* **90**, 436–446 (2021).
183. Rouault, M., Dayan, P. & Fleming, S. M. Forming global estimates of self-performance from local confidence. *Nat. Commun.* **10**, 1141 (2019).
184. Lee, A. L. F., de Gardelle, V. & Mamassian, P. Global visual confidence. *Psychon. Bull. Rev.* **28**, 1233–1242 (2021).
185. De Ruiter, N. M. P., Van Geert, P. L. C. & Kunnen, E. S. Explaining the 'How' of Self-Esteem Development: The Self-Organizing Self-Esteem Model. *Rev. Gen. Psychol.* **21**, 49–68 (2017).
186. Sepulveda, P. *et al.* Visual attention modulates the integration of goal-relevant evidence and not value. *Elife* **9**, (2020).
187. Maloney, L. T. & Mamassian, P. Bayesian decision theory as a model of human visual perception: testing Bayesian transfer. *Vis. Neurosci.* **26**, 147–155 (2009).

188. Morphew, J. W. Changes in metacognitive monitoring accuracy in an introductory physics course. *Metacogn. Learn.* **16**, 89–111 (2021).

189. Nelson, T. O. Metamemory: A Theoretical Framework and New Findings. in *Psychology of Learning and Motivation* (ed. Bower, G. H.) vol. 26 125–173 (Academic Press, 1990).

190. Mazzoni, G., Cornoldi, C., Tomat, L. & Vecchi, T. Remembering the grocery shopping list: A study on metacognitive biases. *Appl. Cogn. Psychol.* **11**, 253–267 (1997).

191. McCarley, J. S. & Gosney, J. Metacognitive Judgments in a Simulated Luggage Screening Task. *Proc. Hum. Fact. Ergon. Soc. Annu. Meet.* **49**, 1620–1624 (2005).

192. Son, L. K. & Metcalfe, J. Metacognitive and control strategies in study-time allocation. *J. Exp. Psychol. Learn. Mem. Cogn.* **26**, 204–221 (2000).

193. Jemstedt, A., Kubik, V. & Jönsson, F. U. What moderates the accuracy of ease of learning judgments? *Metacognition and Learning* **12**, 337–355 (2017).

194. Luna, K., Martín-Luengo, B. & Albuquerque, P. B. Do delayed judgements of learning reduce metamemory illusions? A meta-analysis. *Q. J. Exp. Psychol.* **71**, 1626–1636 (2018).

195. Metcalfe, J. & Finn, B. Evidence that judgments of learning are causally related to study choice. *Psychon. Bull. Rev.* **15**, 174–179 (2008).

196. Chang, M. & Brainerd, C. J. Association and dissociation between judgments of learning and memory: A Meta-analysis of the font size effect. *Metacogn Learn* **17**, 443–476 (2022).

197. Souchay, C., Isingrini, M. & Espagnet, L. Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology* **14**, 299–309 (2000).

198. Gruneberg, M. M. & Monks, J. 'Feeling of knowing' and cued recall. *Acta Psychol.* **38**, 257–265 (1974).

199. Kelemen, W. L., Frost, P. J. & Weaver, C. A., 3rd. Individual differences in metacognition: evidence against a general metacognitive ability. *Mem. Cognit.* **28**, 92–107 (2000).

200. Meyniel, F., Schlunegger, D. & Dehaene, S. The Sense of Confidence during

Probabilistic Learning: A Normative Account. *PLoS Comput. Biol.* **11**, e1004305 (2015).

201. Barthelmé, S. & Mamassian, P. Evaluation of objective uncertainty in the visual system. *PLoS Comput. Biol.* **5**, e1000504 (2009).
202. de Gardelle, V. & Mamassian, P. Does Confidence Use a Common Currency Across Two Visual Tasks? *Psychol. Sci.* **25**, 1286–1288 (2014).
203. Persaud, N., McLeod, P. & Cowey, A. Post-decision wagering objectively measures awareness. *Nat. Neurosci.* **10**, 257–261 (2007).
204. Schurger, A. & Sher, S. Awareness, loss aversion, and post-decision wagering. *Trends in cognitive sciences* vol. 12 209–10; author reply 210 (2008).
205. Seth, A. K. Post-decision wagering measures metacognitive content, not sensory consciousness. *Conscious. Cogn.* **17**, 981–983 (2008).
206. Dienes, Z. & Seth, A. Gambling on the unconscious: a comparison of wagering and confidence ratings as measures of awareness in an artificial grammar task. *Conscious. Cogn.* **19**, 674–681 (2010).
207. Massoni, S., Gajdos, T. & Vergnaud, J.-C. Confidence measurement in the light of signal detection theory. *Front. Psychol.* **5**, 1455 (2014).
208. Goupil, L. & Kouider, S. Behavioral and Neural Indices of Metacognitive Sensitivity in Preverbal Infants. *Curr. Biol.* **26**, 3038–3045 (2016).
209. Cai, Y. *et al.* Time-sensitive prefrontal involvement in associating confidence with task performance illustrates metacognitive introspection in monkeys. *Commun Biol* **5**, 799 (2022).
210. Le Pelley, M. E. Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 686–708 (2012).
211. Beran, M. Animal metacognition: A decade of progress, problems, and the development of new prospects. *Anim. Behav. Cogn.* **6**, 223–229 (2019).
212. Proust, J. P. From comparative studies to interdisciplinary research on metacognition. *Anim. Behav. Cogn.* **6**, 309–328 (2019).

