Trends in Cognitive Sciences

CellPress REVIEWS

Opinion

The Global Workspace Needs Metacognition

Nicholas Shea[1,2] and Chris D. Frith[1,3,*]

The two leading cognitive accounts of consciousness currently available concern global workspace (a form of working memory) and metacognition. There is relatively little interaction between these two approaches and it has even been suggested that the two accounts are rival and separable alternatives. Here, we argue that the successful function of a global workspace critically requires that the broadcast representations include a metacognitive component.

Highlights

The global workspace and metacognition are, respectively, the basis of the two leading cognitive theories of consciousness.

The two theories, which have recently been presented as rivals, are usually pursued separately, but there is no need to choose between them.

There is in fact strong reason to expect items in the global workspace to have a metacognitive accompaniment in the form of a rating of confidence.

Confidence ratings are relied on by the computations that compare, integrate, and compute over representations in the global workspace.

Recent empirical findings support the hypothesis that representations in the global workspace always carry with them a measure of confidence.

Two Rival Theories of Consciousness

While people generally think they know what consciousness is, it has proved difficult to operationalise. Many competing approaches exist. Here, we consider the two most prominent theories that use a cognitive framework. One theory, which evolved from the concept of working memory [1], proposes that, when mental representations are **conscious** (see Glossary), they are in a **global workspace** that is accessible to a variety of cognitive processes [2–4]. The other theory asserts that representations associated with consciousness always have a metacognitive component, such as a degree of **confidence** [5–9].

A common view is that the global workspace and metacognition are two distinctly different ways of making a cognitive system more sophisticated. Each is useful in its own right and they are dissociable: a representation can be **globally available** or **access conscious** without having any metacognitive component and *vice versa* [10]. As a consequence, global availability and metacognition are rarely studied together and there is disagreement about which is more important for understanding consciousness. By contrast, we argue that conscious representations are characterised by both global availability and metacognition. Specifically, we contend that representations in the global workspace always have a metacognitive component.

Distinguishing the Two Theories

The global workspace approach arises from the observation that consciousness appears to make information globally available to a range of widely different mental processes, such as reasoning, recollecting, planning, intention forming, and verbal report. The global workspace is postulated as the functional [2] and neural [3] basis of global availability and, thus, of consciousness. It is assumed that representations enter the workspace after processing has taken place in **domain-specific systems** (perception, emotion, motor control, etc., modulated by attention, which can reflect the goals of the agent). Thus, the global workspace is a form of working memory. Since it is controversial whether consciousness is required to merely maintain a representation in working memory, our focus here is the working memory system that allows representations to be manipulated (Box 1). Global broadcasting enables such manipulation, and that is thought to be an important function of the global workspace.

The link between **metacognition** and consciousness traces back to the longstanding intuition that, if an agent is unable to track or reflect on a particular mental state, then that state cannot be conscious. If this is the case, then some kind of metacognition would be associated with all representations that are conscious. Here, we are particularly concerned with metacognition as the confidence we have in a representation (although other **metacognitive parameters** may

[1]Institute of Philosophy, University of London, Senate House, Malet Street, London, WC1E 7HU, UK
[2]Faculty of Philosophy, University of Oxford, Oxford, OX2 6GG, UK
[3]Wellcome Centre for Human NeuroImaging, University College London, 12 Queen Square, London, WC1N 3BG, UK

*Correspondence:
c.frith@ucl.ac.uk (C.D. Frith).

© 2019 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

CellPress REVIEWS

Box 1. Manipulation of Items in Working Memory

It is controversial whether consciousness (in the sense of global availability) is required merely to maintain a representation in working memory (**WM maintenance**) but less controversial that consciousness is required for a representation to enter into that form of working memory that allows representations to be compared and manipulated (WM manipulation).

Mere maintenance in working memory appears to be a distinct process [91–93]. There is growing evidence that WM maintenance is possible in the absence of sustained neural activity [94]. WM maintenance may exist even for representations that have never been conscious and/or globally broadcast [95,96]. There is in fact little evidence that it is possible to maintain a representation in working memory on which action initiation, reasoning, or decision can be based without that representation having previously been made globally available ([97] *cf*. [98]), but our argument here does not rely on assuming that consciousness is required for a representation to enter WM maintenance.

By contrast, there is little evidence for WM manipulation of representations that have never been made globally available [99–102]. Accordingly, for the purpose of our argument, we assume only that global broadcast is needed for WM manipulation. An important function of global broadcasting is that it allows representations from different domain-specific systems to be put into contact with one another, in a workspace, such that they can be processed together. While nonconscious information can be integrated in the brain in a way that does not depend on working memory or attentional resources [103], the system for WM manipulation allows that to be achieved in a domain-general way, putting representations into contact with one another that have not previously been experienced together.

also have a role). According to theories that link consciousness to metacognition in this specific sense, a conscious representation of some aspect of the world (e.g., a **percept**) is accompanied by, or contains a sense of, certainty or uncertainty; that is, a sense of whether the representation is likely to be correct. Thus, a percept both represents that the world is a certain way, *p*, and at the same time represents itself as having a particular property (e.g., being likely to be correct). Importantly, this need not involve re-representing the content *p*. This is crucially different from the metacognitive account of consciousness put forward by **HOT theories of consciousness** [11].

Of course, cognitive systems can represent probabilities without being metacognitive. For example, when visual systems deploy a probabilistic population code for motion direction [12], what is represented is the probability that the observed motion is in direction *s* (as *s* ranges over 360° of possible motion directions). This does not involve a representation of confidence.

It does not follow from these definitions that global availability and metacognition must be tightly associated. There is no conceptual or necessary connection between them. Furthermore, although studies have shown that some types of metacognition do concern information that is globally available [13–15], other types of metacognition appear not to depend on consciousness [16–18]. Given these findings, we agree that not all metacognitive states are globally available. Our focus is instead on the dissociation in the other direction. Do all globally available representations have a metacognitive component? If true, that claim would be novel, substantive, and contrary to the common claim that the two dissociate [10,19–21].

Connection to Verbal Report

The two theories discussed here both preserve the strong connection that intuitively exists between consciousness and verbal report: an adult suffering no pathology in normal conditions can generally tell you about their conscious states. Both the global workspace and metacognition have a critical role in the ability to give verbal reports about the contents of consciousness ([22] p. 468; [23]).

However, enabling verbal reports is not the only function of the global workspace. Globally available representations are brought together in the workspace (*aka* working memory) so that cognitive work can be done with them [2,3,24].

Similarly, although metacognition does allow people to tell one another about their confidence [23], this is not its only function. Confidence, or uncertainty, is a metacognitive parameter

Glossary

Access conscious: a content is access conscious just in case a representation with that content is globally available.

