Confidence and metacognition

Kiyofumi Miyoshi[a], **Taylor Webb**[b], **Dobromir Rahnev**[c], **and Hakwan Lau**[d], [a] Graduate School of Informatics, Kyoto University, Kyoto, Japan; [b] Department of Psychology, University of California, Los Angeles, CA, United States; [c] School of Psychology, Georgia Institute of Technology, Atlanta, GA, United States; and [d] RIKEN Center for Brain Science, Saitama, Japan

Key points

- Metacognitive monitoring enables adaptive behavioral control by facilitating learning, error detection, resource allocation, and effective interpersonal communication.
- Subjective confidence does not always align with the accuracy of objective decisions, but this divergence may find explanations through the concept of ecological rationality.
- Computational studies of metacognition enable the formal measurement of metacognitive performance and provide insights into the information processing that underlies the formation of confidence.
- The sensory cortex encodes the intensity and uncertainty of sensory information, on which the frontoparietal network represents decision confidence under specific task requirements.
- Confidence remains consistent in various contexts, such as perception, memory, and value-based decision-making, facilitating information utilization across diverse modalities.
- The pursuit of metacognition leads us into the realm of consciousness studies, potentially shedding light on the fundamental origins of our conscious experiences.

Abstract

Humans are not passive recipients of information from the external world. They are “metacognitive” agents, actively engaged in self-referential monitoring of internally expressed information. This introspective insight in turn enables proactive behavioral adjustments, facilitating efficient learning, preventing erroneous actions, and promoting information exchange with others. The study of metacognition has witnessed a recent expansion, incorporating a fusion of computational modeling, neuroimaging, machine learning techniques, and other interdisciplinary approaches. This article serves as an invitation for readers to explore the introspective realm of metacognition, providing an overview of the latest research findings.

https://doi.org/10.1016/B978-0-12-820480-1.00049-8

Introduction

Flexible and adaptive behavior often requires the consideration not only of events in the external world, but also one's own internal mental processes. For example, consider the process of deciding how long to study for an upcoming test, based on a self-assessment of one's own understanding. This process of reflecting on and regulating one's cognitive processes is commonly known as metacognition (e.g., Brown, 1978; Fleming et al., 2012; Flavell, 1979; Nelson and Narens, 1994).

Research on metacognition has been initiated in multiple fields of behavioral science, including perceptual psychophysics, experimental pedagogy, and cognitive psychology (Clarke et al., 1959; Flavell, 1971; Koriat et al., 1980; Nelson and Narens, 1990). While findings across these fields are interrelated and share many commonalities, the breadth of scientific knowledge on metacognition is extensive.

This article specifically focuses on the matter of decision confidence, which refers to a subjective estimate of the likelihood that a decision will be correct. Since decision confidence is a fundamental concept that underlies various metacognitive functions, placing emphasis on it provides the current article with an appropriate scope and structure that extends across related disciplines. But this also means that we would unfortunately be less focused on other important aspects of metacognition, such as those involved in top-down control (Nelson and Narens, 1990, 1994). Those topics may be covered better in other articles of this volume, such as those about attention and cognitive control.

Studying decision confidence is important for at least two reasons. First, the subjective experience of confidence is underpinned by unique computational and neural mechanisms that differ from those involved in objective decision-making (e.g., King and Dehaene, 2014; Shekhar and Rahnev, 2021a). Second, decision confidence has practical significance both for the adaptive control of behavior and information exchange between individuals (e.g., Bahrami et al., 2010; Boldt et al., 2019; Cortese, 2022). In the subsequent sections, we delve deeper into these reasons, highlighting the multifaceted implications of decision confidence. By exploring the internal realm of metacognitive introspection, intriguing research avenues emerge that go beyond the classical paradigm of treating decision-making agents as mere stimulus-response devices.

Subjective confidence versus objective decision

Traditional research on decision-making, especially perceptual psychophysics, has concentrated on investigating how individuals understand the external world. It aims to describe the relationship between the external input individuals receive and the decisions they make based on that input. This type of decision-making is commonly referred to as objective decision-making or type-1 judgment. In addition to objective decision-making, individuals possess the capacity to assess their own internal states, as exemplified by confidence ratings. This self-referential assessment is referred to as metacognitive decision-making or type-2 judgment (Clarke et al., 1959; Galvin et al., 2003).

While these terms are primarily used in the field of perceptual decision-making, similar concepts have been widely applied in other domains, such as the object-level versus meta-level distinction introduced in memory research (Nelson and Narens, 1990, 1994). Although this dichotomous perspective may appear as a simplification of the problem, it carries an important implication that lays the ground for scientific research on metacognition. That is, the problem of subjective introspection, which may seem elusive on its own, can be characterized in its relation to objective decisions (Clarke et al., 1959; Galvin et al., 2003; Nelson and Narens, 1994).

Importantly, a growing body of evidence has shown that confidence is not always a straightforward reflection of objective decision accuracy (e.g., Shekhar and Rahnev, 2021a; Windschitl and Chambers, 2004; Zylberberg et al., 2012). A well-known example is the so-called hard-easy effect, where decision-making agents tend to be under confident in easy tasks and overconfident in difficult tasks (Baranski and Petrusic, 1994; Griffin and Tversky, 1992). Furthermore, through behavioral manipulations, brain stimulation, or lesions, decision confidence can be selectively influenced while the accuracy of objective decisions remains unaffected (Fleming et al., 2014; Maniscalco and Lau, 2015; Rounis et al., 2010; Ruby et al., 2018). Strikingly, selective effects on confidence have also been demonstrated in animal models, through both behavioral manipulations and pharmacological interventions (Lak et al., 2014; Miyamoto et al., 2017; Odegaard et al., 2018; Stolyarova et al., 2019).

These findings emphasize that subjective confidence is not merely a byproduct of decision-making but rather a distinct component, stimulating the investigation of its underlying mechanisms. Moreover, while these observed discrepancies may suggest an apparently suboptimal nature of the metacognitive system, in fact they could also have a rational basis (Miyoshi and Lau, 2020; Webb et al., 2023). By studying the underlying rationale behind these dissociations, we may gain valuable insights into the potential functional role of conscious introspection (Michel, 2023; Morales et al., 2022; Peters, 2022).

Functional roles of confidence

Cognitive functions such as perception, memory, learning, and decision-making are not isolated subsystems that operate on their own. Instead, they are likely regulated by a process of self-directed metacognitive introspection. According to Nelson and Narens (1994), the regulatory functions of metacognition can be categorized into two main components, monitoring and control. Monitoring involves gathering information at the object level, while control entails adjusting ongoing object-level processing based on the acquired information. Confidence, therefore, corresponds to the monitoring of the likelihood that one's decision is correct, which may offer a foundation for adaptive behavioral control.

The influence of confidence, along with related introspective processes, on behavioral adjustments has been studied initially within the field of educational psychology. Researchers in this field demonstrated how individuals allocate their learning time based on estimations of task difficulty and their own knowledge level (e.g., Lovelace, 1984; Son and Metcalfe, 2000; Thiede and Dunlosky, 1999).

In recent years, the functional roles of confidence have been demonstrated across broader fields of behavioral science (e.g., Boldt and Gilbert, 2019; Rollwage et al., 2020; Rollwage and Fleming, 2021). For example, confidence has been identified as a mediator in the information exploration-exploitation trade-off, helping individuals strike a balance between seeking new information and exploiting familiar resources (Boldt et al., 2019). Confidence also mediates multi-sensory cue integration, where visual stimulation that leads to higher subjective confidence exerts greater influence on auditory perception (Gao et al., 2023). Confidence can also serve as a priority signal, guiding individuals to prioritize tasks they feel more confident about (Aguilar-Lleyda et al., 2020). Additionally, confidence has shown potential as a teaching signal for reinforcement learning, enabling learning without the need for explicit rewards (Daniel and Pollmann, 2012; Guggenmos et al., 2016; Cortese et al., 2020). Furthermore, dysfunction in metacognitive regulatory functions has been associated with conditions such as addiction and schizophrenia (Hoven et al., 2019; Rouault et al., 2018a; Rouy et al., 2021). It is therefore likely that the study of decision confidence will yield insights into these pathologies, which could aid in understanding their underlying mechanisms.

In addition to its role in adjusting one's own behavior, confidence itself holds substantial importance in information transmission within a social context. In situations involving collective decision-making with multiple individuals, weighting the opinions of each person based on their confidence levels has been shown to lead to more accurate decision-making (Bahrami et al., 2010). Furthermore, in the context of collecting eyewitness testimony on crime scenes, the confidence expressed by witnesses is considered a crucial indicator of the testimony's credibility (Seale-Carlisle et al., 2019; Wixted and Wells, 2017). Consequently, guidelines have been explored to ensure appropriate methods for collecting and evaluating eyewitness testimony with confidence rating (Wells et al., 2020; Colloff et al., 2021).

As evident from these examples, confidence plays crucial roles in guiding adaptive behavior, operating through computational mechanisms separate from those for objective decision-making. The study of confidence extends beyond the conventional research paradigm of decision-making and encompasses a broader set of research areas, including both laboratory investigations and real-world applications.

Need for theoretical frameworks

The use of theoretical frameworks is essential in studying metacognition. Distinguishing between the most promising frameworks and models is seen as central to the progress of the field (Rahnev et al., 2022). These frameworks, rooted in probabilistic modeling, are employed because decision-making in living organisms is inherently subject to uncertainty. They allow researchers to quantitatively assess task performance and draw statistical inferences about the underlying processes related to metacognition. As a result, they facilitate a continuous cycle of empirical testing and theory development.

Signal detection theory

The signal detection theory (SDT) originated as a method for evaluating radar systems that distinguish meaningful signals from random noise. By the early 1950s, it was adapted for experimental psychology, providing a framework to describe stochastic decision-making behaviors under uncertainty (Peterson et al., 1954; Tanner and Swets, 1954).

SDT assumes that sensory input from the external world gives rise to an internal response, represented as a latent variable for decision-making. Due to inherent noise in the biological brain, even the same sensory input can generate responses of varying strength. **Fig. 1** gives a SDT description of a decision-making task, such as determining whether a visual stimulus is moving left or right against a noisy background. According to the stimulus direction, latent variables across multiple trials can be described by a pair of normal distributions, representing the internal evidence for rightward versus leftward movement. A decision of "right" is made for those trials where observed evidence exceeds an internal criterion value (solid vertical line in **Fig. 1**).

The criterion separates the density of the distributions into four regions, which correspond to the probabilities of different response outcomes (true left, false left, false right, true right). Inversely, by analyzing human response rate data, one can estimate the model parameters and evaluate the sensitivity of objective decision-making as the distance between the two distributions, referred to as d′ (Green and Swets, 1966; Macmillan and Creelman, 2005).

In addition, the SDT framework provides methods for describing how people rate their decision confidence. One of these methods involves using the distance of a latent variable from the objective decision criterion as a basis for confidence (e.g., Galvin et al., 2003). **Fig. 1** illustrates this concept, where criteria for confidence rating (dashed lines) are depicted together with the objective decision criterion (solid line). These criteria partition the distribution density into eight regions, each representing the probability of its associated response outcome (4 response classes mentioned above × 2 classes of confidence).

Through this modeling, one can quantify the metacognitive sensitivity of observers, which indicates how well their confidence distinguishes between their own correct and incorrect decisions. Namely, by analyzing human confidence data for correct and incorrect decisions, metacognitive sensitivity can be measured as the distance between two latent distributions, known as meta-d′

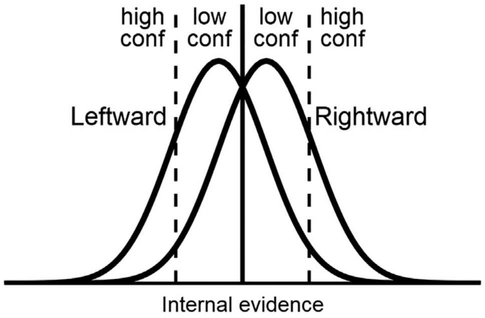

Fig. 1 Signal detection theory model applied to left/right direction discrimination. The distribution depicted on the left (or right) represents the trial-by-trial internal evidence under the physical presence of leftward (or rightward) motion. In each trial, a “left” (or “right”) decision is made if observed evidence falls left (or right) of a decision criterion, depicted as a solid vertical line. The distance from the decision criterion can serve as an indicator of decision confidence, and dashed vertical lines illustrate the criteria to assign low or high confidence ratings.

(Maniscalco and Lau, 2012; Fleming and Lau, 2014). The significance of meta-d′ lies in its direct comparability to d′, allowing for the measurement of metacognitive sensitivity in relation to objective decision accuracy.

While the meta-d′ model is currently considered the standard for metacognitive measurement, recent papers have introduced a variety of measurement models that expand upon the classical SDT framework (Boundy-Singer et al., 2023; Dayan, 2023; Guggenmos, 2022; Miyoshi and Nishida, 2022; Miyoshi et al., 2022; Shekhar and Rahnev, 2021b). These models can be broadly classified into descriptive or process-model approaches. The descriptive approach, including the meta-d′ framework, offers an ad-hoc analysis of observed data, whereas the process-model approach aims to quantify metacognitive performance by explicitly specifying the process through which confidence is computed.

