HAL open science

Consensus goals for the field of visual metacognition

Dobromir Rahnev, Tarryn Balsdon, Lucie Charles, Vincent de Gardelle, Rachel Denison, Kobe Desender, Nathan Faivre, Elisa Filevich, Stephen M Fleming, Janneke Jehee, et al.

To cite this version:

Dobromir Rahnev, Tarryn Balsdon, Lucie Charles, Vincent de Gardelle, Rachel Denison, et al.. Consensus goals for the field of visual metacognition. Perspectives on Psychological Science, 2022, 17 (6), pp.1746-1765. 10.1177/17456916221075615 . hal-03215411v2

HAL Id: hal-03215411

https://hal.science/hal-03215411v2

Submitted on 20 Jul 2022 (v2), last revised 9 Jan 2023 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Consensus goals in the field of visual metacognition

Dobromir Rahnev*, School of Psychology, Georgia Institute of Technology, USA
Tarryn Balsdon, Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France
Lucie Charles, Institute of Cognitive Neuroscience, University College London, UK
Vincent de Gardelle, Paris School of Economics & CNRS, Paris, France
Rachel Denison, Department of Psychological and Brain Sciences, Boston University, USA
Kobe Desender, Brain and Cognition, KU Leuven, Belgium
Nathan Faivre, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France
Elisa Filevich, Bernstein Center for Computational Neuroscience Berlin, Philippstraße 13 Haus 6, 10115 Berlin, Germany
Stephen M. Fleming, Department of Experimental Psychology and Wellcome Centre for Human Neuroimaging, University College London, UK
Janneke Jehee, Donders Institute, Radboud University, Netherlands
Hakwan Lau, Department of Psychology, UCLA
Alan L. F. Lee, Department of Applied Psychology and Wofoo Joseph Lee Consulting and Counselling Psychology Research Centre, Lingnan University, Hong Kong
Shannon M. Locke, Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France
Pascal Mamassian, Laboratoire des systèmes perceptifs, Département d'études cognitives, École normale supérieure, PSL University, CNRS, Paris, France
Brian Odegaard, Department of Psychology, University of Florida, Gainesville, FL USA
Megan Peters, Department of Cognitive Sciences, University of California Irvine, Irvine, CA USA
Gabriel Reyes, Facultad de Psicología, Universidad del Desarrollo, Santiago, Chile
Marion Rouault, Département d'Études Cognitives, École Normale Supérieure, Université Paris Sciences & Lettres (PSL University), Paris, France
Jerome Sackur, Département d'Études Cognitives, École Normale Supérieure, Université Paris Sciences & Lettres (PSL University), Paris, France
Jason Samaha, Department of Psychology, University of California, Santa Cruz
Claire Sergent, Université de Paris, INCC UMR 8002, 75006, Paris, France
Maxine T. Sherman, Sackler Centre for Consciousness Science, University of Sussex, Brighton, UK
Marta Siedlecka, Consciousness Lab, Institute of Psychology, Jagiellonian University, Kraków, Poland
David Soto, Basque Center on Cognition Brain and Language, San Sebastián, Spain. Ikerbasque, Basque Foundation for Science, Bilbao, Spain.
Alexandra Vlassova, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands.
Ariel Zylberberg, Department of Brain and Cognitive Sciences, University of Rochester, USA

Keywords: visual metacognition, confidence, perceptual decision making, goals

***Correspondence**: rahnev@psych.gatech.edu

Supplementary material: https://osf.io/v6qje/

Abstract

Despite the tangible progress in psychological and cognitive sciences over the last several years, these disciplines still trail other more mature sciences in identifying the most important questions that need to be solved. Reaching such consensus could lead to greater synergy across different laboratories, faster progress, and increased focus on solving important problems rather than pursuing isolated, niche efforts. Here, 26 researchers from the field of visual metacognition reached consensus on four long-term and two medium-term common goals. We describe the process that we followed, the goals themselves, and our plans for accomplishing these goals. If this effort proves successful within the next few years, such consensus-building around common goals could be adopted more widely in psychological science.

Introduction

"The trouble with not having a goal is that you can spend your life running up and down the field and never score."

Bill Copeland

The need for common goals in science

There is considerable debate among philosophers about what constitutes progress in science (Feller & Stern, 2007). Nevertheless, two broad themes appear in most accounts. First, scientific progress requires the accumulation of solid, agreed-upon empirical knowledge (Bird, 2007). Second, scientific progress requires theories and models that predict and explain the various empirical findings in a field (Guest & Martin, 2021; Muthukrishna & Henrich, 2019; van Rooij & Baggio, 2021). These two components of scientific progress are in constant interplay with each other: new findings lead to refined theories, which in turn motivate the collection of new and different empirical data to test them.

One factor that may accelerate scientific progress is the existence of common goals in a given discipline. Indeed, if most topics in a field are tackled by only one or a few labs, it becomes difficult to build both an agreed-upon empirical knowledge and robust theories. Such difficulties are apparent to various degrees in many subdisciplines of psychological and cognitive science.

Common goals could have transformative effects on research fields. They can lead to greater synergy among research groups and thus faster progress. In addition, spurious findings are more likely to be weeded out when many groups work toward a common goal. An inspiring example within psychology has been the goal of measuring the replicability of psychological science. The goal has rallied hundreds of laboratories and has led to genuine answers in a few short years (Klein et al., 2018; Open Science Collaboration, 2015) and large-scale collaborations such as the Psychological Science Accelerator. It is clear that this progress would not have been made in the absence of a common goal that served to focus the energies of many researchers. Yet, clearly defined common goals remain largely absent in basic experimental psychology.

Potential drawbacks of common goal setting

Although it is easy to identify potential benefits of common goal setting, it is also possible to think of potential drawbacks. Here we discuss several potential disadvantages of such goal setting that mostly relate to adopting an extreme approach where the common goals completely displace the creativity and innovation of individual researchers. We also explore simple measures to mitigate such drawbacks.

Perhaps the most important drawback is the potential of common goals to stymie innovation. Indeed, if individual researchers abandon their interests and only work on a small set of common goals, many important discoveries may not be made. A healthy level of diversity of goals is important for a discipline (Kording et al., 2018), while an obsession with just a few narrow paths can lead to "tunnel vision." Yet, agreeing on common goals in no way implies that researchers should stop exploring a multitude of research questions and directions. Indeed, we believe that few, if any, researchers would abandon promising leads that fall outside of the common goals. Certainly, none of the current authors plan to do so. Similarly, we doubt that publishers or grant agencies will stop supporting research outside of the common goals and we would certainly discourage them from doing so. In the context of organizations, the existence of a "goals paradox" has been suggested, where both congruence and diversity in organizations' goals are needed for success in collaboration (Vangen & Huxham, 2012). Similarly, goal setting in science should strive to bring about more congruence but not at the expense of diversity.

A second possible concern is that common goal-setting may overturn standard scientific practices. Indeed, research programs often evolve organically around new theories and empirical findings. If this process were fully replaced by explicitly setting goals that scientists should strive to meet, then the organic evolution of research programs would be disrupted. However, the existence of common goals does not prevent researchers from following new leads as in standard scientific practice. Instead, they can help break tendencies to only seek confirmatory evidence for one's favorite theories (Yaron et al., 2021) and enable adversarial collaborations where researchers from different camps work together to resolve their differences (Melloni et al., 2021).

A final potential concern is about the meaning of the word “goal” and what is included under it. We do not think that there is one correct answer and common goals for different fields can be defined on many different levels. Here, we adopt a very broad conception of the term “goal” that encompasses both broad and narrow scientific questions and research directions. These goals can include topics already studied extensively as well as completely new avenues of research. It is possible that a narrower conception of the word “goal” would be more beneficial for more established fields, but such a broad definition seems preferable for newer fields such as ours.

Ultimately, assessing the advantages and disadvantages of common goal setting in science requires data. We are unaware of equivalent efforts in other fields and therefore of relevant data that we can use for this assessment. We hope that the current effort will be one critical data point that can inform our understanding of the value of common goal setting in science.

Creating common goals for the field of visual metacognition

Here, 26 researchers from the field of visual metacognition -- a field of study focused on understanding the subjective evaluation and control in visual perception -- organized around the idea of specifying common goals. We start by giving a brief timeline on the process that we followed, then discuss the specific goals that we agreed on, and end with our strategies for follow-up and evaluation.

The idea for coming up with common goals for our field was born in the summer of 2020. We gathered a group of people working on the topic of confidence and metacognition in perception. We sought to assemble a relatively small group that was diverse in terms of career stage, geographical location, and gender. We did not follow a formal methodology and did not have strict criteria for inclusion when assembling the group, so the authors represent one slice rather than a representative sample of researchers from the field.

To construct an initial list of possible goals, each person was encouraged to submit anonymous entries for what they perceived to be the most important goals in the field. We separated these into two categories: long-term goals, which aim to set a direction for the field and are not expected to be resolved for at least the next ten years, and medium-term goals where concrete progress can be expected in the next five years. This process resulted in 26 long-term goals and 39 medium-term goals. The wording of the goals was then standardized, and all goals were anonymously rated by the same group of researchers on several categories including their importance, clarity, likely success, and likelihood of wide adoption. All proposed goals and raw ratings are included as Supplementary Material. The goals were then sorted based on the answers to the question "Is this goal among the 2-3 goals that should be adopted by the field?" This process resulted in six highly-rated long-term goals and six highly-rated medium-term goals. Everyone was allowed to "rescue" other goals but nobody did. All of these steps were carried out online over approximately four months.

We then held two 3-hour online workshops, three days apart, where we debated the merits of the top-rated goals from both categories. The first workshop covered the long-term goals; the second workshop covered the medium-term goals. In each case, the pros and cons of each goal were thoroughly discussed and one final round of voting took place. Based on these final ratings, each workshop ended with a decision on the consensus goals from each category. The process resulted in four long-term goals and two medium-term goals. The ratings from these meetings are also available as Supplementary material.

Finally, we discussed the best process for following up on these goals, with the discussion starting during the workshop but continuing over the next several months. Writing the current paper served to (1) formalize each goal, (2) publicly announce the goals to both generate commitment and encourage the involvement of the wider research community, and (3) inform researchers from other fields about our process in case other subfields of psychology want to engage in similar goal-setting. All goals, together with the links between them, are graphically presented in Figure 1.