Coarse-graining: a simplification of a system that is a more specific reflection of the microscale details. Effective coarse-graining enables modelling of the behaviour of a system without specifying all of the underlying causes that lead to system state changes.

Confidence: a representation of the probability that a representation is correct or, more broadly, of its reliability or appropriateness. Thus, a degree of certainty or uncertainty.

Conscious: part of a person's subjective experience, here understood as being a matter of global availability (i.e., access consciousness).

Domain-specific systems: perceptual, affective, and motoric systems processing information from a particular domain that are able to perform appropriately through having been exposed to a wealth of experience in the domain during evolution and/or learning (e.g., object perception, reaching and grasping, phonetic perception and production, and mentalizing).

Fluency: the speed and ease with which a representation is retrieved. This applies to perceptions, actions, and memories.

Globally available representation: information represented in a system and format such that it can be used, without undergoing further processing, by a wide range of cognitive processes (e.g., reasoning, planning, intention formation, mentalizing, verbal report, and storage in episodic and semantic memory).

Global workspace: a mechanism that forms the functional and neural basis for representations to be made globally available.

HOT theories of consciousness: claim that having a conscious experience with the content *p* depends on a further representation that one is in a first-order state with the content *p*. Consider the sentences 'John is uncertain' and 'John is sure that Tehran is the capital of Iran'. Both are at the metalevel, but the latter is richer in that it re-represents the content of John's belief. HOT theories claim that consciousness depends on having a

CellPress REVIEWS

associated with representations at all levels of the cognitive hierarchy and has a critical role in computations performed on these representations. For example, the automatic process through which different sources of perceptual evidence are combined (e.g., vision and touch) involves weighting by the relative **precision** of the different signals [25,26]. If conscious representations include a confidence parameter, then it would similarly allow them to be weighed and compared.

Our Hypothesis and a Simple Model

Our Hypothesis

Our hypothesis is that the cognitive work carried out in the workspace calls for metacognition. Metacognitive parameters, such as confidence, enable the computation of a common metric so that information from many different sources can be directly compared and combined [27]. Creation of a common metric is particularly necessary for the efficient functioning of the global workspace. The workspace contains representations broadcast from very different systems (perceptual, motoric, affective, and mnemonic). These need to be made available in a way that allows them to be brought together, so that the representations can be compared and combined to make decisions and reasoned over to plan for action ([28] p. 92). This is the function of estimates of confidence. For example, when people engage in conscious reasoning, they need to weigh items of information by their associated levels of confidence. Our claim is that effective performance of the function served by the global workspace therefore depends on metacognition. If so, we should expect representations in the global workspace always to include, or to be accompanied by, a metacognitive content (e.g., confidence).

> The Hypothesis
>
> Representations in the global workspace always have a metacognitive component

A Simple Model

Our proposal concerns the relationships between consciousness, globally broadcast representations, working memory, and metacognition. Our assumption is that consciousness, in the form of global broadcast, is the gateway to a form of working memory that allows representations to be compared and manipulated (**WM manipulation**; Box 1). An important function of global broadcasting is to make such manipulation possible.

We present our hypothesis in the context of a simple model where there are three types of process: inputs, workspace transitions, and outputs [29–31]: (i) Inputs: first, representations from selected domain-specific systems are broadcast to the workspace in a highly compressed form; (ii) transitions: second, these compressed representations are brought together and manipulated so that the agent can draw conclusions and make decisions; and (iii) outputs: third, these manipulated representations in turn drive actions, including verbal reports, via appropriate domain-specific systems (Figure 1). All three processes can be said to involve 'decisions', either taken by domain-specific mechanisms (e.g., the selection of what to broadcast is sometimes called a perceptual decision) or by the agent. We propose that metacognitive parameters, such as confidence, have a vital role in all three of these processes. Here, we are particularly concerned with the role of metacognition in the manipulation of representations in the workspace.

Functional Argument

Our proposal is that conscious (globally broadcast) representations need to carry with them a measure of confidence. Our argument is essentially functional. We have previously proposed that, when information is shared between people in the service of joint decision-making, it is advantageous if the information is associated with a degree of certainty [23]. In the case of suprapersonal cognitive control, confidence sharing can be used to enable a dyad to make near-optimal use of the information available [32].

thought that re-represents the content of the conscious state.

Metacognition: a representation or evaluation of another cognitive state or process. Confidence and its associates (our main focus here) are forms of metacognition.

Metacognitive parameter: a property of cognitive process or of the content of a mental representation. Metacognitive parameters include confidence (certainty/uncertainty), fluency, familiarity, and, in some circumstances, precision.

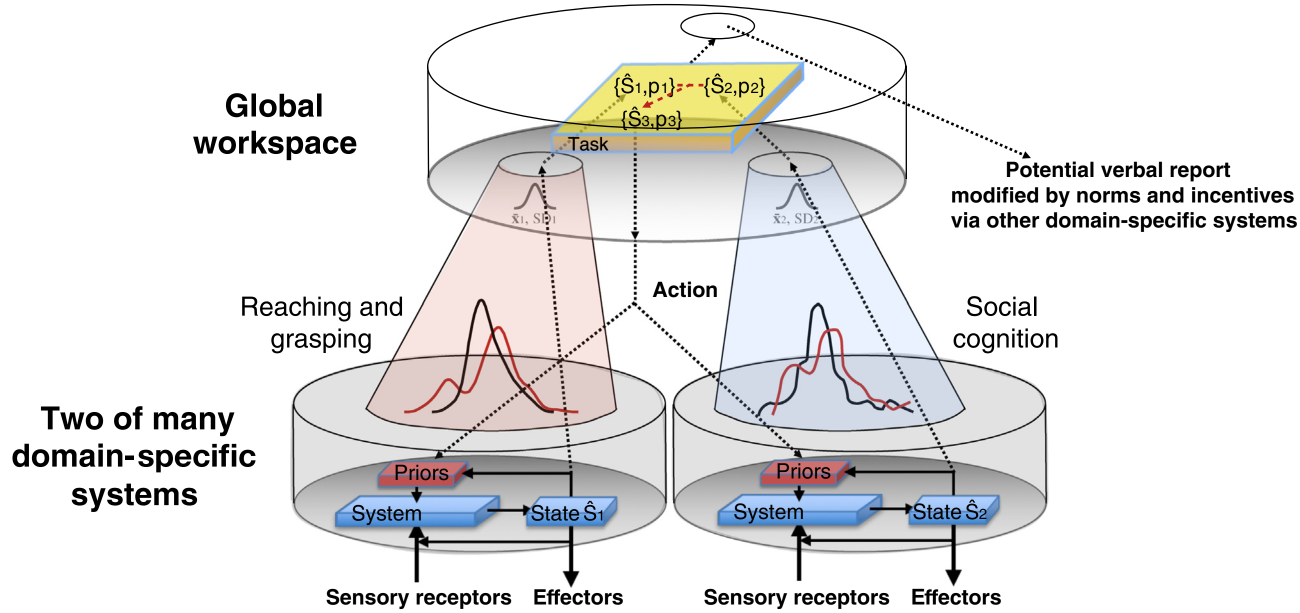
Percept: a conscious perceptual representation of the world, formed on the basis of sensory evidence.