Regardless of the chosen approach, it is essential for metacognition measures to possess specific desirable properties. As in the case of meta-d′, metacognitive measures should be disentangled from objective decision accuracy (as well as other potential confounding factors like individuals’ propensity for high or low confidence ratings). Additionally, these measures should exhibit robust validity and reliability. Rahnev (2023) examined various metacognitive measures and identified their respective strengths and limitations. He found that none of these measures fully satisfied all the desired properties. Furthermore, no measurement approach at present exhibited a clear advantage over the meta-d′ framework, rendering it a reasonable interim option. The measurement of metacognitive performance remains a topic of lively debate, and more research is anticipated in the future.

Sequential sampling framework

Another prominent framework focuses on the temporal accumulation of decision evidence (e.g., Ratcliff et al., 2016). A leading example is the drift diffusion model (Ratcliff, 1978), depicted in Fig. 2 within the context of left/right direction discrimination. In this model, internal evidence is added up at each time step, and a decision is reached when the cumulative sum crosses thresholds set for “left” and “right” decisions.

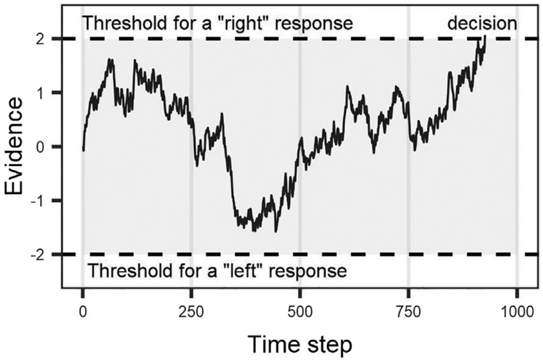

Fig. 2 Drift diffusion model applied to left/right direction discrimination. At each time step, noisy internal evidence about stimulus direction evolves with the positive (or negative) sign favoring rightward (or leftward) movement. A decision is reached when accumulated evidence crosses upper or lower decision boundaries (set at ± 2 in this depiction), representing the “right” and “left” decisions. Here, the means and the standard deviation of the step-by-step evidence variability are set at ± 0.005 and 0.1, respectively. This corresponds to assuming a SDT model of $d' = 0.1$ for each time step. The accumulation process constitutes an optimal decision variable, tracking the log-likelihood ratio (see section **Bayesian confidence**) in this signal detection process over all previous time steps.

This accumulation process aligns with the sequential probability ratio test in statistics, which minimizes expected time steps required to achieve a predefined level of accuracy (e.g., Bogacz et al., 2006). The diffusion model carries significant methodological implications, particularly for neuroscience research, suggesting that the temporal dynamics of optimal decision-making can be represented by unidimensional quantity, such as cumulatively increasing neural firing (Gold and Shadlen, 2000, 2001; Kiani and Shadlen, 2009).

The simple diffusion model, however, encounters a limitation when it comes to explaining decision confidence. In its original form, evidence accumulation in each trial stops at the constant threshold value, resulting in no meaningful variation for confidence ratings. Consequently, this limitation has inspired an idea that additional evidence gathering after making a decision contributes to the generation of confidence (Pleskac and Busemeyer, 2010). The concept of post-decisional confidence construction has received robust support from empirical studies (Moreira et al., 2018; Yu et al., 2015), guiding subsequent investigations of confidence construction mechanisms (Fleming and Daw, 2017; Maniscalco et al., 2021; Navajas et al., 2016). Nevertheless, other research suggests that it is unlikely that confidence is formed exclusively after a decision has been made and instead confidence is likely already being computed during decision formation (Dotan et al., 2018; Chen and Rahnev, 2023; Xue et al., 2023a).

Importantly, diverse model families are present within the sequential sampling framework (e.g., Brown and Heathcote, 2008; Ratcliff and Starns, 2013; Tajima et al., 2019). While the diffusion model family typically assumes a single accumulator that integrates evidence from all stimuli, another prominent family, known as race models, employs a separate accumulator for each stimulus. In race models, each accumulator aims to reach its own threshold, and a decision is made when one of the accumulators reaches its threshold (e.g., Ratcliff et al., 2016).

The race model family has been used to explain confidence since Vickers (1979), proposing that the balance between winning and losing accumulators determines confidence. More recently, efforts have been made to expand race models to enhance our understanding of confidence formation processes. Proposed models include those that incorporate post-decisional evidence collection (Pereira et al., 2020), models that consider the impact of response time (Kiani et al., 2014), and models that can be adapted to multi-alternative choice problems (Ratcliff and Starns, 2013).

The sequential sampling framework offers a means for capturing the temporal dynamics of decision-making processes. Besides, the temporal aspect of metacognitive computation has garnered significant attention in recent neuroscience studies (Cai et al., 2022; Shekhar and Rahnev 2018). Integrating these approaches holds the potential to enhance our understanding of the mechanisms governing confidence generation.

Bayesian confidence

Computational studies of metacognition have sparked a widespread debate concerning whether confidence is calculated through formal Bayesian inference (e.g., Adler and Ma, 2018; Aitchison et al., 2015; Khalvati et al., 2021; Li and Ma, 2020; Xue et al., 2023b). Central to Bayesian inference are two fundamental components, the prior probability and the likelihood function, which correspond to the observer's internal model of the world.

In a binary decision-making situation, the prior probability refers to a subjective belief regarding the occurrence of each stimulus scenario before observing an internal evidence value (e.g., stimuli A and B each having a 50% chance of occurring). Likelihood refers to the probability of observing a specific value of internal evidence under a given stimulus scenario. In the SDT example of **Fig. 1**, the height of each distribution represents the likelihood of internal evidence under its associated stimulus scenario (i.e., stimulus moving left or right). Once internal evidence is obtained, the Bayesian observer calculates the likelihood ratio, the relative heights of the two distributions at the location of the observed value. The posterior probability of a decision being correct is derived by combining the prior probability and the likelihood ratio, which is known as Bayesian confidence (e.g., Pouget et al., 2016).

Bayesian inference has often been discussed in relation to the concept of optimality (Ma, 2010; Rahnev and Denison, 2018). A computation is said to be optimal when it maximizes a particular performance measure. For instance, Bayesian decision-making based on the true knowledge of prior probabilities and likelihoods leads to maximized objective decision accuracy. Likewise, Bayesian confidence, when derived with true knowledge, is optimal in terms of the calibration of metacognitive monitoring since it reflects the true posterior probability of a decision being correct. The agent of performing Bayesian inference with true knowledge is commonly referred to as the Bayesian ideal observer (e.g., Green and Swets, 1966).

It is important to note, however, that Bayesian inference does not always equate to optimality (Ma, 2010). For example, by considering decision-makers with incorrect knowledge of prior probabilities and likelihoods, diverse suboptimal behaviors can be explained within the framework of Bayesian inference (Fleming and Daw, 2017; Hu et al., 2023; Khalvati et al., 2021; Ko and Lau, 2012).

Bayesian inference, in exchange for its flexible descriptive power, imposes challenges on observers by requiring inverse inference through the internal world model. Consequently, studies have explored whether human confidence rating is derived from computationally demanding Bayesian inference or simpler computations. Researchers typically manipulate distributions of stimulus intensity across trials and test whether participants exhibit Bayesian inference utilizing the distributional knowledge. The findings are somewhat mixed, but they suggest that human confidence rating does not fully conform to Bayesian inference (Adler and Ma, 2018; Aitchison et al., 2015; Li and Ma, 2020; Locke et al., 2022; Xue et al., 2023b). This means that even in controlled laboratory settings, where observers could ideally acquire complete knowledge of underlying distributions, human observers do not always adhere to Bayesian norms.

In the real world, obtaining accurate internal models for countless decision-making problems is exceedingly challenging. Consequently, greater emphasis is expected on heuristic computations, or Bayesian inference might be applied upon approximate knowledge of some fundamental structure of the world.

Non-Bayesian computations for confidence

A significant body of research has supported the notion that confidence is based on simpler, non-Bayesian computations (sometimes termed "heuristics"; Miyoshi et al., 2018; Navajas et al., 2017; Rahnev et al., 2011; Zawadzka et al., 2017; Zylberberg et al., 2012). A prominent example is a phenomenon known as positive evidence bias (Gao et al., 2023; Koizumi et al., 2015; Maniscalco et al., 2016; Samaha et al., 2017, 2019). This bias is typically observed in comparing two experimental conditions with different signal intensities but the same signal-to-noise ratio. Despite both conditions usually resulting in consistent decision accuracy, the condition of stronger stimulus intensity generally leads to higher subjective confidence. This phenomenon is often considered indicative of non-Bayesian confidence formation because Bayesian inference, predicated on true distributional knowledge of these conditions, produces a constant level of confidence in cases of matched decision accuracy.

The positive evidence bias, together with corroborative neurobiological findings (Cortese et al., 2016; Odegaard et al., 2018; Peters et al., 2017; Stolyarova et al., 2019), has prompted the notion that confidence primarily arises from the evidence favoring the chosen option, known as a response-congruent evidence (RCE) rule (Maniscalco et al., 2016; Miyoshi et al., 2018; Zylberberg et al., 2012). It may seem counterintuitive that our metacognitive system does not equally utilize all sources of information. However, Miyoshi and Lau (2020) have shown that the RCE rule leads to accurate metacognition when facing the typical evidence structure inherent in various decision-making scenarios (i.e., unequal variance between target and non-target evidence). This finding highlights a crucial avenue for metacognition research, emphasizing the importance of ecological rationality in information processing.

Interestingly, the positive evidence bias can also be explained through Bayesian inference if one assumes the fundamental evidence structure characterized by unequal variance (Miyoshi and Lau, 2020; Webb et al., 2021, 2023). In this case, Bayesian inference operates upon approximate knowledge of natural information structure rather than precise knowledge of specific experimental conditions. Thus, apparently non-Bayesian confidence computations may be viewed in Bayesian terms when considering adaptation to natural representational structures instead of artificial laboratory settings.

It is important to note that there could be various forms of non-Bayesian confidence computations other than the RCE rule (e.g., Boundy-Singer et al., 2023; Gao et al., 2023; Guggenmos, 2022; Navajas et al., 2017; Rahnev et al., 2011; Shekhar and Rahnev, 2022), and not all of them may find a connection to Bayesian confidence. Thus, further research is necessary to unravel the precise computations underlying confidence. Nonetheless, regardless of the approach one takes, acknowledging the rational basis of information processing holds great potential for advancing our understanding of metacognitive systems.

Neural network models of decision confidence

A variety of neural network models have been proposed to explain different aspects of decision confidence. Most notably, these include models that explain how decision confidence may be implemented in dynamic neural circuits (Maniscalco et al., 2021; Song et al., 2017), and models that utilize deep learning techniques to model decision confidence in the context of naturalistic stimuli such as images (Rafiei and Rahnev, 2022; Webb et al., 2021; Webb et al., 2023).

An early effort in this domain comes from the work of Cleeremans et al. (2007), who presented a two-part neural network model, in which a higher-order network learned to estimate the reliability of a first-order network's decisions. More recent work has focused on how decision confidence might be computed in dynamic neural circuits. Song et al. (2017) used reinforcement learning to train recurrent neural networks on a post-decision wagering task. They found that neurons in the network's recurrent hidden layer learned to perform a form of temporal evidence accumulation (see section **Sequential sampling framework**), thus implicitly represented decision confidence, similar to the observed behavior of individual neurons in decision-making brain regions (Kiani and Shadlen, 2009).

One limitation of these results is that they do not address cases in which confidence and objective decisions dissociate, such as the positive evidence bias (see section **Non-Bayesian computations for confidence**). To account for such dissociations, Maniscalco et al. (2021) presented a recurrent accumulator model consisting of two distinct types of accumulators. One type of accumulator represented the cumulative evidence in favor of a particular stimulus type, *independent* of other stimulus types, while another type of accumulator represented the *relative* evidence in favor of one stimulus type over others. This model was able to account for previously observed neural dissociations between decision-making and confidence, by modeling decisions using the relative accumulator units, and confidence using the independent accumulator units.

Other recent work has focused on the question of how decision confidence might be computed from naturalistic stimuli, such as real-world images. Webb et al. (2021, 2023) presented a deep neural network model trained both to classify realistic images, and to predict its own probability of being correct. Surprisingly, this model was found to capture dissociations between confidence and objective decisions, a result that depended on specific aspects of the stimuli on which the model was trained (e.g., variable image contrast). Rafiei and Rahnev (2022) introduced a model that combines a deep neural network architecture with accumulator units inspired by classic race models (Vickers, 1979; see section **Sequential sampling framework**). Importantly, the weights of the network were probabilistic, enabling the model to capture the process of noisy temporal evidence accumulation. This model illustrates how deep neural network modeling techniques can be fruitfully combined with the sequential sampling framework, allowing the model to both operate over real-world images and account for temporal features of decision-making such as response time.

