HHS Public Access

Author manuscript

Neurosci Biobehav Rev. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

Neurosci Biobehav Rev. 2015 August ; 55: 88–97. doi:10.1016/j.neubiorev.2015.04.006.

There are things that we know that we know, and there are things that we do not know we do not know: Confidence in Decision-Making

Piercesare Grimaldi, Ph.D.[1,2,3,4], **Hakwan Lau, Ph.D.**[2,4], and **Michele A. Basso, Ph.D.**[1,3,4]

[1]Departments of Psychiatry and Behavioral Sciences and Neurobiology, University of California at Los Angeles USA

[2]Department of Psychology, University of California at Los Angeles USA

[3]The Semel Institute for Neuroscience, University of California at Los Angeles USA

[4]The Brain Research Institute, University of California at Los Angeles USA

Abstract

Metacognition, the ability to think about our own thoughts, is a fundamental component of our mental life and is involved in memory, learning, planning and decision-making. Here we focus on one aspect of metacognition, namely confidence in perceptual decisions. We review the literature in psychophysics, neuropsychology and neuroscience. Although still a very new field, several recent studies suggest there are specific brain circuits devoted to monitoring and reporting confidence, whereas others suggest that confidence information is encoded within decision-making circuits. We provide suggestions, based on interdisciplinary research, to disentangle these disparate results.

Keywords

metacognition; confidence; consciousness; awareness; monitoring

INTRODUCTION

Thinking about our own thoughts and knowledge - encapsulated by the infamous quote from the former US Secretary of Defense, Donald Rumsfeld and paraphrased in the title of this review - is referred to as metacognition. How we know what we know has captured the interest of philosophers since ancient times. Aristotle, in his *De Alma* (1987), speculated that the act of judging our own thoughts is necessary for remembering.

To whom correspondence is addressed: Piercesare Grimaldi, Ph.D. Departments of Psychiatry and Biobehavioral Sciences and Psychology The Semel Institute for Neuroscience and Human Behavior UCLA 635 Charles Young Drive Box 957332 Los Angeles, CA 90095.

The ability to reflect upon our own thoughts has been considered a logical conundrum for centuries. The idea of recursive monitoring evokes the image of a looker inside the looker, which implies infinite regression. This idea held back advancement of this field for a long time and induced a rational thinker like Descartes, to propose a disembodied self, a soul to solve the problem (Descartes, 1999). Centuries later, French philosopher Auguste Comte thought that the notion that an individual can turn his mental faculties inward is logically impossible (See James, 1983 for a discussion). It is paradoxical, Comte argued, that the mind might divide into two minds to allow self-observation. If there is a looker inside the looker, he reasoned, there needs to be a third looker and so on.

In modern neuroscience we commonly assume that the brain is modular (Fodor, 1983), so it is reasonable to postulate a module specialized in monitoring other modules. Paraphrasing Humphrey (2003): “No one would say that a person cannot use his own eyes to observe his own feet”. Although considering the brain as modular possibly saves us from the need to postulate a mind inside the mind, we still do not know how metacognition is encoded in the brain. Is it implemented by the same circuits that encode decisions or do specialized modules (a looker) monitor the activity of decision-making areas?

Today, the introspective nature of metacognition is considered a core part of what makes us human and a necessity to form the basis of conscious awareness (Terrace and Metcalfe, 2005). The role of metacognition in learning and memory drives research in the field of educational psychology (Bransford et al., 2005). Neuroscientists are also beginning to see metacognition as a topic amenable to inquiry. Here it is important to make a distinction between metacognition and consciousness. Metacognition in perception is linked to mechanisms of conscious perception. The exact relationship between the two may be complicated, and in general, issues surrounding consciousness are controversial (Lau and Rosenthal, 2011). Here we define metacognition as one specific aspect of consciousness, namely the ability for one to introspectively appreciate or monitor the quality of an ongoing perceptual process. Metacognition is not synonymous with consciousness because the latter is associated with a wider plethora of concepts including wakefulness, arousal, self-awareness, control of action, etc. However, metacognition is related to consciousness as is seen in neurological cases of the abolishment of perceptual awareness, such as occurs in blindsight. In these patients, metacognition seems to be impaired: blindsight subjects fail to report confidence in their responses even though their responses, i.e. guesses, reflect good perceptual capacity (Ko and Lau, 2012). In the modern memory literature, metacognition starts to be treated as a scientific subject as early as 1965 (Hart, 1965); see Tulving and Madigan (1970), for a review. In this context, metacognition is often described as either prospective or retrospective (Fleming and Dolan, 2012). Prospective metacognition refers to making judgments or predictions about what information will be available in memory in the future. Retrospective metacognition, in contrast, involves making judgments about a past experience, specifically about whether a memory item has been successfully encoded. How does metacognition as discussed here relate to introspection exactly? Introspection, which has been used for centuries by philosophers to explore our internal word, relies on the insight of the subject. Metacognition, as it is considered in modern literature, depends more on an operational definition, determined by reported levels of confidence in perceptual or memory responses given by the subject. Of course, in assessing metacognition, the hope is

that the subjects' reported level of confidence arises from introspection of the ongoing perceptual or memory process. However, one needs to be careful to rule out that reports of confidence are not driven by other factors such as social pressure too (Asch, 1951).