While we were able to reach a consensus, it should be noted that the process was far from easy. The large number of initially proposed goals demonstrates the existence of a large diversity of topics, approaches, and priorities in the field of visual metacognition (similar diversity exists in related fields such as computational neuroscience; Kording et al., 2018). Zeroing in on only a small minority of goals meant that the great majority of proposed goals were not selected as consensus goals regardless of how strongly the people who proposed them may have felt about them. The two workshops further demonstrated that we did not initially share a common vision for progress in the field. Arriving at a consensus strongly depended on the existence of an abundance of goodwill among the participants and the absence of "warring factions." We include suggestions on optimizing the process of arriving at shared goals in the Supplementary.

A very brief introduction to visual metacognition

We define “visual metacognition” broadly as the study of the subjective evaluation and control of one’s own cognitive processes and behavioral responses during visual perceptual tasks (Nelson & Narens, 1990). Most tasks in the field feature simple perceptual judgments (e.g., discriminating between two possible stimuli such as left- and right-tilted Gabor patches, though more complex tasks such as multi-alternative decisions and estimation tasks are also used). This Type-1, object-level judgment is then supplemented by a Type-2, subjective judgment, usually in the form of a confidence rating. The field has its roots in 19th-century psychophysics (Fechner, 1860; Helmholtz, 1856), which often used confidence ratings to infer the perceptual

experience of the subject (Peirce & Jastrow, 1884). However, the last decade has seen both a substantial growth and a change of focus to understanding self-evaluation itself rather than simply using it as a tool to understand perception (Fleming et al., 2012; Mamassian, 2016; Rahnev, 2021). The field is rapidly maturing and growing, with many investigators from diverse fields such as computational neuroscience, animal neurophysiology, judgment and decision making, and psychometrics becoming increasingly involved. To make the current paper easier to follow for non-specialists, we provide a glossary of common terms that appear in this paper.

Glossary

Term	Definition
Accumulation-to-bound models	A set of models of decision-making that assume an underlying process of accumulation of evidence to a threshold.
Metacognitive bias	An increase or decrease of confidence level despite basic task performance remaining constant.
Metacognitive efficiency	The ability to distinguish between one's own correct and incorrect responses given a certain level of Type-1 performance
Metacognitive noise	A type of noise that affects confidence ratings but not primary decisions.
Metacognitive sensitivity	The ability to distinguish between one's own correct and incorrect responses.
Signal detection theory (SDT)	A theory of perceptual decision making used to model choice behavior (often in two-choice tasks) that considers the across-trial variability in internal evidence for each stimulus category.
Type-1 vs. Type-2 decisions	Type-1 decisions are about the primary task, while Type-2 decisions are about the quality of the Type-1 response.
Type-1 vs. Type-2 task performance	Type-1 task performance indicates how well one's choices predict stimulus identity, whereas Type-2 task performance indicates how well one's subjective ratings predict one's accuracy (i.e., metacognitive sensitivity).

Overview of the consensus goals

We agreed on four long-term and two medium-term goals. All six goals are focused on basic science. This fact largely reflects the current composition and priorities in the field but may also suggest the need for more attention towards applied research in the future. All goals should be accessible to most labs in the field as well as to researchers of all career stages. The selected goals represent a mixture of theoretical and technical components. More specifically, long-term goals 3-4 and medium-term goal 2 are largely theoretical, whereas long-term goals 1-2 and medium-term goal 1 have a dual focus on both technical and theoretical developments. No goal is purely technical -- the models, techniques, and manipulations that different goals seek to develop ultimately gain their significance from their role in answering theoretical questions. Finally, some goals are comparatively narrow (e.g., long-term goals 2 and 4), some are quite broad (e.g., long-term goal 1), and one goal (medium-term goal 2) became broad during our discussion as it was made to encompass three different but related initial entries.

It should be appreciated that the great majority of the initially proposed goals were not selected. These goals varied substantially. A post hoc analysis of these goals categorized only seven of them as closely related to the selected goals, and 49 as unrelated or very remotely related to the selected goals. Some of the most common themes among the non-selected goals included the relationship of metacognition and psychopathology (4 goals), the proper measurement of metacognitive ability (4 goals), the relationship between metacognition and consciousness (3 goals), the neural correlates of visual metacognition (3 goals) and modeling visual metacognition (3 goals). This variability demonstrates the diversity of perspectives, objectives, and methodologies in the field, and thus perhaps further underscores the need for common goal setting.

Long-term goals for the field of visual metacognition

We decided to adopt four long-term goals, and have committed to incorporating them into our research programs. We view these goals as setting a direction and do not expect that any of them will be resolved for at least the next ten years and perhaps beyond. For each goal, we explain why it is important, give a brief background on relevant research and methodologies, and put forward our current thoughts on what needs to be done to ultimately achieve that goal.

Long-term goal 1: Develop falsifiable and detailed computational models of visual metacognition

Why is this goal important?

To achieve progress in our understanding of visual metacognition, a key long-term goal is to develop detailed and falsifiable computational models that explain the implementation of visual metacognition. Both cognitive models that focus on behavior and models that explain data from neural recordings are needed. Although such modeling is a worthy goal in and of itself (by allowing, for example, to predict human behavior; Yarkoni & Westfall, 2017), it is also critical for our theoretical understanding of the mechanisms of visual metacognition. A computational model goes beyond a conceptual, verbal description and translates a specific theory into math making it more precise and unambiguous (Guest & Martin, 2021; van Rooij & Baggio, 2021). Moreover, translating verbal theories into computational models often clarifies the hidden assumptions in the theories. Within the context of visual metacognition, computational modeling can clarify which sources of evidence, internal and external, contribute to reported confidence, reveal the extent to which confidence involves normative computations or heuristics, constrain theories regarding the architecture of metacognition, etc. For such modeling to be useful, models must be sufficiently detailed, provide clear falsifiable hypotheses, and fit actual behavioral and neural data well. Given that modeling of visual metacognition is still in its infancy, this long-term goal is necessarily rather broad by encompassing both cognitive and neural models of any task that involves visual metacognition. We expect that as the field matures, it will become easier and more productive to set narrower modeling goals.

Background

Before providing a roadmap for future developments, we first discuss some of the current models of visual metacognition and their limitations and shortcomings. Much of the early work was inspired by signal detection theory or SDT (Green & Swets, 1966). This framework describes how human observers categorize noisy measurements of a signal by placing a criterion in the measurement space. By imposing additional criteria, the same framework can also be extended to explain how human observers can give a graded evaluation of the quality of their decision (Clarke et al., 1959; Galvin et al., 2003; Maniscalco & Lau, 2012). Thus, within this framework visual metacognition is directly related to the strength of the evidence in that

observers will be more certain about their choice if the evidence sample lies far from the decision criterion.

An important limitation of SDT is that it does not consider within-trial dynamics, but instead only makes predictions about end-of-trial choices. Therefore, such models cannot easily account for influences of speed-accuracy tradeoffs on confidence or allow for changes of mind within the course of a trial (Resulaj et al., 2009). A natural extension of SDT that does consider within-trial dynamics is a class of models based on the accumulation-to-bound principle. Within such models, choices are thought to reflect the noisy accumulation of evidence until a threshold is reached. To account for visual metacognition, several extensions of these models have been proposed. For example, visual metacognition can be quantified as the degree of evidence extracted from additional post-decisional evidence accumulation following the initial boundary crossing (Pleskac & Busemeyer, 2010), as the difference in magnitude between two accumulators (Vickers, 1979), or as the probability that a choice was correct (Kiani & Shadlen, 2009).

An important distinction in current models is that between single-pathway, dual-pathway, and hierarchical models (Fleming & Daw, 2017; Maniscalco & Lau, 2016). According to single-pathway models, a single source of evidence, corrupted with sensory noise, informs both perceptual choices and metacognitive choices. According to dual-pathway models, perceptual and metacognitive choices reflect information corrupted by independent noise sources. Finally, according to hierarchical models, metacognitive choices are based on the corrupted signal that was used to inform the perceptual choice with additional metacognitive noise applied.

The work ahead

As the brief background above shows, several existing models of decision-making can each be extended to incorporate visual metacognition. Yet, many of these models make very similar predictions. For example, one key characteristic of visual metacognition is that choice accuracy usually monotonically increases as a function of decision confidence (Kepecs & Mainen, 2012). However, this pattern is predicted by virtually all theories of visual metacognition. As such, despite being a key aspect of metacognition, such a pattern does not appear informative to distinguish different models. Therefore, the major challenge ahead will be to find ways that allow us to behaviorally differentiate between models of visual metacognition. Two differentiable models will have certain scenarios where they make divergent predictions about behavior. Thus, in addition to giving a computational description of the model, researchers will also need to inspect the models theoretically or by using simulations to identify these key choice contexts where the models are differentiable (Shekhar & Rahnev, 2021a). Preferably, the models should also emphasize biological plausibility in that each algorithmic step can be represented as a neural process (e.g., population coding). These two elements, falsifiability and biological plausibility, would allow for behavioral and neural tests to narrow down the most likely processes underlying visual metacognition, allowing for consensus-building and a greater ability to report and compare fits to metacognitive behavior across studies.

Long-term goal 2: Develop robust protocols to manipulate one's metacognition and investigate if such protocols facilitate adaptive performance

Why is this goal important?

This goal relates to two important questions: what is the function of visual metacognition and can visual metacognition be manipulated experimentally. As already mentioned, metacognition plays both monitoring and regulatory roles (Nelson & Narens, 1990). Research on visual metacognition has paid little attention to its specific functions, although it has been suggested that perceptual confidence might guide perceptual learning (Guggenmos et al., 2016),

associative learning (Hainguerlot et al., 2018), task prioritization (Aguilar-Lleyda et al., 2020), and moderate sensory evidence accumulation (Balsdon et al., 2020). However, in most studies, visual metacognition has not been directly manipulated leaving the causal role of metacognition in behavior unclear. Developing novel protocols to robustly manipulate metacognition will have great methodological, theoretical, and even clinical significance (Moritz & Woodward, 2007).

Background

Manipulations of metacognitive efficiency

Many studies have reported manipulations that modulated metacognitive efficiency. One group of studies used manipulations related to stress. For example, it has been shown that individual predisposition to stress (i.e., cortisol) reactivity, and the administration of cortisol-like drugs, is associated with reduced metacognitive sensitivity (Reyes et al., 2015, 2020). Similarly, other studies suggested that blocking noradrenergic transmission can improve metacognitive efficiency (Allen et al., 2016), and that meditation training can improve metacognition in memory but not in perception (Baird et al., 2014; but see also Schmidt et al., 2019).