Neural correlates of confidence

The theoretical frameworks we have explored provide valuable insights for neuroscience studies of metacognition. They stress the need to differentiate neural substrates related to confidence from those associated with objective decision-making. Additionally, studies can be designed aiming at neural correlates concerning the level of confidence or those associated with metacognitive sensitivity. By combining brain recording and computational modeling, we may achieve a deeper understanding of the neural foundations of metacognition.

Prefrontal and parietal regions

Neural processing for metacognition should involve the management across different processing modules to meet a variety of task demands. The prefrontal cortex (PFC) emerges as a prime candidate for this role, given its intricate connections with diverse areas responsible for perception, memory, emotion, and more (Barbas, 1995; Fuster, 2008; Goldman-Rakic, 1988; Petrides, 2000). Consistent with this idea, early studies have implicated frontal-parietal networks in playing significant roles in conscious introspection (Christoff and Gabrieli, 2000; Dehaene et al., 2003; Del Cul et al., 2009; Fletcher and Henson, 2001; Lau and Passingham, 2006; Lau and Rosenthal, 2011; Shimamura, 2000; Simons et al., 2010). These endeavors have prompted further investigations into frontoparietal regions as potential loci for metacognition.

Pioneering evidence has linked individual differences in metacognitive sensitivity to the anatomy of specific brain regions. A positive correlation has been identified between the gray matter volume in the anterior PFC and metacognitive sensitivity for perceptual discrimination (Fleming et al., 2010; McCurdy et al., 2013). Metacognitive sensitivity was also associated with the white matter microstructure in the genu of the corpus callosum, which connects to the anterior and orbital regions of the PFC (Fleming et al., 2010). Notably, these anatomical measures did not correlate with objective decision accuracy, highlighting the selective involvement of the PFC networks in metacognitive monitoring.

Functional brain recordings have further substantiated the pivotal role of the frontoparietal networks in metacognition (e.g., Cortese et al., 2016; Geurts et al., 2022; Hoven et al., 2022; Masset et al., 2020; Mazor et al., 2020; Morales et al., 2018; Pereira et al., 2020; Peters et al., 2017). For example, through behavior decoding with implanted electrocorticography electrodes, Peters et al. (2017) demonstrated spatiotemporally distinct neural representations for confidence and objective decisions. In a face/house discrimination task, the confidence level was successfully decoded around 450 ms following stimulus onset, contributed by multiple regions including the frontoparietal area. Conversely, the decoding of objective decision occurred approximately 250 ms after stimulus onset and predominantly relied on electrodes situated in the occipital region. Moreover, confidence manifested a stronger reliance on neural evidence for chosen than unchosen options, consistent with the RCE rule (see section **Non-Bayesian computations for confidence**).

Additional evidence comes from a study utilizing neurofeedback through functional magnetic resonance imaging (fMRI) (Cortese et al., 2016). In a random dot motion discrimination task, Cortese et al. constructed a confidence decoder based on multivoxel activation patterns in lateral prefrontal and inferior parietal areas. By manipulating activation in these regions via neurofeedback, they demonstrated bidirectional changes in confidence levels without affecting objective decision accuracy. This selective effect clearly demonstrates the causal involvement of frontoparietal areas in metacognitive monitoring.

The causal involvement of frontoparietal regions in metacognition is further supported by studies employing transcranial magnetic stimulation (TMS) (Cai et al., 2022; Luzio et al., 2022; Martin et al., 2023; Rahnev et al., 2016; Rounis et al., 2010; Ryals et al., 2016; Ruby et al., 2018; Shekhar and Rahnev, 2018; Ye et al., 2018; Xue et al., 2023a). An early study showed that offline TMS to the dorsolateral PFC decreased metacognitive sensitivity without compromising perceptual decision accuracy (Rounis et al., 2010; Ruby et al., 2018). Furthermore, Shekhar and Rahnev (2018) reported that online TMS to the dorsolateral PFC lowered the confidence level, while stimulation to the anterior PFC increased metacognitive sensitivity, suggesting distinct roles of PFC subregions. However, the effect of magnetic stimulation can vary based on stimulation protocols or task characteristics (Cai et al., 2022; Luzio et al., 2022), indicating the need for further investigations into the specific roles of PFC subregions.

Further evidence comes from studies employing pharmacological stimulation in animals. Lak et al. (2014) inactivated rats' orbitofrontal region by injecting GABA-A agonist muscimol during an olfactory discrimination task. The inactivation resulted in a disruption of metacognitive sensitivity while perceptual decision accuracy remained unaffected. Similarly, Miyamoto et al. (2017) injected muscimol into the posterior supra principal dimple and the anterior supplementary eye field of macaques' brain during a recognition memory task. The inactivation of these regions impaired metacognitive sensitivity without impacting objective memory performance, further substantiating the causal involvement of frontal networks in metacognitive monitoring.

Researchers have also investigated the involvement of frontoparietal sensorimotor networks in representing decision confidence (Gold and Shadlen, 2000; Kiani and Shadlen, 2009; Middlebrooks and Sommer, 2012). These studies commonly utilized tasks where monkeys indicated the direction of random dot motion by directing their gaze toward the corresponding direction. In the frontal eye field (FEF) and lateral intraparietal cortex (LIP), neural firing patterns were observed that presumably reflect the temporal accumulation of internal decision evidence (Gold and Shadlen, 2000; Kiani and Shadlen, 2009). Furthermore, the firing rate in the LIP and supplementary eye fields (SEF) correlated with trial-to-trial variations of confidence levels (Kiani and Shadlen, 2009; Middlebrooks and Sommer, 2012). These findings, combined with the theoretical framework of sequential sampling, have led to the view that objective decision and confidence may be represented in the sensorimotor circuits as a unified Bayesian probability (Kiani and Shadlen, 2009; Kiani et al., 2014; Meyniel et al., 2015).

One thing to keep in mind is that these findings on the sensorimotor networks could potentially be influenced by objective decision performance. Odegaard et al. (2018) conducted a subsequent study to address this possibility. They experimentally dissociated objective decision accuracy from confidence levels, and recorded neural activity in the superior colliculus, which is connected to the LIP and SEF. The firing rate in this region primarily reflected subjective confidence rather than integrative Bayesian probability, which aligns with non-Bayesian views of confidence formation.

Involvement of sensory regions

When examining the neural basis of confidence, the sensory cortex often comes up as a key area of interest. One noteworthy finding is that TMS applied to the occipital cortex increased confidence levels in a visual discrimination task (Rahnev et al., 2012). This phenomenon can be explained by considering that the distribution of internal sensory evidence, shown in **Fig. 1**, becomes noisier due to the magnetic stimulation. Namely, the increased variance of the distribution leads to greater probability density surpassing the criterion for high confidence judgment. A similar pattern has been observed when attention is diverted from visual stimuli using spatial cueing (Rahnev et al., 2011). In this case, inattention may result in noisier evidence distributions, again causing greater density to exceed the high confidence criterion. Therefore, it can be inferred that TMS and attentional cueing do not directly alter the representation of confidence per se, but rather interfere with upstream sensory processing, subsequently influencing the reported level of confidence.

To gain a clearer insight into the involvement of sensory regions, it may be helpful to consider the separation of sensory uncertainty and decision confidence (Bang and Fleming, 2018; Pouget et al., 2016). For instance, in random dot motion discrimination, the area MT encodes information about the estimated motion direction and its associated uncertainty (Maunsell and Van Essen, 1983). Consider an example where the motion direction is represented as 15 ± 10 degrees, with 0 degrees indicating vertical upward motion. Here, the observer may exhibit limited confidence when tasked with determining whether the motion is clockwise or counterclockwise relative to the vertical upward direction (0 degrees). Conversely, greater confidence would accompany the assertion that the perceived motion is counterclockwise concerning the rightward direction (90 degrees). Furthermore, when the observer reports the estimated direction as a continuous value rather than discriminating it against a specific direction, the sensory uncertainty of ± 10 degrees would assume the role of decision confidence itself. In this manner, sensory information in upstream brain regions may be reconstructed based on specific task demands, leading to confidence representation in higher brain regions.

The concept of sensory uncertainty accords with the theory of neural population coding (e.g., Abbott and Dayan, 1999; Ma et al., 2006; Paradiso, 1988; Pouget et al., 2002; Salinas and Abbott, 1994; Walker et al., 2020). In the primary visual cortex, for instance, distinct ensembles of neurons exhibit varying sensitivities to diverse stimulus orientations (e.g., Hubel and Wiesel, 1962; Ferster and Miller, 2000). According to the population coding theory, the combined activity of these neurons enables the representation of stimulus orientation and the sensory uncertainty associated with it. Recently, for example, Walker et al. (2020) demonstrated a successful application of a deep neural network to decode orientation information from the population activity in the primary visual cortex of macaques. This model accurately accounted for the monkeys' orientation classification behavior by considering both the decoded orientation information and its corresponding uncertainty. This finding fortifies the presence of sensory uncertainty representation embedded within the population activity of the sensory cortex.

Other contributing regions and factors

In addition to the insights above, numerous brain regions and cognitive factors have been identified as contributors to confidence formation. Regions maintaining close communication with the frontoparietal network have been implicated, such as the dorsal pulvinar (Komura et al., 2013) and the anterior cingulate cortex (Stolyarova et al., 2019). Moreover, investigations have underscored the involvement of the motor system (Faivre et al., 2020; Gajdos et al., 2019; Hobot et al., 2020; Pereira et al., 2020). For instance, Fleming et al. (2015) revealed that the TMS applied to the premotor cortex, for activating the response tied to the unchosen option, reduced confidence for the correct response. Furthermore, the influence of incentive motivation has been indicated, with the possible involvement of the ventromedial PFC (Hoven et al., 2022; Lebreton et al., 2018, 2019). The literature has also explored other pertinent factors, including arousal (Allen et al., 2016; Hauser et al., 2017), positive mood (Koellinger and Treffers, 2015), and the level of worry (Massoni, 2014).

These studies underscore that confidence is more than the estimated probability of a decision being correct. Rather, it manifests as a subjective experience shaped by a complex interplay of various factors. By delving into the mechanisms of its emergence, we may further enrich our understanding of the fundamental essence of confidence.

Metacognition across different cognitive domains

Our primary focus thus far has centered on metacognition for perceptual judgments. Nonetheless, research concerning metacognition has garnered substantial interest across a spectrum of cognitive domains.

Metamemory

The domain of memory research has a distinguished focus on metacognitive behavior. As mentioned earlier, studies in applied educational psychology have pursued functional roles of metacognition, such as optimizing learning time based on the self-assessment of item difficulty (e.g., Lovelace, 1984; Son and Metcalfe, 2000; Thiede and Dunlosky, 1999). A noteworthy concept in this context is prospective metacognition, which involves predicting future memory performance based on the present state of learning. The research methodologies and frameworks developed in these studies have extended beyond the scope of metamemory, yielding significant contributions to the broader landscape of metacognition studies (e.g., Nelson and Narens, 1990, 1994).

Another avenue of metamemory research has originated from cognitive psychology and neuroscience. Studies have illuminated two types of memory, explicit and implicit memory, characterized by the presence or absence of introspective memory awareness (e.g., Tulving, 1972, 1985; Squire and Zola-Morgan, 1988; Squire and Dede, 2015). Essential to explicit memory is the hippocampus and its adjacent regions, with damage to these areas impairs conscious memory introspection (Cohen and Squire, 1980; Scoville and Milner, 1957; Annese et al., 2014). Damage to the frontal cortex also disrupts explicit memory retrieval (e.g., Davidson et al., 2006; Wheeler et al., 1995), further emphasizing its significance in conscious introspection.

Implicit memory encompasses various phenomena such as perceptual priming (e.g., Tulving and Schacter, 1990) and perceptual learning (e.g., Gilbert and Li, 2012). Perceptual priming typically occurs when the same stimulus is repeatedly presented, leading to faster stimulus processing without conscious awareness of memory retrieval. Reduced activity in sensory areas is often associated with perceptual priming (Grill-Spector et al., 2006; Schacter et al., 2007), presumably reflecting heightened efficiency in neural computations (e.g., Desimone, 1996). Perceptual learning emerges through iterative stimulus exposures that progressively enhance stimulus detection or discrimination accuracy without conscious memory awareness. Perceptual learning is considered to involve plastic changes in sensory cortical circuits (e.g., Gilbert et al., 2009; Gilbert and Li, 2012).

The coexistence of conscious and unconscious forms of memory may offer distinct adaptive advantages. Implicit memory can guide efficient actions in familiar environments, while explicit memory enables us to mentally reconstruct events that are not currently present, empowering mental abilities such as reasoning and future planning (Atance and O'Neill, 2001; Schacter et al., 2017). Furthermore, our stored memories from the past are considered foundational for present experiences. Thus, theorists have proposed a perspective on how implicit and explicit memory would interplay in shaping the subjective quality of conscious experience (Lau et al., 2022).