Although metacognition broadly refers to the psychological processes used to plan, monitor and assess an individual's own thoughts, knowledge and performance, here we shall focus on retrospective metacognition, and more specifically, on one aspect of it, namely confidence about the correctness of a decision.

In this review, we explore recent work in the field. We begin by introducing a definition of confidence and how it is measured. We then provide a brief history of ideas that form the basis for current approaches designed to understand confidence in decision-making. Next, we describe recent experimental data that are beginning to unravel the neural basis of confidence in decision-making. Finally, in light of existing conflicting data, we provide some suggestions for how to untangle the question of how the brain encodes confidence. Because much of the work on the neuronal basis of confidence capitalizes on developments in our understanding of oculomotor circuitry and visual perceptual decision-making, we will focus primarily on confidence in the context of vision.

What is confidence and how can it be measured?

Confidence is a belief about the validity of our own thoughts, knowledge or performance and relies on a subjective feeling (Luttrell, 2013). However, in the past decades, several methods have been developed to measure confidence objectively. Until recently, confidence was studied mainly in humans. Recent work, which we review below, suggests that non-human animals may also experience some levels of confidence in decision-making. In this paragraph, we will first review how confidence is measured in humans and then discuss the recent advances in how confidence is measured in animals.

In studies of human perceptual decision-making, confidence is often measured with retrospective judgment. Subjects give a confidence rating right after a report about a perceptual experience and therefore must base their confidence judgment on the memory of their initial response. For example, a subject might first perform some perceptual task such as reporting their perception of an ambiguous object (do you see a vase or a face?). Then the subject would immediately declare how confident s/he felt about that decision.

Similar to measures of confidence using open-ended ratings, several scales have been developed to measure confidence more quantitatively. The most commonly used is confidence rating. In this scale, the subject is asked to report confidence on a continuous scale ranging from 0% or complete uncertainty to 100% or complete certainty. Alternatively, it can be assessed with discrete fixed levels, or a simple binary choice (confident/not confident, Cheesman and Merikle, 1986; Dienes and Perner, 1999). However the use of ratings has been criticized because some subjects may find it not intuitive or they may be poorly motivated to accurately report their confidence (Persaud et al., 2007). To overcome these limitations, post-decision wagering has been introduced, in which subjects bet money or tokens on their own decisions (Persaud et al., 2007; Ruffman et al., 2001). In this context, subjects should ideally bet low when they are not confident and bet high when

they are confident, in order to maximize gain. This task is more engaging and more intuitive for most participants. However, it has been noted that wagering can be influenced by individual propensity to risk (Fleming and Dolan, 2010) and that subjects tend to use only the ends of the scale, probably in order to maximize gains (Sandberg et al., 2010), thus suffering from low sensitivity for intermediate ranges. In an attempt to develop a scale that has both the sensitivity of confidence ratings and the intuitiveness of post decision wagering, the feeling of warmth scale has been developed (Metcalfe, 1986; Wierzchon et al., 2012). In this scale subjects report their confidence as a temperature, ranging from cold (not confident) to hot (very confident), with intermediate options (e.g. chilly or warm). The perceptual awareness scale (Ramsøy and Overgaard, 2004) and the Sergent-Dehaene scale (Sergent and Dehaene, 2004) are also commonly used and were developed to judge the degree of visibility in visual tasks, ranging from no visibility at all to clear perception, with discrete intermediate levels (perceptual awareness scale), or a continuous spectrum (Sergent-Dehaene scale). When applied to confidence, however, these two scales end up being very similar to confidence rating. An extensive discussion of the properties and sensitivities of the different scales is beyond the scope of this review. For a rigorous comparison see, Sandberg et al., (2010) and Wierzchon et al., (2012).