Other studies examined the effects of manipulations of cognitive load or direct stimulation of the prefrontal cortex on metacognitive efficiency. Loading the capacity of working memory systems has been shown to impair metacognitive performance for perceptual decisions (Maniscalco & Lau, 2015; Schmidt et al., 2019; but see Konishi et al., 2020). This effect may reflect the necessary role of neural circuitry involving the dorsolateral prefrontal cortex that is shared among both working memory and metacognition (Feredoes et al., 2011). Relatedly, transcranial magnetic stimulation (TMS) of the dorsolateral prefrontal cortex (Rounis et al., 2010; but see Bor et al., 2017) or anterior prefrontal cortex (Rahnev et al., 2016; Ryals et al., 2016; Shekhar & Rahnev, 2018) have also shown modulations of metacognition.

Other manipulations shown to affect metacognition include experience-dependent training in a visual imagery task (Rademaker & Pearson, 2012), the engagement of visual attention or expectation (Mei et al., 2020; Sherman et al., 2015), and changing the order of Type-1 and Type-2 confidence responses (Wierzchoń et al., 2014). Currently, there is mixed evidence on whether metacognition can be improved using feedback (Carpenter et al., 2019; de Gardelle et al., 2020; Haddara & Rahnev, 2021).

Manipulations of confidence

Several studies have attempted to selectively modulate the overall level of confidence while holding Type-1 performance and/or metacognitive efficiency constant. By causally and selectively modulating confidence, such an approach can be useful for understanding the function that perceptual confidence plays for other aspects of behavior. One popular manipulation is the positive evidence bias, in which the signal and noise components of a visual stimulus are both increased while keeping the signal-to-noise ratio approximately intact (Zylberberg et al., 2012). This paradigm has been used to show that increasing confidence does not facilitate cognitive control (Koizumi et al., 2015) or working memory (Samaha et al., 2016), thus constraining theories on how confidence relates to other higher-order cognitive processes.

However, other work has documented significant effects of confidence on other aspects of behavior. For example, increasing perceptual confidence (independently of accuracy) in a first decision biases evidence accumulation for one's subsequent decision in favor of the initial choice (Rollwage et al., 2020). Relatedly, selectively boosting confidence increased both the attractive and repulsive serial biases typically observed across trials in visual perception tasks (Samaha et al., 2019). Confidence manipulations have also been shown to influence one's decisions to seek additional information (Desender et al., 2018). These effects suggest that

confidence in a perceptual decision, independent of decision accuracy, modulates how perceptual evidence is used to guide subsequent behavior.

The work ahead

The main challenge ahead is three-fold: validating existing manipulations of metacognitive efficiency and confidence, finding novel ways to manipulate metacognition in a way that produces generalizable effects on cognition and behavior, and developing a sound understanding of when, why, and how these effects occur. Further research is needed to test the effect of different types of feedback signals (e.g., based on the accuracy of confidence judgments) or brain markers of metacognitive skill (e.g., via neurofeedback training; Cortese et al., 2016). Another promising direction is to further develop existing neurostimulation interventions (i.e., based on TMS, transcranial direct current stimulation, or pharmacological interventions) to target the mechanisms of metacognition in a way that produces reliable changes in confidence that impact subsequent behavioral performance. We can expect progress on several of these fronts already in the next five years and have consequently discussed whether the whole goal here should be in the medium-term category. Yet, we felt that the current goal is long-term since it is important to develop multiple manipulations of metacognition, investigate whether each facilitates adaptive performance, and compare the results. This process is likely to take time. Ultimately, this line of work should reveal whether metacognitive interventions can support adaptive behavioral performance across different sensory modalities and cognitive tasks, and whether these interventions are sufficiently strong and long-lasting to allow clinical applications.

Long-term goal 3: Determine the computations underlying confidence in tasks of increasingly higher complexity

Why is this goal important?

In the real world, confidence accompanies a wide variety of decisions and is used not only as a form of self-reflection but also as a way to shape how we plan subsequent actions, learn from past errors, and communicate our decisions to others. Characterizing these processes with tasks of increasingly higher complexity will allow us to broaden our conceptualization of visual metacognition. Important next steps include examining confidence in decisions between more than two alternatives, decisions that unfold over prolonged time scales, and decisions that require actively seeking information (Desender et al., 2018; Rouault et al., 2021). In addition, increased task complexity is necessary for understanding the relationship between confidence and other forms of visual metacognition, such as introspection about task strategy, decision time, and the conscious experience of sensory stimuli (see long-term goal 4).

Background

Confidence has usually been studied by asking people to evaluate their performance on simple two-choice tasks. Typical tasks include deciding whether a stimulus is novel or familiar, comparing the orientation of two visual stimuli, or reporting the net direction of motion of randomly moving dots (Kiani & Shadlen, 2009). Focusing the study of confidence on binary decisions has made it possible to relate confidence to decision accuracy and decision time (Kiani et al., 2014). It has also led to the development of precise computational models of confidence in binary decisions (Maniscalco & Lau, 2016; Shekhar & Rahnev, 2021b; Vickers, 1979), and enabled the study of confidence in non-human animals (Kepecs et al., 2008; Kiani & Shadlen, 2009; Masset et al., 2020).

The study of confidence in simple perceptual decisions has laid solid foundations for expansion to tasks that more closely resemble its formation and use in the real world (Rahnev, 2020). Confidence affects how we plan subsequent actions, which has been studied with tasks that

comprise multiple sub-decisions - akin to real-world decisions like preparing a dish or finding a route to a destination. In a task in which two correct decisions were required to obtain a reward, van den Berg et al. (2016) showed that participants adjusted the speed and accuracy of a second decision depending on their confidence in the first. This establishes a role for confidence in regulating the speed-accuracy tradeoff for subsequent decisions, a strategy that maximizes overall reward (Balsdon et al., 2020). The study of tasks in which different sources of information have to be combined to make a decision has shown that confidence is also used to infer the cause of an error. Purcell & Kiani (2016) showed that human participants integrate expected accuracy (or confidence) over multiple decisions to infer when a strategy that was useful in the past is no longer effective, and neural correlates of confidence-guided strategy selection have been found in monkeys (Sarafyazd & Jazayeri, 2019). This line of research highlights how confidence in propositions that span multiple individual decisions ("I'm good at this task") can be built from confidence in individual decisions ("I made this decision correctly") (Lee et al., 2021; Mamassian, 2020; Rouault et al., 2019; Zylberberg et al., 2018).

Confidence also affects how we communicate our decisions to others and how we weigh their opinions. Bahrami et al. (2010) showed that two decision-makers facing the same decision can achieve better performance than each one alone if they can exchange their confidence judgments. Confidence and metacognition influence how we judge the intention and expertise of other agents (Pescetelli & Yeung, 2021) and decide whether to seek advice or information before committing to a decision (Rouault et al., 2021). These studies have leveraged what has been learned about confidence from the study of isolated decisions to approach the more complex functions of confidence.

The work ahead

Despite recent efforts, a gap remains between the tasks used to study confidence and the complexity of both the kinds of perceptual decisions and confidence evaluations characteristic of everyday life.

In realistic contexts, percepts are formed by combining multiple cues, often weighted by their reliability (Trommershäuser et al., 2011). It is unclear whether people have metacognitive access to the uncertainty associated with low-level cues or only to the final unified percept (Deroy et al., 2016). The primary task can also have many more than two decision alternatives. Even simple extensions from binary to ternary decisions have shown that, similar to findings in executive function (Collins & Koechlin, 2012), metacognition may be limited to tracking only the best two alternatives (H.-H. Li & Ma, 2020). A related question is whether confidence only encodes a few discrete levels (Lisi et al., 2020; Zhang & Maloney, 2012) or a continuous representation of perceptual evidence (Swets et al., 1961). Paradigms involving visual search (Gajdos, Régner, et al., 2019), tracking moving stimuli (Locke et al., 2020), and active sampling (Rouault et al., 2021) can reveal the complex interplay of different cues to confidence (Boldt et al., 2017). Another aspect is determining which cues contribute to global and prospective confidence estimates (Lee et al., 2021; Mamassian, 2020; Mei et al., 2020; Rouault et al., 2019; Siedlecka et al., 2016), and how they may interact with “local” confidence in a single decision.

Normative models posit that confidence tracks the probability of a decision being correct. However, observers have been found to deviate from optimal computations (Rahnev & Denison, 2018). Relating confidence to other forms of introspection, such as observers reporting on their cognitive strategy, decision-time, or even stimulus visibility, is important for building a comprehensive theory of metacognition. Finally, the development of implicit measures of confidence would be particularly useful for the study of confidence in non-human animals (beyond the use of response times and willingness to wait for a reward; Kepecs et al., 2008;

Masset et al., 2020). It has been shown that confidence is reflected in neural markers such as pupil dilation (Allen et al., 2016; Balsdon et al., 2020; Lempert et al., 2015; Urai et al., 2017), and the P300 component (Zakrzewski et al., 2019) and central parietal positivity (Boldt et al., 2019; Herding et al., 2019) obtained from electroencephalographic recordings. Further research is necessary to understand how one or a combination of these measures could be used to assess metacognitive accuracy, and how they are related to the neural computation of confidence.

Specific directions that are especially promising for immediate progress are suggested in medium-term goal 1, which is functionally equivalent to the current long-term goal. In addition, understanding the computations underlying confidence in tasks of increasing complexity will require continuous progress on modeling confidence (see long-term goal 1 and medium-term goal 2) with the ultimate goal that models of metacognition should generalize across paradigms to contribute to a unified framework.

Long-term goal 4: Determine the nature of the relationship between perceptual metacognition and perceptual consciousness

Why is this goal important?

Perceptual metacognition and perceptual consciousness are traditionally seen as closely linked; however, their relationship is not fully understood and varies dramatically across theoretical frameworks. So-called first-order theories of consciousness (e.g., Block, 2007; Lamme, 2000) posit that only recurrent activity in early sensory areas is required for consciousness and that metacognition is a post-perceptual cognitive process with no direct link with phenomenal experience. By contrast, according to higher-order theories (HOT), perceptual consciousness is linked to higher-order reflective processes that represent or monitor first-order contents stemming from sensory responses (Lau & Rosenthal, 2011). However, the meta-level representations and self-reflective processes that are critical for conscious experience in HOT need not be similar to the components of metacognitive confidence (Brown et al., 2019), and, as we will review below, metacognition can be dissociated from perceptual consciousness. The global neuronal workspace model distinguishes components of consciousness based on the global availability of information within cognitive and action systems, and self-monitoring or metacognition (Dehaene, 2014). Corroborating this distinction, a recent paper suggests that the network that subtends such global availability during conscious perception takes a different form according to whether participants are requested to decide on their perception or not (Sergent et al., 2021). However, attempts have been made to explain the role of metacognition within this framework (Shea & Frith, 2019) by suggesting that confidence is a key feature of the representations held in the global workspace, which affords a common currency to integrate information from different sensory systems (de Gardelle & Mamassian, 2014; Faivre et al., 2018) and cognitive processes that may be re-used to guide subsequent behavior and mental function.