Confidence rating has also been widely adopted in memory research. Particularly in recognition memory studies, scholars have utilized SDT and sequential sampling frameworks to explore the mechanisms of memory judgment and metamemory monitoring (e.g., Kellen and Klauer, 2015; Miyoshi et al., 2018; Osth et al., 2018; Ratcliff and Starns, 2013; Rotello et al., 2000; Wixted, 2007; Yonelinas, 1994). Additionally, researchers have investigated the relationship between confidence rating and the accuracy of memory judgment (e.g., Tekin et al., 2021), which carries significant implications in practical contexts such as eyewitness testimony (Seale-Carlisle et al., 2019; Wixted and Wells, 2017).

In the realm of neuroscience, neurons in the human posterior parietal cortex have been identified to express confidence in recognition memory judgment (Rutishauser et al., 2018). These neurons selectively increased their firing rate for either low or high confidence. Interestingly, the activity of these confidence neurons was independent of stimulus familiarity (i.e., whether the stimulus had been previously studied or not). Additionally, as mentioned earlier, pharmacological investigations on monkeys revealed the critical role of the parietal-frontal regions in accurate metamemory monitoring (Miyamoto et al., 2017).

Domain-generality of metacognitive monitoring

Further studies have delved into various domains, such as auditory perception (Ais et al., 2016; de Gardelle et al., 2016; Faivre et al., 2018), tactile perception (Arbuzova et al., 2022; Fairhurst et al., 2018; Faivre et al., 2018; Klever et al., 2023), time perception (Yallak and Balcı, 2021), and value-based decision-making (e.g., Brus et al., 2021; De Martino et al., 2013; Sepulveda et al., 2020). A key concept underlying these studies is the potential role of confidence as a common currency, facilitating comparability across different cognitive domains (de Gardelle and Mamassian, 2014; de Gardelle et al., 2016). By utilizing a modality-independent scale of confidence, one can prioritize more reliable information when making decisions involving multisensory inputs (de Gardelle et al., 2016; Klever et al., 2023). Moreover, subjective confidence associated with each modality plays a vital role during multisensory integration, affecting the relative weighting assigned to distinct modalities (Gao et al., 2023).

Another important question pertains to the potential shared mechanisms that underlie metacognition across diverse cognitive domains. Studies have explored whether individuals who exhibit good metacognitive accuracy in one domain also excel in other domains, thus addressing the domain-generality of metacognitive sensitivity (Ais et al., 2016; Faivre et al., 2018; Lee et al., 2018; Mazancieux et al., 2020). Faivre et al. (2018) reported that individuals' metacognitive sensitivity correlated among visual, auditory, and tactile tasks. Likewise, correlations in metacognitive sensitivity have been established between visual perception and memory domains (Lee et al., 2018; Mazancieux et al., 2020). These behavioral correlations suggest that confidence formation in different domains relies, to some degree, on shared underlying mechanisms.

The positive evidence bias (see section **Non-Bayesian computations for confidence**) has also been observed across a range of modalities, including visual perception, auditory perception, recognition memory, and value-based decision (Gao et al., 2023; Koizumi et al., 2015; Miyoshi et al., 2018; Odegaard et al., 2018; Sakamoto and Miyoshi, 2023; Samaha et al., 2017, 2019; Zawadzka et al., 2017). This suggests the operation of the RCE confidence rule across broad cognitive domains (Zylberberg et al., 2012; Maniscalco et al., 2016). As previously elaborated, the RCE computation affords efficient metacognition by aligning with the typical information structure prevalent in various decision-making scenarios (Miyoshi and Lau, 2020). Ecological rationality of this kind might serve as a foundation for the potential domain-generality of metacognition.

Research has also examined the neural bases of metacognition across different cognitive domains (e.g., Baird et al., 2013, 2015; Fleming et al., 2014; Heereman et al., 2015; Rouault et al., 2018b; Ye et al., 2018). Although the frontoparietal areas are generally considered central hubs for metacognitive monitoring, studies have implicated more nuanced involvements of their subregions. For instance, McCurdy et al. (2013) showed that metacognitive sensitivity for visual perception correlated with the gray matter volume in the anterior PFC, while that for recognition memory was associated with the volume in the precuneus. Additionally, in predicting confidence for visual perception and recognition memory, Morales et al. (2018) demonstrated domain-specific fMRI activities in the anterior PFC, while more broadly distributed domain-general activities were observed within the frontoparietal network. However, findings across diverse studies are not always consistent (see Rouault et al., 2018b), necessitating further research to achieve a more comprehensive understanding of this topic.

One consideration to note is that the generality of metacognitive sensitivity often does not hold when shifting between different task formats (e.g., Yes/No and forced-choice formats; see Lee et al., 2018; Miyoshi and Nishida, 2022). Furthermore, even within the realm of visual perception, distinct neural correlates of confidence have been observed for different task formats (Mazor et al., 2020). Therefore, it is crucial to ensure a consistent task format when attempting comparisons across different cognitive domains.

Metacognition beyond simple confidence rating

The study of metacognitive monitoring gives rise to a research field that encompasses various adjacent realms. This section will provide an overview of topics closely linked to the evaluation of confidence.

Error detection

One crucial component of metacognitive monitoring is the detection of erroneous decisions (Rabbitt, 1966; Rabbitt and Rodgers, 1977). This phenomenon is typically observed as "change of mind" behavior, where individuals swiftly amend their initial decision after an incorrect choice. Researchers have proposed that error detection involves continuous information gathering subsequent to the initial decision-making (Resulaj et al., 2009; van den Berg et al., 2016; Yeung and Summerfield, 2012). Within the framework of Bayesian decision-making, this behavior can be understood as a recalibration of confidence based on post-decisional information acquisition (Fleming and Daw, 2017). Specifically, in 2-choice tasks, the Bayesian observer can revise their initial decision once their confidence falls below 50%.

Studies utilizing electroencephalography have identified neural markers of error detection. These markers include the error-related negativity, a negative potential detected at frontocentral sites within 100 ms after an incorrect response. This is succeeded by the error positivity, a positive component localized at the parietal area (Debener et al., 2005; Falkenstein et al., 1991; Gehring et al., 1993). The error positivity has been observed only when individuals are aware of their errors, emphasizing its close association with explicit error recognition (Nieuwenhuis et al., 2001; Yeung and Summerfield, 2012). In contrast, the connection between error-related negativity and conscious error awareness remains somewhat inconclusive. While some studies suggest its role in influencing error awareness (Kirschner et al., 2021; Wessel et al., 2011), others solely associate it with latent post-error processes (Nieuwenhuis et al., 2001). A close connection is inferred between error detection mechanisms and confidence formation processes, and a seamless comprehension of these two holds the potential for further advancing the literature on metacognitive monitoring.

Visual awareness

Another commonly used metacognitive measure in visual tasks is visibility rating (e.g., Dehaene et al., 2006; Rausch and Zehetleitner, 2016). Unlike confidence rating, which evaluates response correctness under specific task demands, visibility rating solely concerns the subjective clarity of visual stimuli. Visibility and confidence closely align in detection tasks, reflecting similar facets of visual awareness. However, high visibility may not necessarily correspond to high confidence, particularly when discriminating subtle differences between highly visible stimuli. Consequently, studies have explored the possible interplay between visibility and confidence (Hellmann et al., 2023; King and Dehaene, 2014; Rausch et al., 2021).

The investigation of visual awareness holds particular significance in bridging the study of metacognition to consciousness research. This connection is exemplified by a phenomenon known as blindsight (Pöppel et al., 1973; Weiskrantz et al., 1974). Blindsight gained fame through the case of patient GY, who experienced damage to the left primary visual cortex, resulting in the loss of visual awareness for stimuli presented in the right visual field (e.g., Stoerig and Barth, 2001).

Remarkably, despite the lack of visual awareness, GY exhibited residual sensitivity in discrimination tasks within the affected visual field (de Gelder et al., 1999; Kentridge et al., 1997; Weiskrantz et al., 1995; Zeki and Ffytche, 1998). Blindsight serves as a vital example that distinguishes phenomenological awareness from task performance, offering a foothold for elucidating the computational and neural underpinnings of consciousness (Ko and Lau, 2012; Miyoshi and Lau, 2020; Persaud et al., 2011; Schmid et al., 2010; Yoshida and Isa, 2015). Nonetheless, diverse perspectives exist regarding the relevance of blindsight to consciousness research, and lively discussions are still ongoing (Lau, 2022; Michel, 2023; Michel and Lau, 2021; Peters, 2022; Phillips, 2021).

Perceptual reality monitoring

Another key concept linking metacognition and consciousness is perceptual reality monitoring. Research on consciousness involves the comparison of different theories regarding the prerequisites for consciousness to arise (e.g., Baars, 1993; Brown et al., 2019; Dehaene et al., 2017; Lamme, 2006). One crucial point of debate is whether self-monitoring is necessary for the occurrence of conscious perception (e.g., Lau and Rosenthal, 2011). A recent theory posits that automatic self-monitoring of sensory information veracity plays a pivotal role in giving rise to conscious awareness (Lau, 2019; Lau et al., 2022). It is proposed that higher-order brain regions, such as the prefrontal and parietal regions, engage in an automatic assessment of the veracity of representation patterns in sensory regions; conscious perception arises only when these patterns are deemed veridical signals resulting from external stimulation.

Recent studies have demonstrated that external visual input and internally generated mental imagery are associated with similar activation patterns in the sensory cortex (Dijkstra et al., 2019; Pearson, 2019; Shen et al., 2019; Xie et al., 2020). However, mental imagery is rarely confused with external input, except in cases of hallucinations or dreams (Hahamy et al., 2021; Koenig-Robert and Pearson, 2020). Generally, conscious visual awareness arises exclusively from external visual input, presumably due to the workings of perceptual reality monitoring mechanisms (Dijkstra et al., 2022). Nevertheless, it is possible to deliberately create situations that induce confusion between the two, enabling investigations into the mechanisms underlying perceptual reality monitoring and conscious awareness (Dijkstra et al., 2021; Dijkstra and Fleming, 2023; Perky, 1910).

Emerging studies have introduced potential computations that underlie perceptual reality monitoring, such as generative adversarial processing (Cushing et al., 2023; Gershman, 2019). Moreover, the frontoparietal regions have been implicated as a neural basis for perceptual reality monitoring, which overlaps with the known neural correlate of confidence (Dijkstra and Fleming, 2023; Lau et al., 2022). One question concerns the relationship between perceptual reality monitoring and the formation of decision confidence. Confidence may appear distinct from automatic perceptual reality monitoring due to its conscious nature, but its formation appears effortless and incidental to sensory processing. If reality monitoring relies on sensory strength thresholding (Dijkstra and Fleming, 2023), it may share mechanisms with confidence formation during visual detection. However, their relationship may be more complex in broader decision-making situations. Consequently, a unified understanding of metacognition across various processing hierarchies emerges as a crucial avenue for future research.

Conclusion

Beyond the realm of external reality lies another terrain of metacognition that extends into our inner world. Rooted in distinct information processing separate from objective decision-making, this internal sphere guides our adaptive behavior, nurtures communion with others, and serves as a gateway to a myriad of adjacent research fields. By achieving a deeper understanding of this introspective world, we can unveil the profound tapestry of human existence that eludes the confines of classical decision-making research.

Acknowledgment

The authors have no conflicts of interest relevant to the content of this article.

References

Abbott, L.F., Dayan, P., 1999. The effect of correlated variability on the accuracy of a population code. Neural Comput. 11, 91—101. https://doi.org/10.1162/089976699300016827.

Adler, W.T., Ma, W.J., 2018. Comparing Bayesian and non-Bayesian accounts of human confidence reports. PLoS Comput. Biol. 14, e1006572. https://doi.org/10.1371/journal.pcbi.1006572.

Aguilar-Lleyda, D., Lemarchand, M., de Gardelle, V., 2020. Confidence as a priority signal. Psychol. Sci. 31, 1084—1096. https://doi.org/10.1177/0956797620925039.

Ais, J., Zylberberg, A., Barttfeld, P., Sigman, M., 2016. Individual consistency in the accuracy and distribution of confidence judgments. Cognition 146, 377—386. https://doi.org/10.1016/j.cognition.2015.10.006.

Aitchison, L., Bang, D., Bahrami, B., Latham, P.E., 2015. Doubly Bayesian analysis of confidence in perceptual decision-making. PLoS Comput. Biol. 11, e1004519. https://doi.org/10.1371/journal.pcbi.1004519.

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

Annese, J., Schenker-Ahmed, N.M., Bartsch, H., Maechler, P., Sheh, C., Thomas, N., Kayano, J., Ghatan, A., Bresler, N., Frosch, M.P., Klaming, R., Corkin, S., 2014. Postmortem examination of patient H.M.'s brain based on histological sectioning and digital 3D reconstruction. Nat. Commun. 5, 3122. https://doi.org/10.1038/ncomms4122.

Arbuzova, P., Guo, S., Koß, C., Kurvits, L., Faivre, N., Kühn, A.A., Filevich, E., Ganos, C., 2022. No evidence of impaired visual and tactile metacognition in adults with tourette disorder. Parkinsonism Relat. Disord. 97, 29—33. https://doi.org/10.1016/j.parkreldis.2022.02.019.