Efforts directed towards measuring confidence in non-human animals started as early as two decades ago. Studying confidence in animals is obviously more difficult, but the advantage is that it opens the door to investigations of complex, cognitive phenomena at the neuronal level. Historically, Descartes thought that language was necessary for reflective thinking, so he excluded non-human animals from having the capacity for metacognition (Descartes, 1999). This belief persists, causing research on metacognition in non-human animals to face skepticism; however studies discussed below indicate that this skepticism may be ill placed.

One early study of confidence in animals was done by Shields and coworkers (Shields et al., 1997). Trained apes, monkeys and dolphins performed a classification task in which visual stimuli were labeled as being in one category or another. The visual stimuli were gradually blended into one another making the categorical discrimination more or less difficult. For some trials, the animals were offered a sure choice, which yielded a small but certain reward. Other trials were rewarded only if the stimulus was correctly categorized. The experimenters reasoned that if the animals were aware of not knowing the correct answer, they would choose the sure choice on those trials when the category of the stimulus presented was ambiguous. Indeed, the animals chose the sure choice more frequently in ambiguous conditions, consistent with the idea that the animals were aware of not knowing the correct answer.

A criticism of the Shield et al.'s work is that the animals could have associated the intermediate stimulus categories with the sure choice. If they did this, then the sure choice option was simply a third stimulus category, and not an indicator of low confidence. To address this potential pitfall, Hampton (Hampton, 2001) devised a prospective memory confidence task. Trained monkeys performed a delayed-match-to-sample task. In this task, an image, referred to as the target, appeared at the beginning of a trial. At the end of the trial, after a delay, animals were required to select the target that reappeared with another series of images (distractors). To evaluate if the monkeys remembered the target, after two thirds of

the delay, the monkeys received an option to accept the test or decline it. If they accepted it, and they made a correct match, they received a large reward. However, if they made a mistake by choosing a distractor as a match, then they received no reward. If the monkeys declined the test, they received a small reward, regardless of whether they chose the target or a distractor as the match. The investigators reasoned that if the monkeys believed they would perform well, they would accept the test, choose the correct target and receive a large reward. However, if they were uncertain, they would decline the test, and opt for the small but certain reward. Therefore, in this task, the monkeys made a prospective judgment about how they were likely to perform on the test. An additional strength of the task design was that four stimuli were used as possible targets and were selected as targets randomly each day. Because the stimulus set changed across sessions, the monkeys could not associate one particular stimulus with the likelihood of correct responding. Rather, they had to rely on their memory of the sample stimulus to decide whether to accept or decline the next step of the task. In addition, the monkeys could not use cues such as their own reaction time to estimate the likelihood of a correct response, because they had to decide to take the test or not before the match choice was required. As predicted, monkeys opted out when they did not remember, consistent with the hypothesis that they were less confident on those trials. Indeed, they performed better in this task then in a similar forced task, when they did not have the option to decline the test.

In a similar spirit, Son and Kornell (2005) trained rhesus macaque monkeys to distinguish the length of two lines. After the monkeys made their decision, consisting of choosing the longest line, they were required to rate their confidence in their decision by making a bet, that is, a retrospective task. Two betting options were represented by two choice targets. If the monkeys chose the low bet target, they received a small reward, regardless of whether their previous response on the discrimination task was right or wrong. If they chose the high bet target, they received a large reward for correct responses and no reward for incorrect responses. Monkeys generally chose low rewards more frequently in difficult discrimination trials indicating that they knew when they did not know. The same monkeys engaged in the same betting strategy during a dot-density discrimination task, showing that they could generalize their reports of confidence to different tasks. Similar approaches have been used to study confidence in smaller mammals such as rodents. Foothe and Crystal (2007) trained rats to discriminate the duration of sounds. In each trial, the rats were able to choose if they wanted to take a test or not. Similar to the monkeys, rats chose to avoid the test when the stimulus was ambiguous.

As mentioned above, a major criticism of this kind of research is that the animals might learn to automatically optimize behavior, without using metacognition, this is especially likely in Shields et al's experiment (Shields et al., 1997) where the sure choice may be interpreted as a third choice. However, this seems unlikely at least in Son and Kornell's study (Son and Kornall, 2005), because several strategies were used to ensure that the animals performed the general exercise of confidence rating, rather than just producing specific patterns of behavior that may superficially resemble this. The behavior of the animals was consistent across different visual categories and the same behavior was observed consistently across different experimental designs such as a sure choice/opt out

strategy (Shields et al., 1997), a betting strategy (Son and Kornell, 2005) and an escape strategy (the possibility to avoid being tested, Hampton, 2001).

In the case of Hampton's experiment (Hampton, 2001), non-metacognitive strategies were also carefully ruled out. Strategies based on external cues, such as those based on the features of the stimulus were ruled out because the target was different in every session. The animal could not have used its own reaction times as a measure of certainty, since the escape choice was given before the decision stage.