Empirical studies often assume a link between metacognition and consciousness, as metacognitive judgments are often used to make inferences about consciousness (e.g., Norman & Price, 2015). However, there is no agreement on whether such measures exhaustively capture all conscious contents and whether they allow for differentiating conscious from unconscious perception (e.g., Seth et al., 2008; Timmermans & Cleeremans, 2015). It has also been proposed that different types of metacognitive assessments measure different phenomena. So-called introspective or first-order judgments (e.g., visibility judgments) are thought to refer directly to one's visual experience, while second-order judgments (e.g., confidence ratings) refer to the evaluation of one's perceptual decision accuracy (Sandberg et al., 2011). Looking for dissociations between these two processes sheds light on whether an

accurate metacognitive assessment of perceptual performance depends on conscious perception (Jachs et al., 2015) or whether it can indicate the presence of conscious experience that cannot be verbalized and reported (Vandenbroucke et al., 2014).

Understanding the relationship between visual consciousness and metacognition, and pinpointing their common and distinct factors, will help both to better understand the nature and function of each construct and further develop theories in each field. Below we review the existing evidence for dissociations between perceptual consciousness and metacognition, focusing on how metacognitive judgments are made for information that is consciously experienced or not, and then provide an overview of the few studies that have attempted to examine the two phenomena simultaneously.

Background

Several lines of evidence suggest that conscious access may not be needed for the successful deployment of metacognition. For instance, Charles and colleagues (2013) assessed perceptual and metacognitive sensitivity in a number classification task across different levels of stimulus visibility. Their results showed that metacognitive processing of visual targets reported as unseen exceeded chance levels. Jachs et al. (2015) replicated these results and found that perceptual sensitivity strongly depended on visibility, while metacognitive sensitivity did so to a much lower extent. In addition, there is evidence that confidence judgments are diagnostic of visual memory accuracy even when participants display chance-level sensitivity in their first-order recognition judgments (Rosenthal et al., 2016; Scott et al., 2014). Finally, when attentional resources are constrained and participants report not seeing the target stimulus, confidence responses can discriminate between actual misses and correct rejections (Kanai et al., 2010; Meuwese et al., 2014). This dissociation between visibility and metacognition is consistent with there being a lower information threshold to make confidence estimates relative to phenomenological reports of visual experience (Zehetleitner & Rausch, 2013).

Our understanding of perceptual metacognition has mostly improved through the analysis of confidence ratings regarding discrimination tasks. Although discrimination tasks offer several practical advantages to compute metacognitive performance, only detection tasks allow a contrastive analysis of perceptual consciousness whereby the behavioral and neural responses evoked by seen vs. unseen stimuli are compared (Baars, 1997). Therefore, a simultaneous evaluation of perceptual consciousness and metacognition requires the collection of confidence ratings regarding the absence vs. presence of stimuli, which only a few studies have done. This is particularly important given that the neural underpinnings of metacognition for discrimination and detection differ qualitatively (Mazor et al., 2020). Among the studies that examined confidence in detection, an emerging pattern is that metacognitive performance is lower when judging stimulus absence vs. stimulus presence (Kanai et al., 2010; Meuwese et al., 2014), potentially in line with an asymmetric contribution of positive and negative evidence to confidence (Peters et al., 2017; Zylberberg et al., 2012) and/or unequal-variance SDT (Kellij et al., 2021; Mazor et al., 2021; Miyoshi & Lau, 2020). While the interplay between perceptual consciousness and metacognition is abundantly discussed at a theoretical level, empirical evidence bearing on this relationship is much scarcer. This interplay derives naturally from models assuming a common mechanism underlying detection and confidence responses. Recently, such a model was proposed considering a stimulus as consciously detected when a leaky evidence accumulation process reached a threshold and deriving confidence as the distance between the maximum of accumulated evidence and that threshold (Pereira et al., 2021). This latter definition of confidence notably explains how stimulus absence may be monitored and accounts for an asymmetry between positive and negative evidence mentioned above.

The work ahead

Future research needs to provide an account of how phenomenal experience, visibility, and confidence relate to computational models of human vision (Denison et al., 2020), generate and test novel predictions, and ultimately refine existing theories of consciousness. Among the hurdles of the work ahead, we note the need to match the level of performance when addressing the neurocognitive mechanisms supporting perceptual awareness and confidence (Morales et al., 2019), and develop novel paradigms that can concurrently assess both, without them being confounded with cognitive functions that are associated with reporting (e.g., attention, decision making, verbal report, response selection). There have been recent developments of so-called no-report paradigms to study the neural basis of perceptual consciousness while minimizing such confounds (Block, 2019; Tsuchiya et al., 2015) but there are currently no similar no-report paradigms for the concurrent assessment of metacognitive confidence and perceptual consciousness. While the present discussion focused on conscious contents, another line of research should also assess how metacognitive monitoring operates across distinct levels of consciousness or vigilance states.

Medium-term goals for the field of visual metacognition

In addition to the four long-term goals that set a general direction for research, we identified two medium-term goals. These medium-term goals are expected to yield progress within the timeframe of the next five years (i.e., we expect measurable progress by the end of 2026). For each of the two goals, we explain how it relates to the four long-term goals, where immediate progress appears most likely, and what we hope to achieve in the next five years. Unlike in the section on long-term goals, here we do not give extensive background for each goal since this background has already been covered in the related long-term goals.

Medium-term goal 1: Expand beyond confidence in two-choice tasks and develop models of confidence for such tasks

Why is this goal important and how does it relate to the long-term goals above?

This goal is strongly related to long-term goal 3, so much so that it can be considered a medium-term version of long-term goal 3. The present medium-term goal is also related to long-term goal 1, which outlined several models (signal detection theory, accumulation-to-bound models, single vs dual channel models) that are currently popular in explaining visual metacognition. Notably, most of these models are designed and tested in experiments where observers rate their confidence in a two-choice task. As a consequence, it is unclear whether the current models of visual metacognition can account for decision confidence in more complex cases, such as tasks with multiple alternatives or continuous judgments. Developing models that can explain visual metacognition in more complex tasks is of critical importance, not just because such models will have more ecological validity (and therefore will have wider explanatory power), but also because they may allow evaluation of the assumptions in current models in more challenging contexts. This will help researchers achieve long-term goal 1 by widening the scope of our models to a broader range of decision scenarios and providing more opportunities for divergence in model predictions. In addition, any progress on this goal will also contribute to the more general long-term goal 3.

The work ahead

Current models of visual metacognition, which mostly apply to two-choice tasks, are inherently limited in scope but it is not necessarily clear how they should be extended. Below, we present what we consider to be the four most promising directions where immediate progress can be made.

First, the most straightforward extension of current models would be to expand them from two-choice tasks to *n*-alternative choice tasks. For example, accumulation-to-bound models that can account for behavior in *n*-alternative choice tasks have been described (Ratcliff & Starns, 2013). Similarly, Li & Ma (2020) have proposed several plausible models for *n*-alternative decisions. Thus, a clear target for future developments would be to continue with these previous attempts and/or expand existing models, testing each model's validity in capturing behavior in *n*-alternative choice tasks in a wide range of perceptual tasks.

A second more ambitious target is to expand current models so that they can explain confidence when estimating a continuous quantity, such as the confidence one has that the orientation of a stimulus was correctly reproduced. In such cases, asking the observer to report the probability they were correct seems unsatisfactory as the observer will rarely be perfect in their report. Instead, their confidence should reflect the degree of error in the estimate. Several studies have already collected data on tasks that involve estimating a continuous quantity (e.g., Graf et al., 2005; Yallak & Balcı, 2021; Yoo et al., 2018) and several such datasets are available in the Confidence Database (Rahnev et al., 2020). The next steps would involve building models of visual metacognition that explain confidence ratings in such tasks.

Third, one step further would be to examine visual metacognition of ongoing perception. Due to the subjective nature of metacognitive reports, visual metacognition is usually queried jointly with or shortly after a choice. However, this does not imply that observers have no metacognitive experiences during the choice formation itself. In fact, there is some evidence that metacognition emerges online during choice formation (Dotan et al., 2018) and that it even controls the termination of the choice formation process (Balsdon et al., 2020). Such online expressions of metacognition pose a challenge for current models of visual metacognition, which usually describe metacognition as a (post-decision) read-out of the decision process. Thus, a clear target for future work will be to develop protocols that allow for robust online measurement of metacognition, and models that can explain such reports.

Finally, the fourth target for model developments is to explain perception-action interactions. There is increasing interest in examining visual behavior in dynamic scenarios where perception and action are both at play (Bonnen et al., 2015; Huk et al., 2018), which increases decision complexity. Thus, confidence can emerge as part of perception and action loops, such as reaching a series of targets or tracking just one (Locke et al., 2020). Rather than a simple button press, the response can be highly varied or of a continuous nature. As we mentioned previously, capturing temporal dynamics and expressing confidence for continuous estimates are highly limited in the currently available frameworks.

What will achieving the goal look like?

Achieving this goal would mean that researchers interested in visual metacognition are no longer limited by the task they use. Nowadays, a lot of interesting research that is done in the field of visual metacognition falls outside the scope of existing models, especially if the experiment does not consider a simple two-choice task. Concrete progress would be having identified one or more robust paradigms for decision scenarios beyond the standard two-choice version (e.g., n-alternative choice, continuous estimates, ongoing perception, or perceptuomotor interactions), with one or more accompanying computational models of metacognition. Ideally, these computational models would be more general and adaptable to different decision scenarios, including the standard two-choice tasks favored today. This may be achieved by generalizing existing metacognitive models (i.e., SDT or accumulation-to-bound models) or with other decision-making frameworks (e.g., Bayesian frameworks).

Medium-term goal 2: Determine the computations underlying confidence and what factors influence these computations

Why is this goal important and how does it relate to the long-term goals above?