Atance, C.M., O'Neill, D.K., 2001. Episodic future thinking. Trends Cogn. Sci. 5, 533—539. https://doi.org/10.1016/S1364-6613(00)01804-0.

Baars, B.J., 1993. A Cognitive Theory of Consciousness. Cambridge University Press, New York.

Bahrami, B., Olsen, K., Latham, P.E., Roepstorff, A., Rees, G., Frith, C.D., 2010. Optimally interacting minds. Science 329, 1081—1085. https://doi.org/10.1126/science.1185718.

Baird, B., Cieslak, M., Smallwood, J., Grafton, S.T., Schooler, J.W., 2015. Regional white matter variation associated with domain-specific metacognitive accuracy. J. Cogn. Neurosci. 27, 440—452. https://doi.org/10.1162/jocn_a_00741.

Baird, B., Smallwood, J., Gorgolewski, K.J., Margulies, D.S., 2013. Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. J. Neurosci. 33, 16657—16665. https://doi.org/10.1523/JNEUROSCI.0786-13.2013.

Bang, D., Fleming, S.M., 2018. Distinct encoding of decision confidence in human medial prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. 115, 6082—6087. https://doi.org/10.1073/pnas.1800795115.

Baranski, J.V., Petrusic, W.M., 1994. The calibration and resolution of confidence in perceptual judgments. Percept. Psychophys. 55, 412—428. https://doi.org/10.3758/BF03205299.

Barbas, H., 1995. Anatomic basis of cognitive-emotional interactions in the primate prefrontal cortex. Neurosci. Biobehav. Rev. 19, 499—510. https://doi.org/10.1016/0149-7634(94)00053-4.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., Cohen, J.D., 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. Psychol. Rev. 113, 700—765. https://doi.org/10.1037/0033-295X.113.4.700.

Boldt, A., Blundell, C., De Martino, B., 2019. Confidence modulates exploration and exploitation in value-based learning. Neurosci. Conscious. 2019. niz004. https://doi.org/10.1093/nc/niz004.

Boldt, A., Gilbert, S.J., 2019. Confidence guides spontaneous cognitive offloading. Cogn. Res. Princ. Implic. 4, 45. https://doi.org/10.1186/s41235-019-0195-y.

Boundy-Singer, Z.M., Ziemba, C.M., Goris, R.L.T., 2023. Confidence reflects a noisy decision reliability estimate. Nat. Hum. Behav. 7, 142—154. https://doi.org/10.1038/s41562-022-01464-x.

Brown, A.L., 1978. Knowing when, where and, how to remember: a problem of metacognition. In: Glaser, R. (Ed.), Advances in Instructional Psychology. Erlbaum Associates, Inc., Hillsdale.

Brown, R., Lau, H., LeDoux, J.E., 2019. Understanding the higher-order approach to consciousness. Trends Cogn. Sci. 23, 754—768. https://doi.org/10.1016/j.tics.2019.06.009.

Brown, S.D., Heathcote, A., 2008. The simplest complete model of choice response time: linear ballistic accumulation. Cogn. Psychol. 57, 153—178. https://doi.org/10.1016/j.cogpsych.2007.12.002.

Brus, J., Aebersold, H., Grueschow, M., Polania, R., 2021. Sources of confidence in value-based choice. Nat. Commun. 12, 7337. https://doi.org/10.1038/s41467-021-27618-5.

Cai, Y., Jin, Z., Zhai, C., Wang, H., Wang, J., Tang, Y., Kwok, S.C., 2022. Time-sensitive prefrontal involvement in associating confidence with task performance illustrates metacognitive introspection in monkeys. Commun. Biol. 5, 799. https://doi.org/10.1038/s42003-022-03762-6.

Chen, S., Rahnev, D., 2023. Confidence response times: challenging postdecisional models of confidence. J. Vis. 23, 1—10. https://doi.org/10.1167/jov.23.7.11.

Christoff, K., Gabrieli, J.D.E., 2000. The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. Psychobiology 28, 168—186. https://doi.org/10.3758/BF03331976.

Clarke, F.R., Birdsall, T.G., Tanner Jr., W.P., 1959. Two types of ROC curves and definitions of parameters. J. Acoust. Soc. Am. 31, 629—630. https://doi.org/10.1121/1.1907764.

Cleeremans, A., Timmermans, B., Pasquali, A., 2007. Consciousness and metarepresentation: a computational sketch. Neural Netw. 20, 1032—1039. https://doi.org/10.1016/j.neunet.2007.09.011.

Cohen, N.J., Squire, L.R., 1980. Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. Science 210, 207—210. https://doi.org/10.1126/science.7414331.

Colloff, M.F., Wilson, B.M., Seale-Carlisle, T.M., Wixted, J.T., 2021. Optimizing the selection of fillers in police lineups. Proc. Natl. Acad. Sci. U. S. A. 118, e2017292118. https://doi.org/10.1073/pnas.2017292118.

Cortese, A., 2022. Metacognitive resources for adaptive learning. Neurosci. Res. 178, 10—19. https://doi.org/10.1016/j.neures.2021.09.003.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., Lau, H., 2016. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nat. Commun. 7, 13669. https://doi.org/10.1038/ncomms13669.

Cortese, A., Lau, H., Kawato, M., 2020. Unconscious reinforcement learning of hidden brain states supported by confidence. Nat. Commun. 11, 4429. https://doi.org/10.1038/s41467-020-17828-8.

Cushing, C.A., Dawes, A.J., Hofmann, S.G., Lau, H., LeDoux, J.E., Taschereau-Dumouchel, V., 2023. A generative adversarial model of intrusive imagery in the human brain. PNAS Nexus 2, pgac265. https://doi.org/10.1093/pnasnexus/pgac265.

Daniel, R., Pollmann, S., 2012. Striatal activations signal prediction errors on confidence in the absence of external feedback. Neuroimage 59, 3457—3467. https://doi.org/10.1016/j.neuroimage.2011.11.058.

Davidson, P.S.R., Troyer, A.K., Moscovitch, M., 2006. Frontal lobe contributions to recognition and recall: linking basic research with clinical evaluation and remediation. J. Int. Neuropsychol. Soc. 12, 210—223. https://doi.org/10.1017/S1355617706060334.

Dayan, P., 2023. Metacognitive information theory. Open Mind 7, 392—411. https://doi.org/10.1162/opmi_a_00091.

de Gardelle, V., Le Corre, F., Mamassian, P., 2016. Confidence as a common currency between vision and audition. PLoS One 11, e0147901. https://doi.org/10.1371/journal.pone.0147901.

de Gardelle, V., Mamassian, P., 2014. Does confidence use a common currency across two visual tasks? Psychol. Sci. 25, 1286—1288. https://doi.org/10.1177/0956797614528956.

de Gelder, B., Vroomen, J., Pourtois, G., Weiskrantz, L., 1999. Non-conscious recognition of affect in the absence of striate cortex. Neuroreport 10, 3759—3763. https://doi.org/10.1097/00001756-199912160-00007.

De Martino, B., Fleming, S.M., Garrett, N., Dolan, R.J., 2013. Confidence in value-based choice. Nat. Neurosci. 16, 105—110. https://doi.org/10.1038/nn.3279.

Debener, S., Ullsperger, M., Siegel, M., Fiehler, K., Cramon, D.Y. von, Engel, A.K., 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. J. Neurosci. 25, 11730—11737. https://doi.org/10.1523/JNEUROSCI.3286-05.2005.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., Sergent, C., 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. Trends Cogn. Sci. 10, 204—211. https://doi.org/10.1016/j.tics.2006.03.007.

Dehaene, S., Lau, H., Kouider, S., 2017. What is consciousness, and could machines have it? Science 358, 486—492. https://doi.org/10.1126/science.aan8871.

Dehaene, S., Sergent, C., Changeux, J.-P., 2003. A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc. Natl. Acad. Sci. U. S. A. 100, 8520—8525. https://doi.org/10.1073/pnas.1332574100.

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., Slachevsky, A., 2009. Causal role of prefrontal cortex in the threshold for access to consciousness. Brain 132, 2531—2540. https://doi.org/10.1093/brain/awp111.

Desimone, R., 1996. Neural mechanisms for visual memory and their role in attention. Proc. Natl. Acad. Sci. U. S. A. 93, 13494—13499. https://doi.org/10.1073/pnas.93.24.13494.

Dijkstra, N., Bosch, S.E., Gerven, M.A.J. van, 2019. Shared neural mechanisms of visual perception and imagery. Trends Cogn. Sci. 23, 423—434. https://doi.org/10.1016/j.tics.2019.02.004.

Dijkstra, N., Fleming, S.M., 2023. Subjective signal strength distinguishes reality from imagination. Nat. Commun. 14, 1627. https://doi.org/10.1038/s41467-023-37322-1.

Dijkstra, N., Kok, P., Fleming, S.M., 2022. Perceptual reality monitoring: neural mechanisms dissociating imagination from reality. Neurosci. Biobehav. Rev. 135, 104557. https://doi.org/10.1016/j.neubiorev.2022.104557.

Dijkstra, N., Mazor, M., Kok, P., Fleming, S., 2021. Mistaking imagination for reality: congruent mental imagery leads to more liberal perceptual detection. Cognition 212, 104719. https://doi.org/10.1016/j.cognition.2021.104719.

Dotan, D., Meyniel, F., Dehaene, S., 2018. On-line confidence monitoring during decision making. Cognition 171, 112—121. https://doi.org/10.1016/j.cognition.2017.11.001.

Fairhurst, M.T., Travers, E., Hayward, V., Deroy, O., 2018. Confidence is higher in touch than in vision in cases of perceptual ambiguity. Sci. Rep. 8, 15604. https://doi.org/10.1038/s41598-018-34052-z.

Faivre, N., Filevich, E., Solovey, G., Kühn, S., Blanke, O., 2018. Behavioral, modeling, and electrophysiological evidence for supramodality in human metacognition. J. Neurosci. 38, 263—277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017.

Faivre, N., Vuillaume, L., Bernasconi, F., Salomon, R., Blanke, O., Cleeremans, A., 2020. Sensorimotor conflicts alter metacognitive and action monitoring. Cortex 124, 224—234. https://doi.org/10.1016/j.cortex.2019.12.001.

Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447—455. https://doi.org/10.1016/0013-4694(91)90062-9.

Ferster, D., Miller, K.D., 2000. Neural mechanisms of orientation selectivity in the visual cortex. Annu. Rev. Neurosci. 23, 441—471. https://doi.org/10.1146/annurev.neuro.23.1.441.

Flavell, J.H., 1979. Metacognition and cognitive monitoring: a new area of cognitive—developmental inquiry. Am. Psychol. 34, 906—911. https://doi.org/10.1037/0003-066X.34.10.906.

Flavell, J.H., 1971. First discussant's comments: what is memory development the development of? Hum. Dev. 14, 272—278. https://doi.org/10.1159/000271221.

Fleming, S.M., Daw, N.D., 2017. Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation. Psychol. Rev. 124, 91—114. https://doi.org/10.1037/rev0000045.

Fleming, S.M., Dolan, R.J., Frith, C.D., 2012. Metacognition: computation, biology and function. Philos. Trans. R. Soc. B Biol. Sci. 367, 1280—1286. https://doi.org/10.1098/rstb.2012.0021.

Fleming, S.M., Lau, H.C., 2014. How to measure metacognition. Front. Hum. Neurosci. 8, 443. https://doi.org/10.3389/fnhum.2014.00443.

Fleming, S.M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., Lau, H., 2015. Action-specific disruption of perceptual confidence. Psychol. Sci. 26, 89—98. https://doi.org/10.1177/0956797614557697.

Fleming, S.M., Ryu, J., Golfinos, J.G., Blackmon, K.E., 2014. Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. Brain 137, 2811—2822. https://doi.org/10.1093/brain/awu221.

Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R.J., Rees, G., 2010. Relating introspective accuracy to individual differences in brain structure. Science 329, 1541—1543. https://doi.org/10.1126/science.1191883.

Fletcher, P.C., Henson, R.N.A., 2001. Frontal lobes and human memory: insights from functional neuroimaging. Brain 124, 849—881. https://doi.org/10.1093/brain/124.5.849.

Fuster, J.M., 2008. The Prefrontal Cortex, fourth ed. Academic Press, Amsterdam.

Gajdos, T., Fleming, S.M., Saez Garcia, M., Weindel, G., Davranche, K., 2019. Revealing subthreshold motor contributions to perceptual confidence. Neurosci. Conscious. 2019, niz001. https://doi.org/10.1093/nc/niz001.

Galvin, S.J., Podd, J.V., Drga, V., Whitmore, J., 2003. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. Psychon. Bull. Rev. 10, 843—876. https://doi.org/10.3758/BF03196546.

Gao, Y., Xue, K., Odegaard, B., Rahnev, D., 2023. Common computations in automatic cue combination and metacognitive confidence reports. bioRxiv. https://doi.org/10.1101/2023.06.07.544029.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. Psychol. Sci. 4, 385—390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x.