Although this does not mean that all studies here summarized are immune to the criticism, the evidence taken together does suggest that non-human animals are able to estimate and report their confidence.

How is confidence quantified?

Although the studies described above provide important insight into the nature of confidence, one limitation is that they used relatively simple quantitative analysis, such as how often an ambiguous stimulus was followed by a sure choice (Shields et al., 1997) or how often a correct choice was endorsed with a high confidence judgment (Son and Kornell, 2005). To study the “meaningfulness” of confidence ratings, researchers often assess the across-trial correlation between accuracy and confidence; a high degree of correlation is interpreted as an indication of a high degree of reliability of the measurement of confidence (Kornell et al., 2007), or in short, high metacognitive sensitivity.

However, the raw statistical correlation between accuracy and confidence is not necessarily a precise measure of metacognitive sensitivity. The correlation is a function of both how well the subject can distinguish between correct and incorrect trials, as well as the overall propensity to give high or low confidence responses. If subjects are very liberal, reporting high confidence frequently, or if they are very conservative, often reporting low confidence, then the interpretation of the correlation strength is limited. As it happens for quantifying perceptual judgment, likewise, the application of signal detection theory (SDT) in this case provides a more reliable approach to quantifying metacognitive sensitivity (Galvin et al., 2003; Maniscalco and Lau, 2012).

In a SDT perspective, metacognitive sensitivity is modeled in a conceptually similar manner to discrimination performance (Galvin et al., 2003; Maniscalco and Lau, 2012). In traditional SDT, we imagine two distributions of internal responses and the observer has to judge which distribution the stimulus presented belongs to. In SDT, the observer sets a criterion (C in Figure 1A). All signals greater than C are assigned to one distribution and all those smaller than C are assigned to the other distribution (S1 and S2 in Figure 1A). Moving C along the X-axis, simulating changes in criterion, produces a receiver operating characteristic curve (ROC; Figure 1B) that gives the measurement of the discriminability of the two populations of signals across all possible criterion levels. In a similar way, we can assess the subject's ability to discriminate between correct and incorrect trials by rating their confidence level. Let us assume that the observer also sets two criteria Co1 and Co2 to determine confidence judgments. All responses greater than Co2 will be endorsed with high confidence, for choice S2 and all the responses smaller than Co1 will have high confidence

for choice S1. All the responses between Co1 and Co2 will have low confidence rating. Now let us consider only the right side of the graph (Figure 1C). Like for criterion C, moving Co2 over the range of x values produces an ROC curve for confidence (Figure 1D). This method offers an unbiased way to quantify confidence sensitivity, i.e. one's ability to distinguish between correct and incorrect trials.

Whereas this kind of SDT analysis of confidence has been applied to numerous human studies (for a recent review, see Fleming and Lau, 2014), it has received little attention in the animal literature. In fact, with a few exceptions (Kepecs et al., 2008; and Kiani and Shadlen, 2009), most animal studies on confidence typically do not apply formal psychophysics or computational models. As we will argue later in this review, we think that applying the tools of psychophysics may turn out to be critical for future investigation of the neural mechanisms of confidence.

How is confidence coded in the brain: Do we need a looker inside the brain?

In the current neuroscience literature there is a fair amount of confusion regarding how confidence is encoded in the brain. Some data indicate that confidence may be encoded by the same circuits involved in decision-making, others that confidence is monitored by dedicated structures. Before reviewing evidence in favor of one or another hypothesis, it is useful to indicate the structures that are commonly considered to be involved in decision-making. In general, it is thought that perceptual decisions evolve within sensorimotor regions of the brain that control action. For example, decision-related activity appears in the lateral intraparietal (LIP) area (Kim and Shadlen, 1999; Platt and Glimcher, 1999; Roitman and Shadlen, 2002), frontal eye field (FEF, Gold and Shadlen, 2000; Kim and Shadlen, 1999) and superior colliculus (Horwitz and Newsome, 1999, 2001; Kim and Basso, 2008, 2010), when eye movements are used to report decisions, and in the parietal reach region, when hand movements are used to report decisions (Scherberger and Andersen, 2007). Recently, also the caudate nucleus of the basal ganglia was identified as carrying decision-related information (Ding and Gold, 2010).