213. Dunlosky, J. & Ariel, R. Self-regulated learning and the allocation of study time. in *Advances in Research and Theory* (ed. Ross, B. H.) vol. 54 103–140 (Elsevier, 2011).
214. Finley, J. R., Tullis, J. G. & Benjamin, A. S. Metacognitive Control of Learning and Remembering. in *New Science of Learning: Cognition, Computers and Collaboration in Education* (eds. Khine, M. S. & Saleh, I. M.) 109–131 (Springer New York, 2010).
215. Nelson, T. O. A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychol. Bull.* **95**, 109–133 (1984).
216. Guggenmos, M. Reverse engineering of metacognition. *Elife* **11**, (2022).
217. Fourneret, P. & Jeannerod, M. Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* **36**, 1133–1140 (1998).
218. Blakemore, S. J., Wolpert, D. M. & Frith, C. D. Abnormalities in the awareness of action. *Trends Cogn. Sci.* **6**, 237–242 (2002).
219. Custers, R. & Aarts, H. The unconscious will: how the pursuit of goals operates outside of conscious awareness. *Science* **329**, 47–50 (2010).
220. Locke, S. M., Mamassian, P. & Landy, M. S. Performance monitoring for sensorimotor confidence: A visuomotor tracking study. *Cognition* **205**, 104396 (2020).
221. Sinanaj, I., Cojan, Y. & Vuilleumier, P. Inter-individual variability in metacognitive ability for visuomotor performance and underlying brain structures. *Conscious. Cogn.* **36**, 327–337 (2015).
222. Pereira, M., Skiba, R., Cojan, Y., Vuilleumier, P. & Bègue, I. Optimal confidence for unaware visuomotor deviations. *bioRxiv* 2021.10.22.465492 (2021) doi:10.1101/2021.10.22.465492.
223. Jovanovic, L. & López□Moliner, J. Contrasting contributions of movement onset and duration to self□evaluation of sensorimotor timing performance. *European Journal of* (2021).
224. Charles, L., Chardin, C. & Haggard, P. Evidence for metacognitive bias in perception of voluntary action. *Cognition* **194**, 104041 (2020).
225. Babinski, J. Contribution a l'étude des troubles mentaux dans l'hémiplégie organique

(anosognosie). *Rev. Neurol.* (1914).

226. Berti, A., Làdavas, E. & Della Corte, M. Anosognosia for hemiplegia, neglect dyslexia, and drawing neglect: clinical findings and theoretical considerations. *J. Int. Neuropsychol. Soc.* **2**, 426–440 (1996).

227. Jehkonen, M., Laihosalo, M. & Kettunen, J. E. Impact of neglect on functional outcome after stroke – a review of methodological issues and recent research findings. *Restor. Neurol. Neurosci.* **24**, 209–215 (2006).

228. Nimmo-Smith, I., Marcel, A. J. & Tegnér, R. A diagnostic test of unawareness of bilateral motor task abilities in anosognosia for hemiplegia. *J. Neurol. Neurosurg. Psychiatry* **76**, 1167–1169 (2005).

229. Jenkinson, P. M. & Fotopoulou, A. Understanding Babinski's anosognosia: 100 years later. *Cortex; a journal devoted to the study of the nervous system and behavior* vol. 61 1–4 (2014).

230. Souchay, C. Metamemory in Alzheimer's disease. *Cortex* **43**, 987–1003 (2007).

231. Amanzio, M. *et al.* Impaired awareness of movement disorders in Parkinson's disease. *Brain Cogn.* **72**, 337–346 (2010).

232. Jenkinson, P. M., Edelstyn, N. M. J., Stephens, R. & Ellis, S. J. Why are some Parkinson disease patients unaware of their dyskinesias? *Cogn. Behav. Neurol.* **22**, 117–121 (2009).

233. Riggs, S. E., Grant, P. M., Perivoliotis, D. & Beck, A. T. Assessment of cognitive insight: a qualitative review. *Schizophr. Bull.* **38**, 338–350 (2012).

234. Lhermitte, F. 'Utilization behaviour'and its relation to lesions of the frontal lobes. *Brain* **106**, 237–255 (1983).

235. Agnew, S. K. & Morris, R. G. The heterogeneity of anosognosia for memory impairment in Alzheimer's disease: A review of the literature and a proposed model. *Aging Ment. Health* **2**, 7–19 (1998).

236. Morris, R. G. & Hannesdottir, K. Loss of 'awareness' in Alzheimer's disease. *Cognitive neuropsychology of Alzheimer's disease* 275–296 (2004).

237. Schacter, D. L. Toward a cognitive neuropsychology of awareness: implicit knowledge and anosognosia. *J. Clin. Exp. Neuropsychol.* **12**, 155–178 (1990).
238. Clare, L., Marková, I. S., Roth, I. & Morris, R. G. Awareness in Alzheimer's disease and associated dementias: theoretical framework and clinical implications. *Aging Ment. Health* **15**, 936–944 (2011).
239. Fotopoulou, A. *et al.* The role of motor intention in motor awareness: an experimental study on anosognosia for hemiplegia. *Brain* **131**, 3432–3442 (2008).
240. Moro, V., Pernigo, S., Zapparoli, P., Cordioli, Z. & Aglioti, S. M. Phenomenology and neural correlates of implicit and emergent motor awareness in patients with anosognosia for hemiplegia. *Behav. Brain Res.* **225**, 259–269 (2011).