Gershman, S.J., 2019. The generative adversarial brain. Front. Artif. Intell. 2, 18. https://doi.org/10.3389/frai.2019.00018.

Geurts, L.S., Cooke, J.R.H., Van Bergen, R.S., Jehee, J.F.M., 2022. Subjective confidence reflects representation of Bayesian probability in cortex. Nat. Hum. Behav. 6, 294—305. https://doi.org/10.1038/s41562-021-01247-w.

Gilbert, C.D., Li, W., 2012. Adult visual cortical plasticity. Neuron 75, 250—264. https://doi.org/10.1016/j.neuron.2012.06.030.

Gilbert, C.D., Li, W., Piech, V., 2009. Perceptual learning and adult cortical plasticity. J. Physiol. 587, 2743—2751. https://doi.org/10.1113/jphysiol.2009.171488.

Gold, J.I., Shadlen, M.N., 2001. Neural computations that underlie decisions about sensory stimuli. Trends Cogn. Sci. 5, 10—16. https://doi.org/10.1016/S1364-6613(00)01567-9.

Gold, J.I., Shadlen, M.N., 2000. Representation of a perceptual decision in developing oculomotor commands. Nature 404, 390—394. https://doi.org/10.1038/35006062.

Goldman-Rakic, P.S., 1988. Topography of cognition: parallel distributed networks in primate association cortex. Annu. Rev. Neurosci. 11, 137—156. https://doi.org/10.1146/annurev.ne.11.030188.001033.

Green, D.M., Swets, J.A., 1966. Signal Detection Theory and Psychophysics. John Wiley, Oxford.

Griffin, D., Tversky, A., 1992. The weighing of evidence and the determinants of confidence. Cogn. Psychol. 24, 411—435. https://doi.org/10.1016/0010-0285(92)90013-R.

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. 10, 14—23. https://doi.org/10.1016/j.tics.2005.11.006.

Guggenmos, M., 2022. Reverse engineering of metacognition. Elife 11, e75420. https://doi.org/10.7554/eLife.75420.

Guggenmos, M., Wilbertz, G., Hebart, M.N., Sterzer, P., 2016. Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. Elife 5, e13388. https://doi.org/10.7554/eLife.13388.

Hahamy, A., Wilf, M., Rosin, B., Behrmann, M., Malach, R., 2021. How do the blind "see"? The role of spontaneous brain activity in self-generated perception. Brain 144, 340—353. https://doi.org/10.1093/brain/awaa384.

Hauser, T.U., Allen, M., Purg, N., Moutoussis, M., Rees, G., Dolan, R.J., 2017. Noradrenaline blockade specifically enhances metacognitive performance. Elife 6, e24901. https://doi.org/10.7554/eLife.24901.

Heereman, J., Walter, H., Heekeren, H.R., 2015. A task-independent neural representation of subjective certainty in visual perception. Front. Hum. Neurosci. 9, 551. https://doi.org/10.3389/fnhum.2015.00551.

Hellmann, S., Zehetleitner, M., Rausch, M., 2023. Simultaneous modeling of choice, confidence, and response time in visual perception. Psychol. Rev. https://doi.org/10.1037/rev0000411.

Hobot, J., Koculak, M., Paulewicz, B., Sandberg, K., Wierzchoń, M., 2020. Transcranial magnetic stimulation-induced motor cortex activity influences visual awareness judgments. Front. Neurosci. 14, 580712. https://doi.org/10.3389/fnins.2020.580712.

Hoven, M., Brunner, G., De Boer, N.S., Goudriaan, A.E., Denys, D., Van Holst, R.J., Luigjes, J., Lebreton, M., 2022. Motivational signals disrupt metacognitive signals in the human ventromedial prefrontal cortex. Commun. Biol. 5, 244. https://doi.org/10.1038/s42003-022-03197-z.

Hoven, M., Lebreton, M., Engelmann, J.B., Denys, D., Luigjes, J., Van Holst, R.J., 2019. Abnormalities of confidence in psychiatry: an overview and future perspectives. Transl. Psychiatr. 9, 268. https://doi.org/10.1038/s41398-019-0602-7.

Hu, X., Yang, C., Luo, L., 2023. Decision criteria in signal detection model are not based on the objective likelihood ratio. J. Exp. Psychol. Gen. Advance online publication. https://doi.org/10.1037/xge0001438.

Hubel, D.H., Wiesel, T.N., 1962. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. Lond. 160, 106—154. https://doi.org/10.1113/jphysiol.1962.sp006837.

Kellen, D., Klauer, K.C., 2015. Signal detection and threshold modeling of confidence-rating ROCs: a critical test with minimal assumptions. Psychol. Rev. 122, 542—557. https://doi.org/10.1037/a0039251.

Kentridge, R.W., Heywood, C.A., Weiskrantz, L., 1997. Residual vision in multiple retinal locations within a scotoma: implications for blindsight. J. Cogn. Neurosci. 9, 191—202. https://doi.org/10.1162/jocn.1997.9.2.191.

Khalvati, K., Kiani, R., Rao, R.P.N., 2021. Bayesian inference with incomplete knowledge explains perceptual confidence and its deviations from accuracy. Nat. Commun. 12, 5704. https://doi.org/10.1038/s41467-021-25419-4.

Kiani, R., Corthell, L., Shadlen, M.N., 2014. Choice certainty is informed by both evidence and decision time. Neuron 84, 1329—1342. https://doi.org/10.1016/j.neuron.2014.12.015.

Kiani, R., Shadlen, M.N., 2009. Representation of confidence associated with a decision by neurons in the parietal cortex. Science 324, 759—764. https://doi.org/10.1126/science.1169405.

King, J.-R., Dehaene, S., 2014. A model of subjective report and objective discrimination as categorical decisions in a vast representational space. Philos. Trans. R. Soc. B Biol. Sci. 369, 20130204. https://doi.org/10.1098/rstb.2013.0204.

Klever, L., Beyvers, M.C., Fiehler, K., Mamassian, P., Billino, J., 2023. Cross-modal metacognition: visual and tactile confidence share a common scale. J. Vis. 23, 1—16. https://doi.org/10.1167/jov.23.5.3.

Ko, Y., Lau, H., 2012. A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. Philos. Trans. R. Soc. B Biol. Sci. 367, 1401—1411. https://doi.org/10.1098/rstb.2011.0380.

Koellinger, P., Treffers, T., 2015. Joy leads to overconfidence, and a simple countermeasure. PLoS One 10, e0143263. https://doi.org/10.1371/journal.pone.0143263.

Koenig-Robert, R., Pearson, J., 2020. Why do imagery and perception look and feel so different? Philos. Trans. R. Soc. B Biol. Sci. 376, 20190703. https://doi.org/10.1098/rstb.2019.0703.

Koizumi, A., Maniscalco, B., Lau, H., 2015. Does perceptual confidence facilitate cognitive control? Atten. Percept. Psychophys. 77, 1295—1306. https://doi.org/10.3758/s13414-015-0843-3.

Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., Miyamoto, A., 2013. Responses of pulvinar neurons reflect a subject's confidence in visual categorization. Nat. Neurosci. 16, 749—755. https://doi.org/10.1038/nn.3393.

Koriat, A., Lichtenstein, S., Fischhoff, B., 1980. Reasons for confidence. J. Exp. Psychol. 6, 107—118. https://doi.org/10.1037/0278-7393.6.2.107.

Kirschner, H., Humann, J., Derrfuss, J., Danielmeier, C., Ullsperger, M., 2021. Neural and behavioral traces of error awareness. Cogn. Affect. Behav. Neurosci. 21, 573—591. https://doi.org/10.3758/s13415-020-00838-w.

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

Lamme, V.A.F., 2006. Towards a true neural stance on consciousness. Trends Cogn. Sci. 10, 494—501. https://doi.org/10.1016/j.tics.2006.09.001.

Lau, H., 2022. In: Consciousness We Trust: The Cognitive Neuroscience of Subjective Experience. Oxford University Press, Oxford.

Lau, H., 2019. Consciousness, metacognition, & perceptual reality monitoring. PsyArXiv. https://doi.org/10.31234/osf.io/ckbyf.

Lau, H., Michel, M., LeDoux, J.E., Fleming, S.M., 2022. The mnemonic basis of subjective experience. Nat. Rev. Psychol. 1, 479—488. https://doi.org/10.1038/s44159-022-00068-6.

Lau, H., Rosenthal, D., 2011. Empirical support for higher-order theories of conscious awareness. Trends Cogn. Sci. 15, 365—373. https://doi.org/10.1016/j.tics.2011.05.009.

Lau, H.C., Passingham, R.E., 2006. Relative blindsight in normal observers and the neural correlate of visual consciousness. Proc. Natl. Acad. Sci. U. S. A. 103, 18763—18768. https://doi.org/10.1073/pnas.0607716103.

Lebreton, M., Bacily, K., Palminteri, S., Engelmann, J.B., 2019. Contextual influence on confidence judgments in human reinforcement learning. PLoS Comput. Biol. 15, e1006973. https://doi.org/10.1371/journal.pcbi.1006973.

Lebreton, M., Langdon, S., Slieker, M.J., Nooitgedacht, J.S., Goudriaan, A.E., Denys, D., van Holst, R.J., Luigjes, J., 2018. Two sides of the same coin: monetary incentives concurrently improve and bias confidence judgments. Sci. Adv. 4, eaaq0668. https://doi.org/10.1126/sciadv.aaq0668.

Lee, A.L.F., Ruby, E., Giles, N., Lau, H., 2018. Cross-domain association in metacognitive efficiency depends on first-order task types. Front. Psychol. 9, 2464. https://doi.org/10.3389/fpsyg.2018.02464.

Li, H.-H., Ma, W.J., 2020. Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. Nat. Commun. 11, 2004. https://doi.org/10.1038/s41467-020-15581-6.

Locke, S.M., Landy, M.S., Mamassian, P., 2022. Suprathreshold perceptual decisions constrain models of confidence. PLoS Comput. Biol. 18, e1010318. https://doi.org/10.1371/journal.pcbi.1010318.

Lovelace, E.A., 1984. Metamemory: monitoring future recallability during study. J. Exp. Psychol. Learn. Mem. Cogn. 10, 756—766. https://doi.org/10.1037/0278-7393.10.4.756.

Luzio, P.D., Tarasi, L., Silvanto, J., Avenanti, A., Romei, V., 2022. Human perceptual and metacognitive decision-making rely on distinct brain networks. PLoS Biol. 20, e3001750. https://doi.org/10.1371/journal.pbio.3001750.

Ma, W.J., 2010. Signal detection theory, uncertainty, and Poisson-like population codes. Vision Res. 50, 2308—2319. https://doi.org/10.1016/j.visres.2010.08.035.

Ma, W.J., Beck, J.M., Latham, P.E., Pouget, A., 2006. Bayesian inference with probabilistic population codes. Nat. Neurosci. 9, 1432—1438. https://doi.org/10.1038/nn1790.

Macmillan, N.A., Creelman, C.D., 2005. Detection Theory: A User's Guide, second ed. Lawrence Erlbaum Associates Publishers, Mahwah.

Maniscalco, B., Lau, H., 2015. Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task. Neurosci. Conscious 2015, niv002. https://doi.org/10.1093/nc/niv002.

Maniscalco, B., Lau, H., 2012. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious. Cogn. 21, 422—430. https://doi.org/10.1016/j.concog.2011.09.021.

Maniscalco, B., Odegaard, B., Grimaldi, P., Cho, S.H., Basso, M.A., Lau, H., Peters, M.A.K., 2021. Tuned inhibition in perceptual decision-making circuits can explain seemingly suboptimal confidence behavior. PLoS Comput. Biol. 17, e1008779. https://doi.org/10.1371/journal.pcbi.1008779.

Maniscalco, B., Peters, M.A.K., Lau, H., 2016. Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. Atten. Percept. Psychophys. 78, 923—937. https://doi.org/10.3758/s13414-016-1059-x.

Martin, A., Lane, T.J., Hsu, T.-Y., 2023. DLPFC-PPC-cTBS effects on metacognitive awareness. Cortex 167, 41—50. https://doi.org/10.1016/j.cortex.2023.05.022.

Masset, P., Ott, T., Lak, A., Hirokawa, J., Kepecs, A., 2020. Behavior- and modality-general representation of confidence in orbitofrontal cortex. Cell 182, 112—126. https://doi.org/10.1016/j.cell.2020.05.022.

Massoni, S., 2014. Emotion as a boost to metacognition: how worry enhances the quality of confidence. Conscious. Cogn. 29, 189—198. https://doi.org/10.1016/j.concog.2014.08.006.

Maunsell, J.H., Van Essen, D.C., 1983. Functional properties of neurons in middle temporal visual area of the macaque monkey: I. Selectivity for stimulus direction, speed, and orientation. J. Neurophysiol. 49, 1127—1147. https://doi.org/10.1152/jn.1983.49.5.1127.