This goal is a combination of three separate goals (see Supplementary) that were similar enough to warrant combining them. The goal, therefore, has three separate components, which are to understand: (1) what are the computations underlying confidence, (2) how do different sources of uncertainty influence metacognitive processes (regardless of whether the first-order decision is affected too), and (3) what processes (if any) selectively affect confidence while leaving the first-order decision unperturbed. The three components are interrelated such that progress on one of them is likely to translate into progress on the rest. Overall, the goal here is to understand the computations behind confidence, especially via the effects of experimental manipulations. As such, this goal will advance long-term goals 1-3 (related to developing models, developing manipulations, and determining confidence computations for complex tasks). While less directly related, progress on this goal may also have implications for long-term goal 4 (uncovering the relationship between metacognition and consciousness). This goal is therefore central to the field of visual metacognition and is likely to have wide-ranging implications.

The work ahead

There are several aspects of this goal where substantial progress can be made in the next five years. We discuss what we perceive as the most important directions related to understanding confidence computation and identifying the factors that influence this computation.

<u>What are the computations underlying confidence?</u>

This question is often phrased as “What does confidence reflect?” There are several competing hypotheses in the field with relatively little agreement at present. One common view is the Bayesian notion that confidence reflects the posterior probability of being correct (Aitchison et al., 2015; Fleming & Daw, 2017; Meyniel et al., 2015; Pouget et al., 2016). In other words, people base their confidence ratings on the probability that their response is correct even if this computation is noisy or biased. Another common view grounded in signal detection theory and accumulation-to-bound models is that confidence directly reflects signal strength (Bang et al., 2019; Green & Swets, 1966; Maniscalco & Lau, 2016). Here, confidence is derived from an abstract evidence axis without computing the probability that a response would be correct. Other alternatives include the view that confidence reflects the evidence for the chosen option while ignoring the evidence for all unchosen alternatives (Koizumi et al., 2015; Maniscalco et al., 2016; Peters et al., 2017; Samaha et al., 2016; Zylberberg et al., 2012) or that it reflects the difference in posterior probability of the two most likely alternatives (H.-H. Li & Ma, 2020). Several papers have compared directly two or more of these alternatives (Adler & Ma, 2018; Aitchison et al., 2015; H.-H. Li & Ma, 2020) but a consensus is yet to emerge. We believe that substantial progress is possible in the next five years on distinguishing between these possibilities.

<u>What factors influence the confidence computation and how?</u>

There is vibrant literature on the factors that influence confidence computation (reviewed in Shekhar & Rahnev, 2021a). Here we briefly mention the factors that have received the greatest attention and then discuss what we perceive as the most promising next steps.

Perhaps the most widely studied factors that affect confidence computations are stimulus variability and attention. However, the exact effects of each of these factors remain controversial. For example, increased variability has been found to lead both to higher-than-

expected and lower-than-expected confidence (Bertana et al., 2021; Boldt et al., 2017; de Gardelle & Mamassian, 2015; Spence et al., 2016, 2018; Zylberberg et al., 2014, 2016). Similarly, different manipulations of attention have been found to either increase or decrease confidence and visibility ratings (Denison et al., 2018; Kurtz et al., 2017; Rahnev et al., 2011, 2012; Recht et al., 2019; Wilimzig et al., 2008; Zizlsperger et al., 2012). These studies have used different designs, manipulations, and sometimes collected different metacognitive measurements (e.g., confidence vs. visibility), making it difficult to pinpoint the reasons for the divergent results. Many other factors have been investigated by relatively fewer studies. For example, confidence is influenced by the confidence on previous trials (Aguilar-Lleyda et al., 2021; Rahnev et al., 2015), motor preparation and execution (Fleming et al., 2015; Gajdos, Fleming, et al., 2019), visual field location (M. K. Li et al., 2018; Solovey et al., 2015), the strength of decision-congruent evidence (Koizumi et al., 2015; Maniscalco et al., 2016; Peters et al., 2017; Samaha et al., 2016; Zylberberg et al., 2012), stimulus visibility (Rausch et al., 2018), and decision time (Kiani et al., 2014).

Despite the large number of factors already identified, many other factors that affect the confidence computation are likely yet to be discovered. A mechanistic understanding of confidence would strongly benefit (and perhaps require) the identification of all critical factors, and therefore the search should continue. The next five years can be expected to add more to the list above. Nevertheless, it also appears that the field has reached a point where more emphasis needs to be given on firmly establishing the knowledge that (we think) we have already gained. For example, few of the studies cited above have been independently replicated and there has not been much consideration of the effect sizes for each of the factors influencing confidence. Therefore, in the next five years, more attention should be paid to replicating existing effects and clarifying the effect size of each.

What will achieving the goal look like?

It is not reasonable to think that five years from now we will know the precise computations underlying confidence and all the ways it is influenced. However, it is reasonable to expect a growing emphasis on empirically adjudicating between different proposals of what confidence reflects, perhaps with an emerging consensus at least for simple two-choice experimental designs. Similarly, it is reasonable to expect the emergence of high-powered replication attempts of the different factors that influence confidence. We will consider the goal "achieved" if both of these expectations are met or at least measurable progress has been made. Such progress will have a large effect as it will ensure that the field is on a sure footing and well-positioned to build cumulative knowledge.

Final thoughts and next steps

Having described the four long-term and two medium-term consensus goals, we end with a short section where we discuss what we learned, as well as our plans for tracking and assessing progress towards achieving the goals listed here.

<u>Thoughts on the process and results of goal setting</u>

One of the greatest difficulties we encountered was with formulating clear and precise evaluation criteria for each goal. Indeed, currently, there is substantial latitude left for each goal. Naturally, given the generality of the long-term goals, deciding on evaluation criteria for each has been particularly challenging, though we have tried hard to establish specific evaluation criteria for the two medium-term goals. We think that difficulties with establishing concrete and rigorous evaluation criteria are likely unavoidable, especially for a relatively new field such as visual metacognition. Time will tell whether goal setting in psychological science is worthwhile

only in the context of precise landmarks and evaluation criteria, or if it can have value even if such landmarks and evaluation criteria are less well defined.

Notably, the issues of replicability, estimation of effect sizes, and the use of appropriate sample sizes were only explicitly discussed in medium-term goal 2. This perhaps reflects a perception among the authors that replicability of findings in the field is likely to be relatively high, though there have been relatively few replication studies thus far to formally test this impression. Nevertheless, given the ongoing replication crisis in psychology and related disciplines (Open Science Collaboration, 2015), it may be important to pay more attention to these issues going forward.

Finally, it should be noted that we did not discuss "truths" in the field. In other words, we did not discuss which previous findings within the field are established beyond reasonable doubt and which are not. Such efforts are likely to be fruitful (e.g., see the paper on "benchmarks" in working memory by Oberauer et al., 2018) and may also be worth undertaking.

Tracking and assessing progress

We expect that formalizing these consensus goals will catalyze progress in the field, foster collaboration, and increase the chance of solving the most important problems in the field. Nevertheless, we recognize that formalizing these goals may have a limited influence without a system for tracking and assessing the progress made. It has been argued that progress in science is achieved only when a community of scientists is willing and able to hold each other accountable for the quality of their work (Ravetz, 1971). At the same time, any formal system of evaluation of individual papers or findings is likely to be inflexible and runs the risk of simply reflecting the opinions of authority figures. Any system of tracking and assessing progress should not be overly onerous (i.e., should not require an exorbitant amount of time and resources to maintain), or else it will likely be quickly abandoned.

Based on these considerations, we have decided to institute several mechanisms to help us track and assess progress towards the long- and medium-term goals that we set. First, we have created a Slack channel intended to allow for informal conversations on issues related to each goal. We invite everyone who has an interest in any of these goals to subscribe and actively participate in the ongoing discussions (link to join: bit.ly/3wsPoyl). Second, papers relevant to each long- and medium-term goal will be tracked using an online community-powered spreadsheet (bit.ly/3CJvmCA). We encourage everyone publishing relevant papers to add their papers to this spreadsheet. To obtain help with either the Slack channel or the spreadsheet, one can email visual.meta.goals@gmail.com. Third, we plan to organize a regular meeting or conference specifically for the field of visual metacognition. Fourth, we intend to write a follow-up paper in approximately five years that will assess progress towards both the long- and medium-term goals. Finally, we encourage new papers to explicitly state which of these long- and medium-term goals their findings are relevant to. This practice would be especially important for null results. Such explicit references will make future reviews and meta-analyses on the topics related to these goals substantially easier and more accurate.

Conclusion

Scientific progress requires the accumulation of agreed-upon empirical knowledge and robust theories. We believe that common goals can accelerate such progress by ensuring both a reliable body of empirical findings and the development of theories that explain existing data and make new predictions. Here 26 researchers from the field of visual metacognition agreed on such consensus goals. We identified four long-term and two medium-term goals, as well as a

process for tracking and assessing progress. Only time will tell how this effort will impact our field. We hope that the formulation of these goals will enable researchers from across the field to focus our energies, increase the quality of our research, ensure that we build solid cumulative knowledge in our field, and foster more collaboration. At the very least, it should be a useful experiment that provides insight into the forces that drive science and can stir it into states of higher or lower impact. If this effort proves successful, consensus goal setting can become a model for many fields of psychological science and beyond.