Precision: the inverse of the variance of a signal. Low precision may reflect variation in the world and/or noise in processing. When precision is relied on in cue combination (see main text), systems treat it as reflecting noise in different perceptual channels, hence uncertainty. Treated that way, precision becomes a metacognitive parameter.

Scale-free: a representation that is independent of an absolute metric (e.g., percentage scores or standard scores).

WM maintenance: a form of working memory that maintains a representation in an active state. Current evidence suggests a representation can be maintained in working memory without remaining conscious (Box 1 in the main text).

WM manipulation: a form of working memory that allows representations to be integrated, manipulated, and altered (i.e., a workspace). Experiments suggest that, for such manipulation to occur, the representations must be, or have been, in consciousness.



Trends in Cognitive Sciences

Figure 1. A Simple Workspace Model with Inputs, Transitions, and Outputs. Input processes: the inputs to the workspace come from a variety of domain-specific systems. The two shown here (reaching and grasping, and social cognition) would be required for solving the delicate problem of 'when and how to shake hands'. These perceptuomotor systems process a wealth of information nonconsciously in the service of action. They deal in rich probabilistic hypotheses, perhaps even a full probability distribution over available possibilities ([78,79] but see also [80]). Through constant interaction with the world, domain-specific systems have learnt how to update probabilistic hypotheses in real time, approximating or perhaps implementing inferences that are Bayesian or otherwise optimal [81,82]. What is selected for global broadcast is a representation of a single state of affairs (*cf*. [83]), for example a maximum likelihood estimate, rather than a whole range of alternate possibilities [84–86]. These representations are a highly compressed summary of some of this information, involving substantial data reduction. The compression is achieved by coarse-graining, and by leaving some variables and states out of the representation in such a way that the effective information in the representation is increased [87]. This permits better communication of information within the workspace. Simplification is also needed to solve complex, multistep planning problems [88]. In the task involved in our example, one representation broadcast to the work space concerns the ease with which a hand shake could be achieved, the other concerns an inference about whether this person would expect a handshake. Workspace transitions: within the workspace, these compressed representations are manipulated and compared in order to draw conclusions or take decisions [46,54]. Estimates of confidence in these representations will have an important role in reaching decisions. This is discussed further in the main text. Output processes: the representations in the workspace are taken up by the appropriate domain-specific systems, leading to alterations in their state estimates. This will sometimes lead to action (e.g., the hand is shaken) and also a potential verbal output (e.g., reporting why I thought it appropriate, or not, to shake hands). The verbal reports frequently concern the subject's confidence about an item in the workspace. This would include confidence in the inputs (e.g., a percept) as well as in the outputs (e.g., a decision). How confidence is reported will be modified by the social context in which the report is made [89], taking into account strategic considerations [90]. Verbal reports can modify representations in the workspaces of others [23].

The functional consideration that favours confidence sharing between individuals applies equally when cognitive control is operating within an individual. The manipulation function of the workspace requires representations that conflict or support one another to be weighed and integrated so that conclusions can be drawn and decisions reached. That can be done more effectively if the representations contain, or are accompanied by, a rating of confidence or uncertainty (or other relevant metacognitive parameter). Computational models show that the benefits justify the extra cost of tracking confidence, and there is extensive evidence that this principle is relied on by other cognitive systems (Box 2).

Our thesis is not that working memory manipulations must involve metacognition. A system could reason and take decisions with representations without tracking their reliability. However, for these decisions to be optimal, there is a functional reason to expect metacognition to be involved in the workspace. It is likely that a computational principle, relied on by many other cognitive processes and which would allow the workspace to perform its functions more effectively, would be

Box 2. Computational Basis of the Functional Argument

When a cognitive process is presented with items of information that conflict or support one another, a computational route to integrating them optimally is to weigh each item of information by its respective reliability. This requires a confidence parameter to be assigned to each: (i) weighting by confidence operates as an optimal or near-optimal basis for integration within a sensory modality [104], across two sensory modalities [25,26,105], and in motor control [106]; (ii) therefore, the brain must somehow represent information about uncertainty for use in its computations [78]. Signals that appear to have this role are widely observed in the brain [107–110]; (iii) computations that weight information by confidence can be used to solve problems of robotic control [111]. Confidence-involving computations have recently begun to gain prominence in artificial intelligence research [112]; (iv) **coarse-graining** and other kinds of simplification of the right kind enable better communication of the causal structure of systems than at the microscale [87]. Optimising the simplification requires metacognitive parameters; (v) a common currency for representations (e.g., scale-free) enables direct comparison of representations. Computation of a common currency involves a metacognitive parameter (e.g., standard scores) [27,35]; (vi) achievement of complex, multistep planning requires heuristics that efficiently simplify, approximate, and hierarchically decompose hard tasks into simpler subtasks. Optimising the cost of computing the values of choices requires access to metacognitive parameters [88,113]; and (vii) thus, the functional argument for confidence marking, which a prominent global workspace theorist has argued applies to information shared between individuals ([24] pp. 111–112,247), is equally applicable when representations have to be integrated and weighed within an individual in the global workspace.

implemented in the workspace. Therefore, we should expect representations in the global workspace to involve metacognition, in the form of confidence, or some other relevant metacognitive parameter.

This is an argument for confidence being relied on by the WM manipulation processes that weigh, integrate and decide; that is, for the confidence information itself being globally available. Thus, we can think of an item in the workspace as representing something about the world and at the same time representing the accuracy of that representation. The content it makes globally available includes both aspects.

In defining metacognition earlier, we saw that not all probabilistic representations are metacognitive. Our functional argument supports the existence of metacognitive contents, not simply probabilistic contents, in the global workspace. As with cue combination, appropriate weighing and integrating in the workspace should largely be based on the reliability of the cognitive processes by which percepts were formed and selected for global broadcast (similarly for affective, motoric, and mnemonic representations). Confidence estimates are also relied for cognitive control. On both grounds [33] they fall within the purview of metacognition. Evidence of their intimate relation to confidence reports and load sensitive processing supports this conclusion, as do the brain areas involved (Box 3).