Mazancieux, A., Dinze, C., Souchay, C., Moulin, C.J.A., 2020. Metacognitive domain specificity in feeling-of-knowing but not retrospective confidence. Neurosci. Conscious 2020. https://doi.org/10.1093/nc/niaa001.

Mazor, M., Friston, K.J., Fleming, S.M., 2020. Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. Elife 9, e53900. https://doi.org/10.7554/eLife.53900.

McCurdy, L.Y., Maniscalco, B., Metcalfe, J., Liu, K.Y., De Lange, F.P., Lau, H., 2013. Anatomical coupling between distinct metacognitive systems for memory and visual perception. J. Neurosci. 33, 1897—1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013.

Meyniel, F., Sigman, M., Mainen, Z.F., 2015. Confidence as Bayesian probability: from neural origins to behavior. Neuron 88, 78—92. https://doi.org/10.1016/j.neuron.2015.09.039.

Michel, M., 2023. Confidence in consciousness research. WIREs Cogn. Sci. 14, e1628. https://doi.org/10.1002/wcs.1628.

Michel, M., Lau, H., 2021. Is blindsight possible under signal detection theory? Comment on Phillips (2021). Psychol. Rev. 128, 585—591. https://doi.org/10.1037/rev0000266.

Middlebrooks, P.G., Sommer, M.A., 2012. Neuronal correlates of metacognition in primate frontal cortex. Neuron 75, 517—530. https://doi.org/10.1016/j.neuron.2012.05.028.

Miyamoto, K., Osada, T., Setsuie, R., Takeda, M., Tamura, K., Adachi, Y., Miyashita, Y., 2017. Causal neural network of metamemory for retrospection in primates. Science 355, 188—193. https://doi.org/10.1126/science.aal0162.

Miyoshi, K., Kuwahara, A., Kawaguchi, J., 2018. Comparing the confidence calculation rules for forced-choice recognition memory: a winner-takes-all rule wins. J. Mem. Lang. 102, 142—154. https://doi.org/10.1016/j.jml.2018.06.001.

Miyoshi, K., Lau, H., 2020. A decision-congruent heuristic gives superior metacognitive sensitivity under realistic variance assumptions. Psychol. Rev. 127, 655—671. https://doi.org/10.1037/rev0000184.

Miyoshi, K., Nishida, S., 2022. GGSDT: a unified signal detection framework for confidence data analysis. bioRxiv. https://doi.org/10.1101/2022.10.28.514329.

Miyoshi, K., Sakamoto, Y., Nishida, S., 2022. On the assumptions behind metacognitive measurements: implications for theory and practice. J. Vis. 22, 1—15. https://doi.org/10.1167/jov.22.10.18.

Morales, J., Lau, H., Fleming, S.M., 2018. Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. J. Neurosci. Off. J. Soc. Neurosci. 38, 3534—3546. https://doi.org/10.1523/JNEUROSCI.2360-17.2018.

Morales, J., Odegaard, B., Maniscalco, B., 2022. The neural substrates of conscious perception without performance confounds. In: De Brigard, F., Sinnott-Armstrong, W. (Eds.), Neuroscience and Philosophy. MIT Press, Cambridge.

Moreira, C.M., Rollwage, M., Kaduk, K., Wilke, M., Kagan, I., 2018. Post-decision wagering after perceptual judgments reveals bi-directional certainty readouts. Cognition 176, 40—52. https://doi.org/10.1016/j.cognition.2018.02.026.

Navajas, J., Bahrami, B., Latham, P.E., 2016. Post-decisional accounts of biases in confidence. Curr. Opin. Behav. Sci. 11, 55—60. https://doi.org/10.1016/j.cobeha.2016.05.005.

Navajas, J., Hindocha, C., Foda, H., Keramati, M., Latham, P.E., Bahrami, B., 2017. The idiosyncratic nature of confidence. Nat. Hum. Behav. 1, 810—818. https://doi.org/10.1038/s41562-017-0215-1.

Nelson, T.O., Narens, L., 1994. Why investigate metacognition?. In: Metacognition: Knowing about Knowing. MIT Press, Cambridge. https://doi.org/10.7551/mitpress/4561.001.0001.

Nelson, T.O., Narens, L., 1990. Metamemory: a theoretical framework and new findings. In: Psychology of Learning and Motivation. Elsevier. https://doi.org/10.1016/S0079-7421(08)60053-5.

Nieuwenhuis, S., Ridderinkhof, K.R., Blom, J., Band, G.P.H., Kok, A., 2001. Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. Psychophysiology 38, 752—760. https://doi.org/10.1111/1469-8986.3850752.

Odegaard, B., Grimaldi, P., Cho, S.H., Peters, M.A.K., Lau, H., Basso, M.A., 2018. Superior colliculus neuronal ensemble activity signals optimal rather than subjective confidence. Proc. Natl. Acad. Sci. U. S. A. 115, E1588—E1597. https://doi.org/10.1073/pnas.1711628115.

Osth, A.F., Jansson, A., Dennis, S., Heathcote, A., 2018. Modeling the dynamics of recognition memory testing with an integrated model of retrieval and decision making. Cogn. Psychol. 104, 106—142. https://doi.org/10.1016/j.cogpsych.2018.04.002.

Paradiso, M.A., 1988. A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. Biol. Cybern. 58, 35—49. https://doi.org/10.1007/BF00363954.

Pearson, J., 2019. The human imagination: the cognitive neuroscience of visual mental imagery. Nat. Rev. Neurosci. 20, 624—634. https://doi.org/10.1038/s41583-019-0202-9.

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., Desvachez, A., Blanke, O., Van De Ville, D., Millán, J. del R., 2020. Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. Proc. Natl. Acad. Sci. U. S. A. 117, 8382—8390. https://doi.org/10.1073/pnas.1918335117.

Perky, C.W., 1910. An Experimental study of imagination. Am. J. Psychol. 21, 422—452. https://doi.org/10.2307/1413350.

Persaud, N., Davidson, M., Maniscalco, B., Mobbs, D., Passingham, R.E., Cowey, A., Lau, H., 2011. Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. Neuroimage 58, 605—611. https://doi.org/10.1016/j.neuroimage.2011.06.081.

Peters, M.A.K., 2022. Towards characterizing the canonical computations generating phenomenal experience. Neurosci. Biobehav. Rev. 142, 104903. https://doi.org/10.1016/j.neubiorev.2022.104903.

Peters, M.A.K., Thesen, T., Ko, Y.D., Maniscalco, B., Carlson, C., Davidson, M., Doyle, W., Kuzniecky, R., Devinsky, O., Halgren, E., Lau, H., 2017. Perceptual confidence neglects decision-incongruent evidence in the brain. Nat. Hum. Behav. 1, 0139. https://doi.org/10.1038/s41562-017-0139.

Peterson, W., Birdsall, T., Fox, W., 1954. The theory of signal detectability. Trans. IRE Prof. Group Inf. Theory 4, 171—212. https://doi.org/10.1109/TIT.1954.1057460.

Petrides, M., 2000. The role of the mid-dorsolateral prefrontal cortex in working memory. Exp. Brain Res. 133, 44—54. https://doi.org/10.1007/s002210000399.

Phillips, I., 2021. Blindsight is qualitatively degraded conscious vision. Psychol. Rev. 128, 558—584. https://doi.org/10.1037/rev0000254.

Pleskac, T.J., Busemeyer, J.R., 2010. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychol. Rev. 117, 864—901. https://doi.org/10.1037/a0019737.

Pöppel, E., Held, R., Frost, D., 1973. Residual visual function after brain wounds involving the central visual pathways in man. Nature 243, 295—296. https://doi.org/10.1038/243295a0.

Pouget, A., Deneve, S., Duhamel, J.-R., 2002. A computational perspective on the neural basis of multisensory spatial representations. Nat. Rev. Neurosci. 3, 741—747. https://doi.org/10.1038/nrn914.

Pouget, A., Drugowitsch, J., Kepecs, A., 2016. Confidence and certainty: distinct probabilistic quantities for different goals. Nat. Neurosci. 19, 366—374. https://doi.org/10.1038/nn.4240.

Rabbitt, P., Rodgers, B., 1977. What does a man do after he makes an error? An analysis of response programming. Q. J. Exp. Psychol. 29, 727—743. https://doi.org/10.1080/14640747708400645.

Rabbitt, P.M.A., 1966. Error correction time without external error signals. Nature 212, 438. https://doi.org/10.1038/212438a0.

Rafiei, F., Rahnev, D., 2022. RTNet: a neural network that exhibits the signatures of human perceptual decision making. bioRxiv. https://doi.org/10.1101/2022.08.23.505015.

Rahnev, D., 2023. Measuring metacognition: a comprehensive assessment of current methods. PsyArXiv. https://doi.org/10.31234/osf.io/waz9h.

Rahnev, D., Balsdon, T., Charles, L., de Gardelle, V., Denison, R., Desender, K., Faivre, N., Filevich, E., Fleming, S.M., Jehee, J., Lau, H., Lee, A.L.F., Locke, S.M., Mamassian, P., Odegaard, B., Peters, M., Reyes, G., Rouault, M., Sackur, J., Samaha, J., Sergent, C., Sherman, M.T., Siedlecka, M., Soto, D., Vlassova, A., Zylberberg, A., 2022. Consensus goals in the field of visual metacognition. Perspect. Psychol. Sci. 17, 1746—1765. https://doi.org/10.1177/17456916221075615.

Rahnev, D., Denison, R.N., 2018. Suboptimality in perceptual decision making. Behav. Brain Sci. 41, e223. https://doi.org/10.1017/S0140525X18000936.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., de Lange, F.P., Lau, H., 2011. Attention induces conservative subjective biases in visual perception. Nat. Neurosci. 14, 1513—1515. https://doi.org/10.1038/nn.2948.

Rahnev, D., Nee, D.E., Riddle, J., Larson, A.S., D'Esposito, M., 2016. Causal evidence for frontal cortex organization for perceptual decision making. Proc. Natl. Acad. Sci. U. S. A. 113, 6059—6064. https://doi.org/10.1073/pnas.1522551113.

Rahnev, D.A., Maniscalco, B., Luber, B., Lau, H., Lisanby, S.H., 2012. Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. J. Neurophysiol. 107, 1556—1563. https://doi.org/10.1152/jn.00985.2011.

Ratcliff, R., 1978. A theory of memory retrieval. Psychol. Rev. 85, 59—108. https://doi.org/10.1037/0033-295X.85.2.59.

Ratcliff, R., Smith, P.L., Brown, S.D., McKoon, G., 2016. Diffusion decision model: current issues and history. Trends Cogn. Sci. 20, 260—281. https://doi.org/10.1016/j.tics.2016.01.007.

Ratcliff, R., Starns, J.J., 2013. Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. Psychol. Rev. 120, 697—719. https://doi.org/10.1037/a0033152.

Rausch, M., Hellmann, S., Zehetleitner, M., 2021. Modelling visibility judgments using models of decision confidence. Atten. Percept. Psychophys. 83, 3311—3336. https://doi.org/10.3758/s13414-021-02284-3.

Rausch, M., Zehetleitner, M., 2016. Visibility is not equivalent to confidence in a low contrast orientation discrimination task. Front. Psychol. 7, 591. https://doi.org/10.3389/fpsyg.2016.00591.

Resulaj, A., Kiani, R., Wolpert, D.M., Shadlen, M.N., 2009. Changes of mind in decision-making. Nature 461, 263—266. https://doi.org/10.1038/nature08275.

Rollwage, M., Fleming, S.M., 2021. Confirmation bias is adaptive when coupled with efficient metacognition. Philos. Trans. R. Soc. Lond. B Biol. Sci. 376, 20200131. https://doi.org/10.1098/rstb.2020.0131.

Rollwage, M., Loosen, A., Hauser, T.U., Moran, R., Dolan, R.J., Fleming, S.M., 2020. Confidence drives a neural confirmation bias. Nat. Commun. 11, 2634. https://doi.org/10.1038/s41467-020-16278-6.

Rotello, C.M., Macmillan, N.A., Van Tassel, G., 2000. Recall-to-reject in recognition: evidence from ROC curves. J. Mem. Lang. 43, 67—88. https://doi.org/10.1006/jmla.1999.2701.

Rouault, M., Seow, T., Gillan, C.M., Fleming, S.M., 2018a. Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biol. Psychiatr. 84, 443—451. https://doi.org/10.1016/j.biopsych.2017.12.017.

Rouault, M., McWilliams, A., Allen, M.G., Fleming, S.M., 2018b. Human metacognition across domains: insights from individual differences and neuroimaging. Personal. Neurosci. 1, e17. https://doi.org/10.1017/pen.2018.16.

Rounis, E., Maniscalco, B., Rothwell, J.C., Passingham, R.E., Lau, H., 2010. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cogn. Neurosci. 1, 165—175. https://doi.org/10.1080/17588921003632529.

Rouy, M., Saliou, P., Nalborczyk, L., Pereira, M., Roux, P., Faivre, N., 2021. Systematic review and meta-analysis of metacognitive abilities in individuals with schizophrenia spectrum disorders. Neurosci. Biobehav. Rev. 126, 329—337. https://doi.org/10.1016/j.neubiorev.2021.03.017.