References

Adler, W. T., & Ma, W. J. (2018). Comparing Bayesian and non-Bayesian accounts of human confidence reports. *PLOS Computational Biology*, *14*(11), e1006572. https://doi.org/10.1371/journal.pcbi.1006572

Aguilar-Lleyda, D., Konishi, M., Sackur, J., & de Gardelle, V. (2021). Confidence can be automatically integrated across two visual decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(2), 161–171. https://doi.org/10.1037/xhp0000884

Aguilar-Lleyda, D., Lemarchand, M., & de Gardelle, V. (2020). Confidence as a Priority Signal. *Psychological Science*, *31*(9), 1084–1096. https://doi.org/10.1177/0956797620925039

Aitchison, L., Bang, D., Bahrami, B., & Latham, P. E. (2015). Doubly Bayesian Analysis of Confidence in Perceptual Decision-Making. *PLoS Computational Biology*, *11*(10), 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

Baars, B. (1997). Contrastive phenomenology: A thoroughly empirical approach to consciousness. In N. Block, O. Flanagan, & G. Güzeldere (Eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 187–202). MIT Press. https://philpapers.org/rec/BAACPA

Baird, B., Mrazek, M. D., Phillips, D. T., & Schooler, J. W. (2014). Domain-specific enhancement of metacognitive ability following meditation training. *Journal of Experimental Psychology. General*, *143*(5), 1972–1979. https://doi.org/10.1037/a0036882

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1753), 1–11. https://doi.org/10.1038/s41467-020-15561-w

Bang, J. W., Shekhar, M., & Rahnev, D. (2019). Sensory noise increases metacognitive efficiency. *Journal of Experimental Psychology: General*, *148*(3), 437–452. https://doi.org/10.1037/xge0000511

Bertana, A., Chetverikov, A., van Bergen, R. S., Ling, S., & Jehee, J. F. M. (2021). Dual strategies in human confidence judgments. *Journal of Vision*, *21*(5), 21. https://doi.org/10.1167/jov.21.5.21

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*(5–6), 481–499. https://doi.org/10.1017/S0140525X07002786

Block, N. (2019). What Is Wrong with the No-Report Paradigm and How to Fix It. *Trends in Cognitive Sciences*, *23*(12), 1003–1013. https://doi.org/10.1016/J.TICS.2019.10.001

Boldt, A., de Gardelle, V., & Yeung, N. (2017). The impact of evidence reliability on sensitivity and bias in decision confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(8), 1520–1531. https://doi.org/10.1037/xhp0000404

Boldt, A., Schiffer, A.-M., Waszak, F., & Yeung, N. (2019). Confidence Predictions Affect Performance Confidence and Neural Preparation in Perceptual Decision Making. *Scientific Reports*, *9*(1), 4031. https://doi.org/10.1038/s41598-019-40681-9

Bonnen, K., Burge, J., Yates, J., Pillow, J., & Cormack, L. K. (2015). Continuous psychophysics: Target-tracking to measure visual sensitivity. *Journal of Vision*, *15*(3), 14–14. https://doi.org/10.1167/15.3.14

Bor, D., Schwartzman, D. J., Barrett, A. B., & Seth, A. K. (2017). Theta-burst transcranial magnetic stimulation to the prefrontal or parietal cortex does not impair metacognitive visual awareness. *PLOS ONE*, *12*(2), e0171793. https://doi.org/10.1371/journal.pone.0171793

Brown, R., Lau, H., & LeDoux, J. E. (2019). Understanding the Higher-Order Approach to

Consciousness. *Trends in Cognitive Sciences*, *23*(9), 754–768. https://doi.org/10.1016/J.TICS.2019.06.009
Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal of Experimental Psychology: General*, *148*(1), 51–64. https://doi.org/10.1037/xge0000505
Clarke, F. R., Birdsall, T. G., & Tanner, W. P. (1959). Two Types of ROC Curves and Definitions of Parameters. *The Journal of the Acoustical Society of America*, *31*(5), 629–630. https://doi.org/10.1121/1.1907764
Collins, A., & Koechlin, E. (2012). Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making. *PLoS Biology*, *10*(3), e1001293. https://doi.org/10.1371/journal.pbio.1001293
Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*, *7*, 13669. https://doi.org/10.1038/ncomms13669
de Gardelle, V., Faivre, N., Filevich, E., Reyes, G., Rouy, M., Sackur, J., & Vergnaud, J.-C. (2020). Role of feedback on metacognitive training. *PsychArchives*. https://doi.org/https://doi.org/10.23668/PSYCHARCHIVES.3138
de Gardelle, V., & Mamassian, P. (2014). Does Confidence Use a Common Currency Across Two Visual Tasks? *Psychological Science*, *25*(6), 1286–1288. https://doi.org/10.1177/0956797614528956
de Gardelle, V., & Mamassian, P. (2015). Weighting Mean and Variability during Confidence Judgments. *PloS One*, *10*(3), e0120870. https://doi.org/10.1371/journal.pone.0120870
Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
Denison, R. N., Adler, W. T., Carrasco, M., & Ma, W. J. (2018). Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence. *Proceedings of the National Academy of Sciences*, *115*(43), 11090–11095. https://doi.org/10.1073/pnas.1717720115
Denison, R. N., Block, N., & Samaha, J. (2020). What do models of visual perception tell us about visual phenomenology? In F. De Brigard & W. Sinnot-Armstrong (Eds.), *Neuroscience and Philosophy*. MIT Press.
Deroy, O., Spence, C., & Noppeney, U. (2016). Metacognition in Multisensory Perception. *Trends in Cognitive Sciences*, *20*(10), 736–747. https://doi.org/10.1016/j.tics.2016.08.006
Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771
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
Faivre, N., Filevich, E., Solovey, G., Kühn, S., & Blanke, O. (2018). Behavioral, Modeling, and Electrophysiological Evidence for Supramodality in Human Metacognition. *The Journal of Neuroscience*, *38*(2), 263–277. https://doi.org/10.1523/JNEUROSCI.0322-17.2017
Fechner, G. T. (1860). *Elemente der Psychophysik*. Breitkopf und Härtel.
Feller, I., & Stern, P. C. (2007). *A Strategy for Assessing Science: Behavioral and Social Research on Aging*. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK26378/
Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(42), 17510–17515. https://doi.org/10.1073/pnas.1106439108
Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision performance: A general

Bayesian framework for metacognitive computation. *Psychological Review*, *124*(1), 91–114. https://doi.org/10.1037/rev0000045
Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: computation, biology and function. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1280–1286. https://doi.org/10.1098/rstb.2012.0021
Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-Specific Disruption of Perceptual Confidence. *Psychological Science*, *26*(1), 89–98. https://doi.org/10.1177/0956797614557697
Gajdos, T., Fleming, S., Garcia, M. S., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of Consciousness*, *5*(1), niz001. https://doi.org/10.1101/330605
Gajdos, T., Régner, I., Huguet, P., Hainguerlot, M., Vergnaud, J.-C., Sackur, J., & de Gardelle, V. (2019). Does social context impact metacognition? Evidence from stereotype threat in a visual search task. *PLOS ONE*, *14*(4), e0215050. https://doi.org/10.1371/journal.pone.0215050
Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876. https://doi.org/10.3758/BF03196546
Graf, E. W., Warren, P. A., & Maloney, L. T. (2005). Explicit estimation of visual uncertainty in human motion processing. *Vision Research*, *45*(24), 3050–3059. https://doi.org/10.1016/J.VISRES.2005.08.007
Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley & Sons Ltd.
Guest, O., & Martin, A. E. (2021). How Computational Modeling Can Force Theory Building in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 789–802. https://doi.org/10.1177/1745691620970585
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
Haddara, N., & Rahnev, D. (2021). The impact of feedback on perceptual decision making and metacognition: Reduction in bias but no change in sensitivity. *Psychological Science*.
Hainguerlot, M., Vergnaud, J.-C., & de Gardelle, V. (2018). Metacognitive ability predicts learning cue-stimulus associations in the absence of external feedback. *Scientific Reports*, *8*(1), 5602. https://doi.org/10.1038/s41598-018-23936-9
Helmholtz, H. L. F. (1856). *Treatise on physiological optics*. Thoemmes Continuum.
Herding, J., Ludwig, S., von Lautz, A., Spitzer, B., & Blankenburg, F. (2019). Centro-parietal EEG potentials index subjective evidence and confidence during perceptual decision making. *NeuroImage*, *201*, 116011. https://doi.org/10.1016/j.neuroimage.2019.116011
Huk, A., Bonnen, K., & He, B. J. (2018). Beyond Trial-Based Paradigms: Continuous Behavior, Ongoing Neural Activity, and Natural Stimuli. *Journal of Neuroscience*, *38*(35), 7551–7558. https://doi.org/10.1523/JNEUROSCI.1920-17.2018
Jachs, B., Blanco, M. J., Grantham-Hill, S., & Soto, D. (2015). On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 269. https://doi.org/10.1037/xhp0000026
Kanai, R., Walsh, V., & Tseng, C. H. (2010). Subjective discriminability of invisibility: A framework for distinguishing perceptual and attentional failures of awareness. *Consciousness and Cognition*, *19*(4), 1045–1057. https://doi.org/10.1016/j.concog.2010.06.003
Kellij, S., Fahrenfort, J., Lau, H., Peters, M. A. K., & Odegaard, B. (2021). An investigation of how relative precision of target encoding influences metacognitive performance. *Attention,*

Perception, & Psychophysics, *83*(1), 512–524. https://doi.org/10.3758/s13414-020-02190-0
Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1322–1337. https://doi.org/10.1098/rstb.2012.0037
Kepecs, A., Uchida, N., Zariwala, H. a, & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. https://doi.org/10.1038/nature07200
Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. *Neuron*, *84*(6), 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*(5928), 759–764. https://doi.org/10.1126/science.1169405
Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225
Koizumi, A., Maniscalco, B., & Lau, H. (2015). Does perceptual confidence facilitate cognitive control? *Attention, Perception, & Psychophysics*, *77*(4), 1295–1306. https://doi.org/10.3758/s13414-015-0843-3
Konishi, M., Compain, C., Berberian, B., Sackur, J., & de Gardelle, V. (2020). Resilience of perceptual metacognition in a dual-task paradigm. *Psychonomic Bulletin & Review*, *27*(6), 1–10. https://doi.org/10.3758/s13423-020-01779-8
Kording, K., Blohm, G., Schrater, P., & Kay, K. (2018). Appreciating diversity of goals in computational neuroscience. *PsyArXivArXiv*. https://doi.org/10.31219/OSF.IO/3VY69
Kurtz, P., Shapcott, K. A., Kaiser, J., Schmiedt, J. T., & Schmid, M. C. (2017). The Influence of Endogenous and Exogenous Spatial Attention on Decision Confidence. *Scientific Reports*, *7*(1), 6431. https://doi.org/10.1038/s41598-017-06715-w
Lamme, V. A. F. (2000). Neural Mechanisms of Visual Awareness: A Linking Proposition. *Brain and Mind*, *1*(3), 385–406. https://doi.org/10.1023/A:1011569019782
Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences*, *15*(8), 365–373. https://doi.org/10.1016/j.tics.2011.05.009
Lee, A. L. F., de Gardelle, V., & Mamassian, P. (2021). Global visual confidence. *Psychonomic Bulletin & Review*, 1–10. https://doi.org/10.3758/s13423-020-01869-7
Lempert, K. M., Chen, Y. L., & Fleming, S. M. (2015). Relating Pupil Dilation and Metacognitive Confidence during Auditory Decision-Making. *PloS One*, *10*(5), e0126588. https://doi.org/10.1371/journal.pone.0126588
Li, H.-H., & Ma, W. J. (2020). Confidence reports in decision-making with multiple alternatives violate the Bayesian confidence hypothesis. *Nature Communications*, *11*(1), 2004. https://doi.org/10.1038/s41467-020-15581-6
Li, M. K., Lau, H., & Odegaard, B. (2018). An investigation of detection biases in the unattended periphery during simulated driving. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-018-1554-3
Lisi, M., Mongillo, G., Milne, G., Dekker, T., & Gorea, A. (2020). Discrete confidence levels revealed by sequential decisions. *Nature Human Behaviour*, 1–8. https://doi.org/10.1038/s41562-020-00953-1
Locke, S. M., Mamassian, P., & Landy, M. S. (2020). Performance monitoring for sensorimotor confidence: A visuomotor tracking study. *Cognition*, *205*, 104396.