Common Currency

The function of the workspace is to allow highly disparate representations to be brought together so that cognitive work can be done with them. Incompatible representations drive one another out of the workspace, whereas compatible representations support one another in the workspace synergistically [34]. However, for these representations to be integrated, some form of common currency is required. One solution would be to use **scale-free** (normalised) representations. In statistics, for example, normalisation typically uses a ‘confidence’ measure, such as the standard deviation, to generate a standard score, $(X-\mu)/\sigma$. In this way, confidence can supply a common currency between different sensory modalities [27]. In the weighted confidence-sharing model, which gives a good account of the integration of information across people, confidence ratings are shared in the form of standard scores [32].

Common currency has also been discussed in relation to value. This is necessary for making choices between different types of reward. Ventromedial prefrontal cortex (vmPFC)/orbitofrontal cortex (OFC) has been identified as the region where value is represented in a common currency [35].

Box 3. Further Evidence in Favour of the Hypothesis

Fluency

Fluency affects the metacognitive ratings that are associated with workspace representations. A robust finding across many different tasks is that confidence is boosted by fluency of various kinds (perceptual, conceptual, motoric: [114]). Thus, potential answers presented 100 ms more quickly are more likely to be judged correct [46]. Falsely high confidence engendered by misleading fluency cues probably underlies people's tendency to stick with an initial incorrect answer in reasoning problems [115]. Confidence in memories is also partially determined by the fluency of retrieved information [15].

Learning

Learning about the environment can occur even in the absence of external feedback. This kind of learning is guided by self-generated feedback that is based on subjective confidence ratings [116–118]. Subjective confidence is used to weight prior knowledge and incoming evidence by their respective reliability [110]. Individuals with a better correspondence between confidence judgments and prediction accuracy learn more quickly [119]. Individuals with better metacognitive accuracy for a perceptual decision are also better able to learn novel arbitrary cue–stimulus associations for the stimuli the perceptual decision was based on [120].

Brain

The brain regions implicated in metacognition are also implicated in the manipulation of items in the workspace. This applies, in particular to the frontopolar cortex [Brodmann area (BA) 10], the region of frontal cortex most expanded in the human brain [121,122]. BA 10 is involved with making confidence judgements in both humans [54] and macaques [123]. In humans, BA 10 is also seen as a gateway between sensory representations (whether input driven or self-generated) and downstream supervisory processes, including WM manipulation [124,125], and also as a means for interposing different behavioural plans into dorsolateral PFC so that the currently relevant plan can be executed at the right time [126], including when a behaviour switch is based on accumulating uncertain evidence [127]. Thus, BA 10 is involved in inputting representations into WM manipulation. Its role in metacognitive accuracy suggests that the representations passed through this gateway are confidence labelled, in line with our hypothesis.

Scale-free representations will not be appropriate for domain-specific systems. For example, reaching and grasping requires a precise representation of spatial scale in egocentric coordinates. In working memory, by contrast, representations of space are thought to be in allocentric coordinates [36]. After a delay, grasp scaling shifts from absolute to relative metrics [37]. Allocentric representations provide a common currency for space that is independent of location and, therefore, not suitable for reaching actions. This distinction is consistent with findings that affordance-based, in-the-moment action selection is not a feature of the representations manipulated in working memory [38,39] and is further supported by the double dissociation observed in the reaching behaviour of subjects with visual form agnosia and those with optic ataxia. The former can perform online reaching in-the-moment but are impaired when working memory is required and they have to reach for a target that has disappeared [40]. Those with optic ataxia are impaired at online reaching, but their performance improves after a delay [41–43]. Likewise, in the normal case dual-task interference has greater effects on delayed grasping than does in-the-moment grasping [44].

Sources of Uncertainty

There are several sources of uncertainty that affect manipulation in the workspace. For example, I may see a Hershey Bar™ and be reminded that my nephew likes candy bars. However, I also remember that he is allergic to nuts. All the pieces of information in the workspace relevant to my decision to buy the bar are associated with different levels of confidence. Are Hershey bars nut-free? Is Jack allergic to other things besides nuts? How likely is Jack to have an allergic reaction? Will I feel regret if he does have an allergic reaction? Since the first representation has low confidence and I anticipate regret with high confidence, I decide not to buy the bar.

A major source of uncertainty in the conclusion of a reasoning task can be the reasoning process itself [45]. For instance, in assessing a solution to an anagram, perceptual uncertainty about what the letters are is unlikely to be relevant, compared with uncertainty about the steps that mediate between the anagram and its putative solution [46].

In a perceptual decision-making task, by contrast, the primary source of uncertainty in the decision is uncertainty in the percepts on which it is based [47–52]. Perceptual decision-making is not normally considered to be a working memory task. However, it does often involve comparing or manipulating items. The subject may have to decide whether the preponderant direction of motion of a random dot display lies to the left or right of a given reference line, or whether the first or second array of contrast gratings contains an oddball [53,54]. Thus, the decision is the result of a computation that takes two globally broadcast representations as input. That decision then needs to be held in working memory, together with an associated level of confidence, to form the basis of a subsequent confidence report (the output process).

Evidence That Metacognitive Parameters Are Represented in the Workspace

Our hypothesis makes the strong prediction that every representation subject to manipulation in the workspace should have a metacognitive component. If it were combined with the claim that representations outside the workspace always lack a metacognitive component, the hypothesis would be straightforward to test. However, since confidence is also often found in nonconscious representations [16,17], our contention is that metacognition is ever-present in the global workspace and only sometimes present elsewhere. Although that is a difficult prediction to test exhaustively, there is much positive evidence for it. This evidence comes from studies exploring: (i) explicit reports of confidence; (ii) the relationship between confidence and cognitive load; and (iii) automatic error detection using error-related negativity signals. Box 3 discusses additional evidence in favour of the hypothesis from studies of **fluency**, learning, and brain mechanisms.

Explicit Reports of Confidence

If working memory representations have a metacognitive component, then we would expect confidence ratings in working memory tasks to correlate, to some extent, with task performance. This is the case for visual working memory tasks, suggesting that representations held in working memory do contain more than a point estimate [55,56]. This is well captured by a model in which working memory representations are associated with a precision parameter, which is transformed to give a reported level of confidence [31,57]. Conversely, for behaviour based on representations that are not globally broadcast (e.g., unconscious priming), there is no reason to expect reported confidence to correlate with performance and, indeed, in some paradigms it does not [58,59].

The fact that percepts come with a certainty or uncertainty that feeds into subjects' reported confidence does not imply that the certainty perfectly tracks accuracy or that the subject will show perfect metacognitive efficiency. Confidence reports are a measure but only an imperfect measure of accuracy [60]. Metacognitive accuracy for detection is higher for stimulus-present than for stimulus-absent trials [61]. A ready explanation for this is that percepts carry a (moderately reliable) certainty parameter, on which the subject's confidence judgement is based when there is a stimulus present. When there is no stimulus, there is no percept and, hence, no accompanying certainty parameter, leaving the subject with little basis for making reliable confidence judgements.