Ruby, E., Maniscalco, B., Peters, M.A.K., 2018. On a "failed" attempt to manipulate visual metacognition with transcranial magnetic stimulation to prefrontal cortex. Conscious. Cogn. 62, 34—41. https://doi.org/10.1016/j.concog.2018.04.009.

Rutishauser, U., Aflalo, T., Rosario, E.R., Pouratian, N., Andersen, R.A., 2018. Single-neuron representation of memory strength and recognition confidence in left human posterior parietal cortex. Neuron 97, 209—220. https://doi.org/10.1016/j.neuron.2017.11.029.

Ryals, A.J., Rogers, L.M., Gross, E.Z., Polnaszek, K.L., Voss, J.L., 2016. Associative recognition memory awareness improved by theta-burst stimulation of frontopolar cortex. Cereb. Cortex 26, 1200—1210. https://doi.org/10.1093/cercor/bhu311.

Sakamoto, Y., Miyoshi, K., 2023. A confidence framing effect: flexible use of evidence in metacognitive monitoring. PsyArXiv. https://doi.org/10.31234/osf.io/mpt3e.

Salinas, E., Abbott, L.F., 1994. Vector reconstruction from firing rates. J. Comput. Neurosci. 1, 89—107. https://doi.org/10.1007/BF00962720.

Samaha, J., Iemi, L., Postle, B.R., 2017. Prestimulus alpha-band power biases visual discrimination confidence, but not accuracy. Conscious. Cogn. 54, 47—55. https://doi.org/10.1016/j.concog.2017.02.005.

Samaha, J., Switzky, M., Postle, B.R., 2019. Confidence boosts serial dependence in orientation estimation. J. Vis. 19, 1—13. https://doi.org/10.1167/19.4.25.

Schacter, D.L., Benoit, R.G., Szpunar, K.K., 2017. Episodic future thinking: mechanisms and functions. Curr. Opin. Behav. Sci. 17, 41—50. https://doi.org/10.1016/j.cobeha.2017.06.002.

Schacter, D.L., Wig, G.S., Stevens, W.D., 2007. Reductions in cortical activity during priming. Curr. Opin. Neurobiol. 17, 171—176. https://doi.org/10.1016/j.conb.2007.02.001.

Schmid, M.C., Mrowka, S.W., Turchi, J., Saunders, R.C., Wilke, M., Peters, A.J., Ye, F.Q., Leopold, D.A., 2010. Blindsight depends on the lateral geniculate nucleus. Nature 466, 373—377. https://doi.org/10.1038/nature09179.

Scoville, W.B., Milner, B., 1957. Loss of recent memory after bilateral hippocampal lesions. J. Neurol. Neurosurg. Psychiatr. 20, 11—21. https://doi.org/10.1136/jnnp.20.1.11.

Seale-Carlisle, T.M., Colloff, M.F., Flowe, H.D., Wells, W., Wixted, J.T., Mickes, L., 2019. Confidence and response time as indicators of eyewitness identification accuracy in the lab and in the real world. J. Appl. Res. Mem. Cogn. 8, 420—428. https://doi.org/10.1016/j.jarmac.2019.09.003.

Sepulveda, P., Usher, M., Davies, N., Benson, A.A., Ortoleva, P., De Martino, B., 2020. Visual attention modulates the integration of goal-relevant evidence and not value. Elife 9, e60705. https://doi.org/10.7554/eLife.60705.

Shekhar, M., Rahnev, D., 2018. Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. J. Neurosci. 38, 5078—5087. https://doi.org/10.1523/JNEUROSCI.3484-17.2018.

Shekhar, M., Rahnev, D., 2021a. Sources of metacognitive inefficiency. Trends Cogn. Sci. 25, 12—23. https://doi.org/10.1016/j.tics.2020.10.007.

Shekhar, M., Rahnev, D., 2021b. The nature of metacognitive inefficiency in perceptual decision making. Psychol. Rev. 128, 45—70. https://doi.org/10.1037/rev0000249.

Shekhar, M., Rahnev, D., 2022. How do humans give confidence? A comprehensive comparison of process models of metacognition. PsyArXiv. https://doi.org/10.31234/osf.io/cwrnt.

Shen, G., Horikawa, T., Majima, K., Kamitani, Y., 2019. Deep image reconstruction from human brain activity. PLoS Comput. Biol. 15, e1006633. https://doi.org/10.1371/journal.pcbi.1006633.

Shimamura, A.P., 2000. The role of the prefrontal cortex in dynamic filtering. Psychobiology 28, 207—218. https://doi.org/10.3758/BF03331979.

Simons, J.S., Peers, P.V., Mazuz, Y.S., Berryhill, M.E., Olson, I.R., 2010. Dissociation between memory accuracy and memory confidence following bilateral parietal lesions. Cereb. Cortex 20, 479—485. https://doi.org/10.1093/cercor/bhp116.

Son, L.K., Metcalfe, J., 2000. Metacognitive and control strategies in study-time allocation. J. Exp. Psychol. Learn. Mem. Cogn. 26, 204—221. https://doi.org/10.1037/0278-7393.26.1.204.

Song, H.F., Yang, G.R., Wang, X.-J., 2017. Reward-based training of recurrent neural networks for cognitive and value-based tasks. Elife 6, e21492. https://doi.org/10.7554/eLife.21492.

Squire, L.R., Dede, A.J.O., 2015. Conscious and unconscious memory systems. Cold Spring Harb. Perspect. Biol. 7, a021667. https://doi.org/10.1101/cshperspect.a021667.

Squire, L.R., Zola-Morgan, S., 1988. Memory: brain systems and behavior. Trends Neurosci. 11, 170—175. https://doi.org/10.1016/0166-2236(88)90144-0.

Stoerig, P., Barth, E., 2001. Low-level phenomenal vision despite unilateral destruction of primary visual cortex. Conscious. Cogn. 10, 574—587. https://doi.org/10.1006/ccog.2001.0526.

Stolyarova, A., Rakhshan, M., Hart, E.E., O'Dell, T.J., Peters, M.A.K., Lau, H., Soltani, A., Izquierdo, A., 2019. Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. Nat. Commun. 10, 4704. https://doi.org/10.1038/s41467-019-12725-1.

Tajima, S., Drugowitsch, J., Patel, N., Pouget, A., 2019. Optimal policy for multi-alternative decisions. Nat. Neurosci. 22, 1503—1511. https://doi.org/10.1038/s41593-019-0453-9.

Tanner Jr., W.P., Swets, J.A., 1954. A decision-making theory of visual detection. Psychol. Rev. 61, 401—409. https://doi.org/10.1037/h0058700.

Tekin, E., DeSoto, K.A., Wixted, J.H., Roediger Iii, H.L., 2021. Applying confidence accuracy characteristic plots to old/new recognition memory experiments. Memory 29, 427—443. https://doi.org/10.1080/09658211.2021.1901937.

Thiede, K.W., Dunlosky, J., 1999. Toward a general model of self-regulated study: an analysis of selection of items for study and self-paced study time. J. Exp. Psychol. Learn. Mem. Cogn. 25, 1024—1037. https://doi.org/10.1037/0278-7393.25.4.1024.

Tulving, E., 1985. Memory and consciousness. Can. Psychol. Psychol. Can. 26, 1—12. https://doi.org/10.1037/h0080017.

Tulving, E., 1972. Episodic and semantic memory. In: Organization of Memory. Academic Press, Oxford.

Tulving, E., Schacter, D.L., 1990. Priming and human memory systems. Science 247, 301—306. https://doi.org/10.1126/science.2296719.

van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M.N., Wolpert, D.M., 2016. A common mechanism underlies changes of mind about decisions and confidence. Elife 5, e12192. https://doi.org/10.7554/eLife.12192.

Vickers, D., 1979. Decision Processes in Visual Perception. Academic Press, New York.

Walker, E.Y., Cotton, R.J., Ma, W.J., Tolias, A.S., 2020. A neural basis of probabilistic computation in visual cortex. Nat. Neurosci. 23, 122—129. https://doi.org/10.1038/s41593-019-0554-5.

Webb, T., Miyoshi, K., So, T.Y., Lau, H., 2021. A task-optimized neural network model of decision confidence. Proc. Annu. Meet. Cogn. Sci. Soc. 43.

Webb, T.W., Miyoshi, K., So, T.Y., Rajananda, S., Lau, H., 2023. Natural statistics support a rational account of confidence biases. Nat. Commun. 14, 3992. https://doi.org/10.1038/s41467-023-39737-2.

Weiskrantz, L., Barbur, J.L., Sahraie, A., 1995. Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex (V1). Proc. Natl. Acad. Sci. U. S. A. 92, 6122—6126. https://doi.org/10.1073/pnas.92.13.6122.

Weiskrantz, L., Warrington, E.K., Sanders, M.D., Marshall, J., 1974. Visual capacity in the hemianopic field following a restricted occipital ablation. Brain 97, 709—728. https://doi.org/10.1093/brain/97.1.709.

Wells, G.L., Kovera, M.B., Douglass, A.B., Brewer, N., Meissner, C.A., Wixted, J.T., 2020. Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. Law Hum. Behav. 44, 3—36. https://doi.org/10.1037/lhb0000359.

Wessel, J.R., Danielmeier, C., Ullsperger, M., 2011. Error awareness revisited: accumulation of multimodal evidence from central and autonomic nervous systems. J. Cogn. Neurosci. 23, 3021—3036. https://doi.org/10.1162/jocn.2011.21635.

Wheeler, M.A., Stuss, D.T., Tulving, E., 1995. Frontal lobe damage produces episodic memory impairment. J. Int. Neuropsychol. Soc. 1, 525—536. https://doi.org/10.1017/S1355617700000655.

Windschitl, P.D., Chambers, J.R., 2004. The dud-alternative effect in likelihood judgment. J. Exp. Psychol. Learn. Mem. Cogn. 30, 198—215. https://doi.org/10.1037/0278-7393.30.1.198.

Wixted, J.T., 2007. Dual-process theory and signal-detection theory of recognition memory. Psychol. Rev. 114, 152—176. https://doi.org/10.1037/0033-295X.114.1.152.

Wixted, J.T., Wells, G.L., 2017. The relationship between eyewitness confidence and identification accuracy: a new synthesis. Psychol. Sci. Public Interest 18, 10—65. https://doi.org/10.1177/1529100616686966.

Xie, S., Kaiser, D., Cichy, R.M., 2020. Visual imagery and perception share neural representations in the alpha frequency band. Curr. Biol. 30, 2621—2627. https://doi.org/10.1016/j.cub.2020.04.074.

Xue, K., Zheng, Y., Rafiei, F., Rahnev, D., 2023a. The timing of confidence computations in human prefrontal cortex. bioRxiv. https://doi.org/10.1101/2023.03.21.533662.

Xue, K., Shekhar, M., Rahnev, D., 2023b. Challenging the Bayesian confidence hypothesis. PsyArXiv. https://doi.org/10.31234/osf.io/mf5zp.

Yallak, E., Balcı, F., 2021. Metric error monitoring: another generalized mechanism for magnitude representations? Cognition 210, 104532. https://doi.org/10.1016/j.cognition.2020.104532.

Ye, Q., Zou, F., Lau, H., Hu, Y., Kwok, S.C., 2018. Causal evidence for mnemonic metacognition in human precuneus. J. Neurosci. 38, 6379—6387. https://doi.org/10.1523/JNEUROSCI.0660-18.2018.

Yeung, N., Summerfield, C., 2012. Metacognition in human decision-making: confidence and error monitoring. Philos. Trans. R. Soc. B Biol. Sci. 367, 1310—1321. https://doi.org/10.1098/rstb.2011.0416.

Yonelinas, A.P., 1994. Receiver-operating characteristics in recognition memory: evidence for a dual-process model. J. Exp. Psychol. Learn. Mem. Cogn. 20, 1341—1354. https://doi.org/10.1037/0278-7393.20.6.1341.

Yoshida, M., Isa, T., 2015. Signal detection analysis of blindsight in monkeys. Sci. Rep. 5, 10755. https://doi.org/10.1038/srep10755.

Yu, S., Pleskac, T.J., Zeigenfuse, M.D., 2015. Dynamics of postdecisional processing of confidence. J. Exp. Psychol. Gen. 144, 489—510. https://doi.org/10.1037/xge0000062.

Zawadzka, K., Higham, P.A., Hanczakowski, M., 2017. Confidence in forced-choice recognition: what underlies the ratings? J. Exp. Psychol. Learn. Mem. Cogn. 43, 552—564. https://doi.org/10.1037/xlm0000321.

Zeki, S., Ffytche, D.H., 1998. The Riddoch syndrome: insights into the neurobiology of conscious vision. Brain 121, 25—45. https://doi.org/10.1093/brain/121.1.25.

Zylberberg, A., Barttfeld, P., Sigman, M., 2012. The construction of confidence in a perceptual decision. Front. Integr. Neurosci. 6, 79. https://doi.org/10.3389/fnint.2012.00079.