https://doi.org/10.1016/j.cognition.2020.104396
Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, *2*(1), annurev-vision-111815-114630. https://doi.org/10.1146/annurev-vision-111815-114630
Mamassian, P. (2020). Confidence Forced-Choice and Other Metaperceptual Tasks. *Perception*, *49*(6), 616–635. https://doi.org/10.1177/0301006620928010
Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021
Maniscalco, B., & Lau, H. (2015). Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task. *Neuroscience of Consciousness*, *2015*(1), niv002. https://doi.org/10.1093/nc/niv002
Maniscalco, B., & Lau, H. (2016). The signal processing architecture underlying subjective reports of sensory awareness. *Neuroscience of Consciousness*, *2016*(1), 1–17. https://doi.org/10.1093/nc/niw002
Maniscalco, B., Peters, M. A. K., & Lau, H. (2016). Heuristic use of perceptual evidence leads to dissociation between performance and metacognitive sensitivity. *Attention, Perception & Psychophysics*, *78*(3), 923–937. https://doi.org/10.3758/s13414-016-1059-x
Masset, P., Ott, T., Lak, A., Hirokawa, J., & Kepecs, A. (2020). Behavior- and Modality-General Representation of Confidence in Orbitofrontal Cortex. *Cell*, *182*(1), 112-126.e18. https://doi.org/10.1016/j.cell.2020.05.022
Mazor, M., Friston, K. J., & Fleming, S. M. (2020). Distinct neural contributions to metacognition for detecting, but not discriminating visual stimuli. *ELife*, *9*, 853366. https://doi.org/10.7554/eLife.53900
Mazor, M., Moran, R., & Fleming, S. (2021). Stage 2 Registered Report: Metacognitive asymmetries in visual perception. *PsyArXiv*. https://doi.org/10.31234/OSF.IO/AV9NS
Mei, N., Rankine, S., Olafsson, E., & Soto, D. (2020). Similar history biases for distinct prospective decisions of self-performance. *Scientific Reports*, *10*(1), 5854. https://doi.org/10.1038/s41598-020-62719-z
Melloni, L., Mudrik, L., Pitts, M., & Koch, C. (2021). Making the hard problem of consciousness easier. *Science*, *372*(6545), 911–912. https://doi.org/10.1126/science.abj3259
Meuwese, J. D. I., van Loon, A. M., Lamme, V. A. F., & Fahrenfort, J. J. (2014). The subjective experience of object recognition: comparing metacognition for object detection and object categorization. *Attention, Perception, & Psychophysics*, *76*, 1057–1068. https://doi.org/10.3758/s13414-014-0643-1
Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*, *88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039
Miyoshi, K., & Lau, H. (2020). A decision-congruent heuristic gives superior metacognitive sensitivity under realistic variance assumptions. *Psychological Review*. https://doi.org/10.1037/rev0000184
Morales, J., Odegaard, B., & Maniscalco, B. (2019). The Neural Substrates of Conscious Perception without Performance Confounds. *PsyArXiv*, 1–29. https://doi.org/10.31234/osf.io/8zhy3
Moritz, S., & Woodward, T. S. (2007). Metacognitive training in schizophrenia: from basic research to knowledge translation and intervention. *Current Opinion in Psychiatry*, *20*(6), 619–625. https://doi.org/10.1097/YCO.0b013e3282f0b8ed
Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1
Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 125–141). Academic Press.

Norman, E., & Price, M. C. (2015). Measuring consciousness with confidence ratings. In M. Overgaard (Ed.), *Behavioural methods in consciousness research* (pp. 159–180). Oxford University Press.

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Peirce, C. S., & Jastrow, J. (1884). On Small Differences in Sensation. *Memoirs of the National Academy of Sciences*, *3*, 75–83.

Pereira, M., Megevand, P., Tan, M. X., Chang, W., Wang, S., Rezai, A., Seeck, M., Corniola, M., Momjian, S., Bernasconi, F., Blanke, O., & Faivre, N. (2021). Evidence accumulation relates to perceptual consciousness and monitoring. *Nature Communications*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-23540-y

Pescetelli, N., & Yeung, N. (2021). The role of decision confidence in advice-taking and trust formation. *Journal of Experimental Psychology: General*, *150*(3), 507–526. https://doi.org/10.1037/xge0000960

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. *Nature Human Behaviour*, *1*(7), 0139. https://doi.org/10.1038/s41562-017-0139

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. https://doi.org/10.1037/a0019737

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

Purcell, B. A., & Kiani, R. (2016). Neural Mechanisms of Post-error Adjustments of Decision Policy in Parietal Cortex. *Neuron*, *89*(3), 658–671. https://doi.org/10.1016/j.neuron.2015.12.027

Rademaker, R. L., & Pearson, J. (2012). Training Visual Imagery: Improvements of Metacognition, but not Imagery Strength. *Frontiers in Psychology*, *3*, 224. https://doi.org/10.3389/fpsyg.2012.00224

Rahnev, D. (2020). Confidence in the Real World. *Trends in Cognitive Sciences*, *24*(8), 590–591. https://doi.org/10.1016/J.TICS.2020.05.005

Rahnev, D. (2021). Visual metacognition: Measures, models and neural correlates. *American Psychologist*.

Rahnev, D., Bahdo, L., de Lange, F. P., & Lau, H. (2012). Prestimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *Journal of Neurophysiology*, *108*(5), 1529–1536. https://doi.org/10.1152/jn.00184.2012

Rahnev, D., & Denison, R. N. (2018). Suboptimality in Perceptual Decision Making. *Behavioral and Brain Sciences*, *41*(e223), 1–66. https://doi.org/10.1017/S0140525X18000936

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., … Zylberberg, A. (2020). The Confidence Database. *Nature Human Behaviour*, *4*(3), 317–325. https://doi.org/10.1038/s41562-019-0813-1

Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence Leak in Perceptual Decision Making. *Psychological Science*, *26*(11), 1664–1680.

https://doi.org/10.1177/0956797615595037
Rahnev, D., Maniscalco, B., Graves, T., Huang, E., De Lange, F. P., & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, *14*(12), 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. *Proceedings of the National Academy of Sciences*, *113*(20), 6059–6064. https://doi.org/10.1073/pnas.1522551113
Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychological Review*, *120*(3), 697–719. https://doi.org/10.1037/a0033152
Rausch, M., Hellmann, S., & Zehetleitner, M. (2018). Confidence in masked orientation judgments is informed by both evidence and visibility. *Attention, Perception, and Psychophysics*, *80*, 134–154. https://doi.org/10.3758/s13414-017-1431-5
Ravetz, J. (1971). *Scientific Knowledge and Its Social Problems*. Oxford University Press.
Recht, S., Mamassian, P., & de Gardelle, V. (2019). Temporal attention causes systematic biases in visual confidence. *Scientific Reports*, *9*(1), 11622. https://doi.org/10.1038/s41598-019-48063-x
Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275
Reyes, G., Silva, J. R., Jaramillo, K., Rehbein, L., & Sackur, J. (2015). Self-Knowledge Dim-Out: Stress Impairs Metacognitive Accuracy. *PLOS ONE*, *10*(8), e0132320. https://doi.org/10.1371/journal.pone.0132320
Reyes, G., Vivanco-Carlevari, A., Medina, F., Manosalva, C., de Gardelle, V., Sackur, J., & Silva, J. R. (2020). Hydrocortisone decreases metacognitive efficiency independent of perceived stress. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-71061-3
Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020). Confidence drives a neural confirmation bias. *Nature Communications*, *11*(1), 2634. https://doi.org/10.1038/s41467-020-16278-6
Rosenthal, C. R. R., Andrews, S. K. K., Antoniades, C. A. A., Kennard, C., & Soto, D. (2016). Learning and recognition of a non-conscious sequence of events in human primary visual cortex. *Current Biology*, *26*(6), 834–841. https://doi.org/10.1016/j.cub.2016.01.040
Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), 1141. https://doi.org/10.1038/s41467-019-09075-3
Rouault, M., Weiss, A., Lee, J. K., Bouté, J., Drugowitsch, J., Chambon, V., & Wyart, V. (2021). Specific cognitive signatures of information seeking in controllable environments. *BioRxiv*. https://doi.org/10.1101/2021.01.04.425114
Rounis, E., Maniscalco, B., Rothwell, J. C., Passingham, R. E., & Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cognitive Neuroscience*, *1*(3), 165–175. https://doi.org/10.1080/17588921003632529
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. *Cerebral Cortex*, *26*(3), 1200–1210. https://doi.org/10.1093/cercor/bhu311
Samaha, J., Barrett, J. J., Sheldon, A. D., LaRocque, J. J., & Postle, B. R. (2016). Dissociating Perceptual Confidence from Discrimination Accuracy Reveals No Influence of Metacognitive Awareness on Working Memory. *Frontiers in Psychology*, *7*, 851. https://doi.org/10.3389/fpsyg.2016.00851
Samaha, J., Switzky, M., & Postle, B. R. (2019). Confidence boosts serial dependence in orientation estimation. *Journal of Vision*, *19*(4), 25. https://doi.org/10.1167/19.4.25
Sandberg, K., Bibby, B. M., Timmermans, B., Cleeremans, A., & Overgaard, M. (2011).