The finding that metacognition can be inaccurate also features in work on metamemory. Extensive research on the metacognition of memory has shown that retrieved memories are associated with a level of confidence. That confidence assignment is an imperfect guide to accuracy, since it is based on cues and heuristics, such as fluency of recall, that can be misleading [15]. The fact that metacognitive accuracy is imperfect does not indicate, *pace* Dehaene *et al.* ([10] p. 6), that confidence dissociates from global broadcast. Rather, metamemory research confirms that memories retrieved into the global workspace are accompanied by a measure of confidence, albeit one that is imperfect.

The Relationship between Cognitive Load and Confidence

One way of testing whether a confidence parameter lies in or outside the global workspace is to see whether it is affected by concurrent working memory load. It is well established that cognitive load reduces the precision with which items are coded in working memory [62,63]; thus, if confidence representations are indeed in the workspace, we should expect them to be compromised by cognitive load. Of course, cognitive load has wide-ranging effects on performance, so this is not a very specific prediction. However, there is no across-the-board reason why confidence represented outside the workspace should be directly affected by load.

In line with this prediction, metacognitive efficiency in a perceptual task is compromised by concurrent manipulation of items in an unrelated task (but not by mere maintenance) [64]. Since we hypothesise that percepts carry a confidence parameter in addition to their object-level contents about the world, it should be possible to interfere with metacognition while preserving object-level task performance. Indeed, loading or interfering with working memory can have the effect of reducing confidence even when performance is not affected [60].

Further supporting the relationship between confidence and cognitive load, studies have found that individual differences in working memory capacity are correlated with metacognitive performance [65]. Individuals with higher capacity are more able to adjust their response bias to account for their perceptual sensitivity [66]. Similarly, model-based reasoning, which calls for the manipulation of probabilities in working memory, is stronger in subjects who exercise more cognitive control in standard tasks [67] and is impaired by cognitive load [68]. Perceptual and metacognitive vigilance appear to depend on a shared, limited cognitive resource [69].

Finally, it is known that executive working memory load, as well as visuospatial load, increases the detection threshold in a visual task [70]. This kind of modulation does not appear to extend to representations that are not conscious. In a task in which both low-visibility and high-visibility items contribute to a decision, a top-down modulation that aimed to reduce reliance on incoming evidence only had the effect of reducing reliance on high-visibility stimuli, not on low-visibility stimuli [71]. Thus, the way that WM manipulation affects confidence appears to be specific to representations that have been made globally available (high-visibility stimuli). Conversely, confirming evidence that aids task performance but is only represented unconsciously does not improve metacognitive performance [72].

Automatic Error Detection

Our hypothesis contends that conscious representations always have a metacognitive component. However, it is possible that these are only induced when a confidence judgement is called for. The phenomenon of automatic error detection [16,73] provides evidence that this is not the case. Automatic error detection can be assessed using the classic error-related negativity (ERN) signal recorded from frontocentral electroencephalogram (EEG) electrodes. This signal occurs ~100 ms after an incorrect response. Critically, the ERN is produced even when participants are not instructed to make a confidence judgement, provided that the relevant stimuli have been consciously perceived [16]. Furthermore, the magnitude of the ERN is reduced under cognitive load [74] and increased when the subject has less confidence in a perceptual decision [75]. Error positivity (Pe) also scales with confidence [76]. The fact that the ERN is produced by an automatic process suggests that the confidence signals that drive it are always present when a response is based on conscious stimuli.

Concluding Remarks

On the basis of a functional analysis, we have argued that metacognitive parameters have a critical role in the manipulation of representations that have been globally broadcast to the workspace. We have also highlighted some of the empirical evidence in favour of this proposal. This evidence suggests an extensive entanglement between metacognition and manipulation in working memory.

Of course, future work should continue to investigate this entangled relationship and test specific predictions of our hypothesis (Box 4). For example, it is unclear how the confidence that an agent reports in an explicit decision task (e.g., when making a forced choice between two alternatives) relates to the confidence parameter associated with each of the two percepts that have been broadcast to the global workspace. Fortunately, implicit behavioural and neural signals of confidence and automatic error detection are known and, therefore, future work could systematically investigate how these relate to the levels of confidence explicitly reported by subjects (see Outstanding Questions).

Recently, there have been advances in training people to control patterns of brain activity (using decoded neural feedback [77]). These advances mean that it might be possible to intervene directly on the confidence attached to representations in the workspace. For example, it may be possible to bias a participant's choice in favour of or against the second of two stimuli by simply instructing them before the second stimulus to activate a high- or low-confidence neural pat-

Box 4. Additional Predictions

Cognitive Load

Our hypothesis predicts that loading WM manipulation should always impair metacognitive accuracy. It should also compromise the ability to weight representations by confidence in the course of WM manipulation. Thus, we predict that load will cause subjects to give undue weight to low-confidence representations when taking decisions or reaching conclusions, similar to the way load disposes people to believe what they hear uncritically [128].

Where stimuli that are unseen as well as those that are seen contribute to a perceptual decision, the hypothesis predicts that it is only confidence in consciously experienced stimuli that will be modulated by concurrent working memory load. This could be tested by adding unconscious (noisy) primes into the design of [64], in which the impact of WM manipulation on metacognitive efficiency was first demonstrated.

We have seen that the ERN is higher for low-confidence stimuli [75] and is affected by load [74]. A straightforward further prediction is that the extent to which high- versus low-confidence percepts generate a difference in ERN should be reduced under cognitive load.

Online versus Offline Reaching and Grasping

We have already referred to the distinction between online reaching and action guidance after a delay (under 'Common Currency' in main text). Only the latter is guided by representations in working memory and is strongly affected by dual-task interference. Online reaching and grasping requires precise representations of location in allocentric coordinates, whereas offline reaching and grasping plausibly uses a common currency in allocentric coordinates. The hypothesis predicts that noise should affect these two systems in different ways.

There is evidence that speeded reaching can integrate information over trials in a Bayes optimal manner [129]. Our hypothesis predicts that noise will have a different effect on reaching trajectories in the online case (all evidence is probabilistically integrated into the decision) than in the offline case (only the pared-down location-plus-confidence affects reaching). In the latter, but not the former, the reaching trajectory should show the positive evidence bias (the finding that confidence in a two-alternative forced choice only reflects the amount of positive evidence for the decision, not the balance between evidence for and against the decision [50,51,56]). Such a result would support our characterisation of the nature of representations in the workspace.

Outstanding Questions

How is confidence in a workspace representation computed from probability distributions in the domain-specific systems that broadcast it?

Is there a confidence threshold, below which a representation cannot enter the workspace?

How is the confidence associated with a percept related to its acuity and its reported visibility?

Is the confidence attached to an item in the workspace affected by the confidence attached to other items in the workspace?

Does the confidence rating of a workspace representation affect the way attention is directed?

Does it take more attention to sustain a low- than a high-confidence representation in the global workspace?

How does the level of confidence reported in explicit judgements relate to the confidence accompanying the percepts on which judgements are based, measured implicitly (e.g., through reaction times or automatic error detection)?

Our hypothesis implies that, for conscious percepts, the effect of evidence strength on decision confidence should be affected by load, whereas, for unconscious percepts, it should not. Can unconscious primes be used to test for this asymmetry?

Does the confidence revealed by the way reaching is executed reflect the positive evidence bias?

If participants are trained through decoded neural feedback (DecNef) to produce the neural signature of low confidence, does that reduce the weight placed on the information in a concurrent stimulus when it is integrated into a subsequent decision?

Similarly, will DecNef for a low-confidence pattern increase the rate of correct responses in reasoning problems where the intuitive solution is incorrect?

Is there a common neural mechanism (e.g., in Brodmann area 10) for assigning confidence to representations that are tern. Furthermore, if confidence lowering extended to other tasks, for example the tendency to endorse an intuitive but incorrect answer in a verbal judgement-and-decision problem, then that would show that the mechanism for assigning confidence to representations broadcast to the workspace is domain general.

input to the global workspace, or are there different neural mechanisms of confidence assignment for different domain-specific systems?

In conclusion, our proposal refines our understanding of the nature of both the global workspace and the representations it makes globally available. That brings us one step closer to a better characterisation of consciousness.

Acknowledgements

For helpful comments, the authors would like to thank two anonymous referees, Ned Block, and audiences at the Institute of Philosophy and at the 2018 Annual Conference of the Association for the Scientific Study of Consciousness. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 681422 (MetCogCon).

References

1. Baddeley, A. (1992) Consciousness and working memory. *Conscious. Cogn.* 1, 3–6
2. Baars, B.J. (1988) *A Cognitive Theory of Consciousness,* Cambridge University Press
3. Dehaene, S. and Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79, 1–37
4. Dehaene, S. and Changeux, J.-P. (2011) Experimental and theoretical approaches to conscious processing. *Neuron* 70, 200–227
5. Kunimoto, C. *et al.* (2001) Confidence and accuracy of near-threshold discrimination responses. *Conscious. Cogn.* 10, 294–340
6. Persaud, N. *et al.* (2011) Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *NeuroImage* 58, 605–611
7. Cleeremans, A. (2014) Connecting conscious and unconscious processing. *Cogn. Sci.* 38, 1286–1315
8. Fleming, S.M. and Lau, H.C. (2014) How to measure metacognition. *Front. Hum. Neurosci.* 8, 443
9. Matthews, J. *et al.* (2018) Conscious access in the near absence of attention: critical extensions on the dual-task paradigm. *Philos. Trans. R. Soc. B* 373, 20170352
10. Dehaene, S. *et al.* (2017) What is consciousness, and could machines have it? *Science* 358, 486–492
11. Rosenthal, R. (2005) *Consciousness and Mind,* Clarendon Press
12. Beck, J.M. *et al.* (2008) Probabilistic population codes for Bayesian decision making. *Neuron* 60, 1142–1152
13. Nelson, T.O. (1996) Consciousness and metacognition. *Am. Psychol.* 51, 102–116
14. Koriat, A. (2007) Metacognition and consciousness. In *Cambridge Handbook of Consciousness* (Zelazo, P.D., *et al.*, eds), pp. 289–326, Cambridge University Press
15. Koriat, A. (2015) Metacognition: decision-making processes in self-monitoring and self-regulation. In *The Wiley Blackwell Handbook of Judgment and Decision Making* (Keren, G. and Wu, G., eds), pp. 356–379, Wiley–Blackwell
16. Charles, L. *et al.* (2013) Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage* 73, 80–94
17. Kanai, R. *et al.* (2010) Subjective discriminability of invisibility: a framework for distinguishing perceptual and attentional failures of awareness. *Conscious. Cogn.* 19, 1045–1057
18. Jachs, B. *et al.* (2015) On the independence of visual awareness and metacognition: a signal detection theoretic analysis. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 269–276
19. Timmermans, B. *et al.* (2012) Higher order thoughts in action: consciousness as an unconscious re-description process. *Philos. Trans. R. Soc. B* 367, 1412–1423
20. Fleming, S.M. *et al.* (2012) Metacognition: computation biology and function. *Philos. Trans. R. Soc. B* 367, 1280–1286
21. Snodgrass, M. *et al.* (2009) Access is mainly a second-order process: SDT models whether phenomenally (first-order) conscious states are accessed by reflectively (second-order) conscious processes. *Conscious. Cogn.* 18, 561–564
22. Dehaene, S. (2009) Neural global workspace. In *The Oxford Companion to Consciousness* (Bayne, T., *et al.*, eds), pp. 466–470, Oxford University Press
23. Shea, N.J. *et al.* (2014) Supra-personal cognitive control and metacognition. *Trends Cogn. Sci.* 18, 186–193
24. Dehaene, S. (2014) Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts, Viking (The Penguin Group)
25. Ernst, M.O. and Banks, M.S. (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433
26. Deroy, O. *et al.* (2016) Metacognition in multisensory perception. *Trends Cogn. Sci.* 20, 736–747
27. de Gardelle, V. *et al.* (2016) Confidence as a common currency between vision and audition. *PLoS ONE* 11, e0147901
28. Prinz, J. (2012) The Conscious Brain, Oxford University Press
29. Moran, R. *et al.* (2015) Post choice information integration as a causal determinant of confidence: novel data and a computational account. *Cogn. Psychol.* 78, 99–147
30. Rahnev, D. *et al.* (2016) Causal evidence for frontal cortex organization for perceptual decision making. *Proc. Natl. Acad. Sci. U. S. A.* 113, 6059–6064
31. Fleming, S.M. *et al.* (2018) Neural mediators of changes of mind about perceptual decisions. *Nat. Neurosci.* 21, 617–624
32. Bahrami, B. *et al.* (2010) Optimally interacting minds. *Science* 329, 1081–1085
33. Nelson, T.O. and Narens, L. (1990) Metamemory: a theoretical framework and new findings. *Psychol. Learn. Motiv.* 26, 125–141
34. MacGregor, L.J. *et al.* (2015) Sustained meaning activation for polysemous but not homonymous words: evidence from EEG. *Neuropsychologia* 68, 126–138
35. Levy, D.J. and Glimcher, P.W. (2012) The root of all value: a neural common currency for choice. *Curr. Opin. Neurobiol.* 22, 1027–1038
36. Schenk, T. (2006) An allocentric rather than perceptual deficit in patient D.F. *Nat. Neurosci.* 9, 1369–1370
37. Hu, Y. and Goodale, M.A. (2000) Grasping after a delay shifts size-scaling from absolute to relative metrics. *J. Cogn. Neurosci.* 12, 856–868
38. Pecher, D. *et al.* (2013) The role of affordances for working memory for objects. *J. Cogn. Psychol.* 25, 107–118
39. Still, J.D. and Dark, V.J. (2010) Examining working memory load and congruency effects on affordances and conventions. *Int. J. Hum. Comput. St.* 68, 561–571
40. Goodale, M.A. *et al.* (1994) Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia* 32, 1159–1178
41. Milner, A.D. *et al.* (2001) Grasping the past: delay can improve visuomotor performance. *Curr. Biol.* 11, 1896–1901
42. Schindler, I. *et al.* (2004) Automatic avoidance of obstacles is a dorsal stream function: evidence from optic ataxia. *Nat. Neurosci.* 7, 779