Measuring consciousness: task accuracy and awareness as sigmoid functions of stimulus duration. *Consciousness and Cognition*, *20*(4), 1659–1675. https://doi.org/10.1016/j.concog.2011.09.002
Sarafyazd, M., & Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. *Science*, *364*(6441), eaav8911. https://doi.org/10.1126/science.aav8911
Schmidt, C., Reyes, G., Barrientos, M., Langer, Á. I., & Sackur, J. (2019). Meditation focused on self-observation of the body impairs metacognitive efficiency. *Consciousness and Cognition*. https://doi.org/10.1016/j.concog.2019.03.001
Scott, R. B., Dienes, Z., Barrett, A. B., Bor, D., & Seth, A. K. (2014). Blind Insight : Metacognitive Discrimination Despite Chance Task Performance. *Psychological Science*, *25*(12), 2199–2208. https://doi.org/10.1177/0956797614553944
Sergent, C., Corazzol, M., Labouret, G., Stockart, F., Wexler, M., King, J.-R., Meyniel, F., & Pressnitzer, D. (2021). Bifurcation in brain dynamics reveals a signature of conscious processing independent of report. *Nature Communications 2021 12:1*, *12*(1), 1–19. https://doi.org/10.1038/s41467-021-21393-z
Seth, A. K., Dienes, Z., Cleeremans, A., Overgaard, M., & Pessoa, L. (2008). Measuring consciousness: relating behavioural and neurophysiological approaches. *Trends in Cognitive Sciences*, *12*(8), 314–321. https://doi.org/10.1016/j.tics.2008.04.008
Shea, N., & Frith, C. D. (2019). The Global Workspace Needs Metacognition. *Trends in Cognitive Sciences*, *23*(7), 560–571. https://doi.org/10.1016/j.tics.2019.04.007
Shekhar, M., & Rahnev, D. (2018). Distinguishing the roles of dorsolateral and anterior PFC in visual metacognition. *Journal of Neuroscience*, *38*(22), 5078–5087. https://doi.org/10.1523/JNEUROSCI.3484-17.2018
Shekhar, M., & Rahnev, D. (2021a). Sources of Metacognitive Inefficiency. *Trends in Cognitive Sciences*, *25*(1), 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. *Psychological Review*, *128*(1), 45–70. https://doi.org/10.1037/rev0000249
Sherman, M. T., Seth, A. K., Barrett, A. B., & Kanai, R. (2015). Prior expectations facilitate metacognition for perceptual decision. *Consciousness and Cognition*, *35*, 53–65. https://doi.org/10.1016/j.concog.2015.04.015
Siedlecka, M., Paulewicz, B., & Wierzchoń, M. (2016). But I Was So Sure! Metacognitive Judgments Are Less Accurate Given Prospectively than Retrospectively. *Frontiers in Psychology*, *7*, 218. https://doi.org/10.3389/fpsyg.2016.00218
Solovey, G., Graney, G. G., & Lau, H. (2015). A decisional account of subjective inflation of visual perception at the periphery. *Attention, Perception & Psychophysics*, *77*(1), 258–271. https://doi.org/10.3758/s13414-014-0769-1
Spence, M. L., Dux, P. E., & Arnold, D. H. (2016). Computations Underlying Confidence in Visual Perception. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(5), 671–682. https://doi.org/10.1037/xhp0000179
Spence, M. L., Mattingley, J. B., & Dux, P. E. (2018). Uncertainty information that is irrelevant for report impacts confidence judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(12), 1981–1994. https://doi.org/10.1037/xhp0000584
Swets, J. A., Tanner, W. P., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*(5), 301–340. http://www.ncbi.nlm.nih.gov/pubmed/13774292
Timmermans, B., & Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. In *Behavioral Methods in Consciousness Research* (pp. 21–46). Oxford Scholoarship Online. https://doi.org/10.1093/acprof:oso/9780199688890.003.0003
Trommershäuser, J., Kording, K., & Landy, M. S. (2011). *Sensory Cue Integration* (J. Trommershäuser, K. P. Körding, & M. S. Landy (eds.)). Oxford University Press.
Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. F. (2015). No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences*,

19(12), 757–770. https://doi.org/10.1016/j.tics.2015.10.002
Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 14637. https://doi.org/10.1038/ncomms14637
van den Berg, R., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). Confidence Is the Bridge between Multi-stage Decisions. *Current Biology*, *26*(23), 3157–3168. https://doi.org/10.1016/j.cub.2016.10.021
van Rooij, I., & Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, *16*(4), 682–697. https://doi.org/10.1177/1745691620970604
Vandenbroucke, A. R. E., Fahrenfort, J. J., Sligte, I. G., & Lamme, V. A. F. (2014). Seeing without knowing: neural signatures of perceptual inference in the absence of report. *Journal of Cognitive Neuroscience*, *26*(5), 955–969. https://doi.org/10.1162/jocn_a_00530
Vangen, S., & Huxham, C. (2012). The Tangled Web: Unraveling the Principle of Common Goals in Collaborations. *Journal of Public Administration Research and Theory*, *22*(4), 731–760. https://doi.org/10.1093/JOPART/MUR065
Vickers, D. (1979). *Decision Processes in Visual Perception*. Academic Press.
Wilimzig, C., Tsuchiya, N., Fahle, M., Einhäuser, W., & Koch, C. (2008). Spatial attention increases performance but not subjective confidence in a discrimination task. *Journal of Vision*, *8*(5), 1–10. https://doi.org/10.1167/8.5.7
Yallak, E., & Balcı, F. (2021). Metric error monitoring: Another generalized mechanism for magnitude representations? *Cognition*, *210*(June 2020), 104532. https://doi.org/10.1016/j.cognition.2020.104532
Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393
Yaron, I., Melloni, L., Pitts, M., & Mudrik, L. (2021). The Consciousness Theories Studies (ConTraSt) database: analyzing and comparing empirical studies of consciousness theories. *BioRxiv*, 2021.06.10.447863. https://doi.org/10.1101/2021.06.10.447863
Yoo, A. H., Klyszejko, Z., Curtis, C. E., & Ma, W. J. (2018). Strategic allocation of working memory resource. *Scientific Reports 2018 8:1*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-34282-1
Zakrzewski, A. C., Wisniewski, M. G., Iyer, N., & Simpson, B. D. (2019). Confidence tracks sensory- and decision-related ERP dynamics during auditory detection. *Brain and Cognition*, *129*, 49–58. https://doi.org/10.1016/J.BANDC.2018.10.007
Zehetleitner, M., & Rausch, M. (2013). Being confident without seeing: What subjective measures of visual consciousness are about. *Attention, Perception, & Psychophysics*, *75*(7), 1406–1426. https://doi.org/10.3758/s13414-013-0505-2
Zhang, H., & Maloney, L. T. (2012). Ubiquitous Log Odds: A Common Representation of Probability and Frequency Distortion in Perception, Action, and Cognition. *Frontiers in Neuroscience*, *6*, 1. https://doi.org/10.3389/fnins.2012.00001
Zizlsperger, L., Sauvigny, T., & Haarmeier, T. (2012). Selective attention increases choice certainty in human decision making. *PloS One*, *7*(7), e41136. https://doi.org/10.1371/journal.pone.0041136
Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*(September), 79. https://doi.org/10.3389/fnint.2012.00079
Zylberberg, A., Fetsch, C. R., & Shadlen, M. N. (2016). The influence of evidence volatility on choice, reaction time and confidence in a perceptual decision. *ELife*, *5*(5), e17688. https://doi.org/10.7554/eLife.17688
Zylberberg, A., Roelfsema, P. R., & Sigman, M. (2014). Variance misperception explains

illusions of confidence in simple perceptual decisions. *Consciousness and Cognition*, *27*, 246–253. https://doi.org/10.1016/j.concog.2014.05.012

Zylberberg, A., Wolpert, D. M., & Shadlen, M. N. (2018). Counterfactual Reasoning Underlies the Learning of Priors in Decision Making. *Neuron*, *99*(5), 1083-1097.e6. https://doi.org/10.1016/j.neuron.2018.07.035

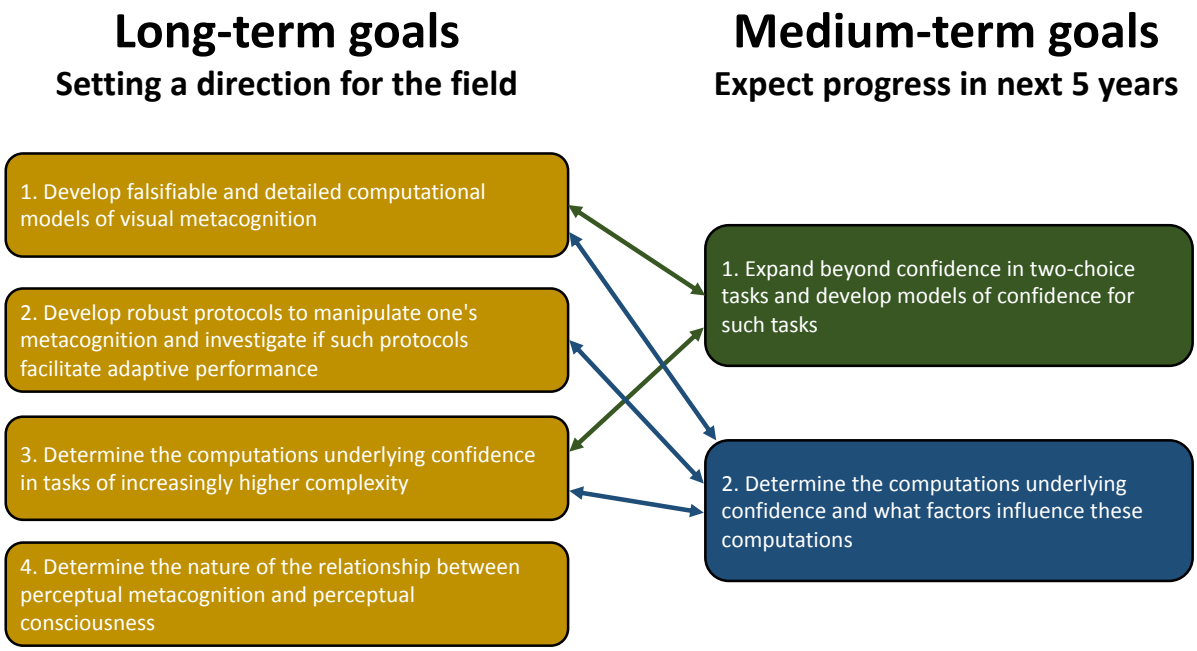

Figure 1. Consensus long- and medium-term goals. The arrows indicate how the four long-term goals are related to each of the two medium-term goals. Long-term goal 4 is the only long-term goal that is not directly connected to either of the medium-term goals, though progress on these medium-term goals could have implications for long-term goal 4 too. The arrows are bidirectional to highlight the facts that (1) progress on the medium-term goals automatically results in progress for the long-term goals, and (2) the broader long-term goals have critical subcomponents represented by the medium-term goals.