43. Rice, N.J. *et al.* (2008) Delay abolishes the obstacle avoidance deficit in unilateral optic ataxia. *Neuropsychologia* 46, 1549–1557
44. Singhal, A. *et al.* (2007) Dual-task interference is greater in delayed grasping than in visually guided grasping. *J. Vis.* 7, 5.1–12
45. Ackerman, R. and Thompson, V.A. (2017) Meta-reasoning: monitoring and control of thinking and reasoning. *Trends Cogn. Sci.* 21, 607–617
46. Topolinski, S. and Reber, R. (2010) Immediate truth – temporal contiguity between a cognitive problem and its solution determines experienced veracity of the solution. *Cognition* 114, 117–122
47. Aitchison, L. *et al.* (2015) Doubly Bayesian analysis of confidence in perceptual decision-making. *PLoS Comput. Biol.* 11, e1004519
48. Bang, J.W. *et al.* (2019) Sensory noise increases metacognitive efficiency. *J. Exp. Psychol. Gen.* 148, 437–452
49. Navajas, J. *et al.* (2017) The idiosyncratic nature of confidence. *Nat. Hum. Behav.* 1, 810–818
50. Zylberberg, A. *et al.* (2012) Decision making during the psychological refractory period. *Curr. Biol.* 22, 1795–1799
51. Maniscalco, B. and Lau, H. (2016) The signal processing architecture underlying subjective reports of sensory awareness. *Neurosci. Conscious.* 2016, niw002
52. Palser, E.R. *et al.* (2018) Altering movement parameters disrupts metacognitive accuracy. *Conscious. Cogn.* 57, 33–40
53. Luu, L. and Stocker, A.A. (2018) Post-decision biases reveal a self-consistency principle in perceptual inference. *eLife* 7, e33334
54. Fleming, S.M. *et al.* (2010) Relating introspective accuracy to individual differences in brain structure. *Science* 329, 1541–1543
55. Rademaker, R.L. *et al.* (2012) Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *J. Vis.* 12, 1–13
56. Peters, M.A. *et al.* (2017) Perceptual confidence neglects decision-incongruent evidence in the brain. *Nat. Hum. Behav.* 1, 0139
57. van den Berg, R. *et al.* (2017) Fechner's law in metacognition: a quantitative model of visual working memory confidence. *Psychol. Rev.* 124, 197–214
58. Koizumi, A. *et al.* (2015) Does perceptual confidence facilitate cognitive control? *Atten. Percept. Psychophys.* 77, 1295–1306
59. Samaha, J. *et al.* (2016) Dissociating perceptual confidence from discrimination accuracy reveals no influence of metacognitive awareness on working memory. *Front. Psychol.* 7, 851
60. Bona, S. and Silvanto, J. (2014) Accuracy and confidence of visual short-term memory do not go hand-in-hand: behavioral and neural dissociations. *PLoS ONE* 9, e90808
61. Meuwese, J.D. *et al.* (2014) The subjective experience of object recognition: comparing metacognition for object detection and object categorization. *Atten. Percept. Psychophys.* 76, 1057–1068
62. Bays, P.M. and Husain, M. (2008) Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854
63. Suchow, J.W. *et al.* (2014) Terms of the debate on the format and structure of visual memory. *Atten. Percept. Psychophys.* 76, 2071–2079
64. Maniscalco, B. and Lau, H. (2015) Manipulation of working memory contents selectively impairs metacognitive sensitivity in a concurrent visual discrimination task. *Neurosci. Conscious.* 2015, niv002
65. Komori, M. (2016) Effects of working memory capacity on metacognitive monitoring: a study of group differences using a listening span test. *Front. Psychol.* 7, 285
66. Lynn, S.K. *et al.* (2016) Working memory capacity is associated with optimal adaptation of response bias to perceptual sensitivity in emotion perception. *Emotion* 16, 155–163
67. Otto, A.R. *et al.* (2015) Cognitive control predicts use of model-based reinforcement learning. *J. Cogn. Neurosci.* 27, 319–333
68. Otto, A.R. *et al.* (2013) The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. *Psychol. Sci.* 24, 751–761
69. Maniscalco, B. *et al.* (2017) Limited cognitive resources explain a trade-off between perceptual and metacognitive vigilance. *J. Neurosci.* 37, 1213–1224
70. De Loof, E. *et al.* (2015) Different effects of executive and visuospatial working memory on visual consciousness. *Atten. Percept. Psychophys.* 77, 2523–2528
71. de Lange, F.P. *et al.* (2011) How awareness changes the relative weights of evidence during human decision-making. *PLoS Biol.* 9, e1001203
72. Vlassova, A. *et al.* (2014) Unconscious information changes decision accuracy but not confidence. *Proc. Natl. Acad. Sci. U. S. A.* 11, 16214–16218
73. Yeung, N. and Summerfield, C. (2012) Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. B Biol. Sci.* 367, 1310–1321
74. Krigolson, O.E. *et al.* (2012) Cognitive load impacts error evaluation within medial-frontal cortex. *Brain Res.* 1430, 62–67
75. Scheffers, M.K. and Coles, M.G.H. (2000) Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. *J. Exp. Psychol. Hum. Percept. Perform.* 26, 141–151
76. Boldt, A. and Yeung, N. (2015) Shared neural markers of decision confidence and error detection. *J. Neurosci.* 35, 3478–3484
77. Cortese, A. *et al.* (2016) Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nat. Commun.* 7, 13669
78. Knill, D.C. and Pouget, A. (2004) The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* 27, 712–719
79. Shen, S. and Ma, W.J. (2016) A detailed comparison of optimality and simplicity in perceptual decision making. *Psychol. Rev.* 123, 452–480
80. Rahnev, D. and Denison, R.N. (2018) Suboptimality in perceptual decision making. *Behav. Brain Sci.* 41, e225
81. Ma, W.J. *et al.* (2006) Bayesian inference with probabilistic population codes. *Nat. Neurosci.* 9, 1432–1438
82. Keshvari, S. *et al.* (2012) Probabilistic computation in human perception under variability in encoding precision. *PLoS ONE* 7, e40216
83. Block, N. (2018) If perception is probabilistic, why does it not seem probabilistic? *Philos. Trans. R. Soc. B* 373, 20170341
84. Stocker, A. and Simoncelli, E.P. (2008) A Bayesian model of conditioned perception. *Adv. Neural Inf. Proces. Syst.* 20, 1409–1416
85. Dehaene, S. and Sigman, M. (2012) From a single decision to a multi-step algorithm. *Curr. Opin. Neurobiol.* 22, 937–945
86. Rahnev, D. (2017) The case against full probability distributions in perceptual decision making. *bioRxiv* 2017, 108944
87. Hoel, E.P. (2017) When the map is better than the territory. *Entropy* 19, 188
88. Huys, Q.J. *et al.* (2015) Interplay of approximate planning strategies. *Proc. Natl. Acad. Sci. U. S. A.* 112, 3098–3103
89. Bang, D. *et al.* (2017) Confidence matching in group decision-making. *Nat. Hum. Behav.* 1, 0117
90. Hertz, U. *et al.* (2017) Neural computations underpinning the strategic management of influence in advice giving. *Nat. Commun.* 8, 2191
91. Owen, A.M. *et al.* (1996) Double dissociations of memory and executive functions in working memory tasks following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. *Brain* 119, 1597–1615
92. Petrides, M. and Milner, B. (1982) Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* 20, 249–262
93. Petrides, M. (2000) The role of the mid-dorsolateral prefrontal cortex in working memory. *Exp. Brain Res.* 133, 44–54
94. Stokes, M.G. (2015) 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405
95. King, J-R. *et al.* (2016) Brain mechanisms underlying the brief maintenance of seen and unseen sensory information. *Neuron* 92, 1122–1134

96. Trübutschek, D. *et al.* (2017) A theory of working memory without consciousness or sustained activity. *eLife* 6, e23871
97. Stein, T. *et al.* (2016) Can working memory be non-conscious? *Neurosci. Conscious.* 2016, niv011
98. Soto, D. and Silvanto, J. (2016) Is conscious awareness needed for all working memory processes? *Neurosci. Conscious.* 2016, niw009
99. Soto, D. *et al.* (2011) Working memory without consciousness. *Curr. Biol.* 21, R912–R913
100. Sklar, A.Y. *et al.* (2012) Reading and doing arithmetic nonconsciously. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19614–19619
101. Soto, D. and Silvanto, J. (2014) Reappraising the relationship between working memory and conscious awareness. *Trends Cogn. Sci.* 18, 520–525
102. Jacobs, C. *et al.* (2018) Visual working memory performance in aphantasia. *Cortex* 105, 61–73
103. Wahn, B. and König, P. (2015) Vision and haptics share spatial attentional resources and visuotactile integration is not affected by high attentional load. *Multisens. Res.* 28, 371–392
104. Jacobs, R.A. (1999) Optimal integration of texture and motion cues to depth. *Vis. Res.* 39, 3621–3629
105. Alais, D. and Burr, D. (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262
106. Körding, K.P. and Wolpert, D.M. (2004) Bayesian integration in sensorimotor learning. *Nature* 427, 244
107. Summerfield, C. and Koechlin, E. (2008) A neural representation of prior information during perceptual inference. *Neuron* 59, 336–347
108. Fetsch, C.R. *et al.* (2012) Neural correlates of reliability-based cue weighting during multisensory integration. *Nat. Neurosci.* 15, 146–154
109. De Martino, B. *et al.* (2017) Social information is integrated into value and confidence judgments according to its reliability. *J. Neurosci.* 37, 6066–6074
110. Meyniel, F. and Dehaene, S. (2017) Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proc. Natl. Acad. Sci. U. S. A.* 114, E3859–E3868
111. Thrun, S. *et al.* (2001) Robust Monte Carlo localization for mobile robots. *Artif. Intell.* 128, 99–141
112. Ghahramani, Z. (2015) Probabilistic machine learning and artificial intelligence. *Nature* 521, 452–459
113. Gershman, S.J. *et al.* (2014) Retrospective revaluation in sequential decision making: a tale of two systems. *J. Exp. Psychol. Gen.* 143, 182–194
114. Unkelbach, C. and Greifeneder, R. (2013) A general model of fluency effects in judgment and decision making. In *The Experience of Thinking: How the Fluency of Mental Processes Influences Cognition and Behaviour* (Unkelbach, C. and Greifender, R., eds), pp. 11–32, Psychology Press
115. Thompson, V.A. *et al.* (2013) Matching bias on the selection task: it's fast and feels good. *Think. Reason.* 19, 431–452
116. Daniel, R. and Pollmann, S. (2012) Striatal activations signal prediction errors on confidence in the absence of external feedback. *NeuroImage* 59, 3457–3467
117. Guggenmos, M. *et al.* (2016) Mesolimbic confidence signals guide perceptual learning in the absence of external feedback. *eLife* 5, e13388
118. Guggenmos, M. and Sterzer, P. (2017) A confidence-based reinforcement learning model for perceptual learning. *bioRxiv* 2017, 136903
119. Frömer, R. *et al.* (2018) I knew that! Confidence in outcome prediction and its impact on feedback processing and learning. *bioRxiv* 2018, 44282
120. Hainguerlot, M. *et al.* (2018) Metacognitive ability predicts learning cue–stimulus associations in the absence of external feedback. *Sci. Rep.* 8, 5602
121. Donahue, C.J. *et al.* (2018) Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. *Proc. Natl. Acad. Sci. U. S. A.* 115, E5183–E5192
122. Semendeferi, K. *et al.* (2001) Prefrontal cortex in humans and apes: a comparative study of area 10. *Am. J. Phys. Anthropol.* 114, 224–241
123. Miyamoto, K. *et al.* (2018) Reversible silencing of the frontopolar cortex selectively impairs metacognitive judgment on non-experience in primates. *Neuron* 97, 980–989
124. Burgess, P.W. *et al.* (2007) The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends Cogn. Sci.* 11, 290–298
125. Burgess, P.W. *et al.* (2007) Function and localization within rostral prefrontal cortex (area 10). *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 362, 887–899
126. Koechlin, E. and Hyafil, A. (2007) Anterior prefrontal function and the limits of human decision-making. *Science* 318, 594–598
127. Boorman, E.D. *et al.* (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62, 733–743
128. Gilbert, D.T. (1991) How mental systems believe. *Am. Psychol.* 46, 107–119
129. Hudson, T.E. *et al.* (2012) Speeded reaching movements around invisible obstacles. *PLoS Comput. Biol.* 8, e1002676