Early evidence revealing the neural basis of confidence came from neuropsychology. Patients with Korsakoff syndrome, a disease due to chronic alcohol abuse and malnutrition, have severe amnesia. Shimamura and Squire (1986) found that these subjects have impaired confidence compared to control amnesic patients. Pathological changes in this disease include a decreased volume of the prefrontal cortex and of the thalamus, suggesting that these two areas may be involved in confidence. More recently Lau and Passingham (2006) gathered further evidence in favor of a role of the prefrontal cortex in visual confidence. They measured metabolic activation with functional magnetic resonance imaging (fMRI) in human subjects performing a metacontrast masking task. In metacontrast masking, a figure that overlaps with the contour of the targets is presented after the target (Breitmeyer, 1984). In this task, subjects were prompted to discriminate between a square and a diamond, which appeared only for a short time before the mask. To dissociate performance and confidence the authors changed the stimulus onset interval (SOA), the time between the appearance of the stimulus and the mask. For certain SOAs, subjects had the same level of performance but reported different confidence levels. The authors found that activity in one prefrontal region

(Brodmann area 46, Figure 2) correlated with reports of confidence but not performance. Along the same lines, Del Cul et al. (2009) used visual backward masking to assess visual awareness in patients with focal prefrontal lesions. In backward masking a mask is presented immediately after the presentation of a stimulus and leads to a failure to consciously perceive the stimulus (Breitmeyer, 1984). The delay between the appearance of the stimulus and the mask varied. For short delays, the target is not generally perceived consciously. However, with increasing SOAs, the stimulus becomes visible abruptly, in an all-or-none manner. The authors found higher perceptual thresholds in patients with prefrontal lesions compared to healthy control subjects. The deficit was evident particularly in subjects with lesions of the anterior prefrontal cortex (Figure 2). Also, as in Lau and Passingham (2006), perceptual discrimination performance was little affected.

To test whether the prefrontal cortex plays a causal role in confidence reporting, Rounis et al. (2010) applied transcranial magnetic stimulation to the prefrontal cortex and found that subject's confidence decreased, even though their perceptual discrimination performance remained unchanged with stimulation. Fleming and colleagues (Fleming et al., 2010) were also able to identify anatomical areas related to confidence but not to perceptual discrimination. In their experiment, they instructed subjects to perform a visual discrimination. At the end of every trial, they asked to give a confidence rating about the answer they just provided. Although the difficulty of the trial was adjusted so that all subjects had the same performance, the authors noticed that some subjects had higher confidence sensitivity than others. To investigate the differences in the brains of these different subjects, they analyzed the structural MRIs of the subjects' brains and found that the volume of the gray matter in the anterior prefrontal cortex was larger. Areas in the temporal lobe were also found to be smaller in subjects with high confidence sensitivity compared to subjects with low confidence sensitivity (Figure 2). With a similar design, McCurdy et al., (2013), found that visual confidence, but not task performance, is correlated with the volume of the frontal polar cortical regions (Figure 2). Recently, Hebart and colleagues (Hebart et al., 2014) found an unexpected neural correlate of confidence in the ventral striatum (Figure 2). The subjects in this study had to report the main motion direction in a random dot motion stimulus, and subsequently indicate their degree of confidence. The authors found that activity in the ventral striatum was increased in high confidence trials providing further evidence for a specialized circuit for confidence, since the striatum is not a structure that is classically associated with perceptual decisions.

In an important study in monkeys, Komura and colleagues (Komura et al., 2013) found that the pulvinar may play a role in monitoring confidence. The monkeys were shown a cloud of red and green random moving dots. At the beginning of the trial, a fixation spot appeared that was either red or green indicating which color moving dots to attend. Monkeys reported the perceived direction of motion of the dots of the same color as the fixation spot that started the trial. Confidence was explicitly measured by giving an opt-out choice. The authors found a higher rate of discharge in neurons of the pulvinar for correct and incorrect trials than for trials when the monkey chose to opt out. The authors also reversibly inactivated the pulvinar with the GABA agonist muscimol, and found that after inactivation, monkeys were less confident about their choice and more likely to opt out. The interpretation of these data, however, is complicated by the finding that the inactivation of

the pulvinar also causes hemineglect (Wilke et al., 2010), thus making it difficult to disentangle confidence from attention. Perceptual performance, however, was unaffected after pulvinar inactivation, suggesting that attentional deficits cannot explain the results. Thus, this study suggests that pulvinar may be a monitoring structure, or that it is part of a pathway that helps to facilitate confidence reporting together with the prefrontal cortex, confirming the early findings by Shimamura and Squire (1986) indicating that the thalamus is involved in metacognition.

A recent study implicates a new cortical area in confidence coding. Middlebrooks and Sommer found a neural correlate of confidence in the supplementary eye field (SEF, Middlebrooks and Sommer, 2012). In this experiment, the authors used a betting paradigm, in which monkeys made eye movements to dimly flashed targets. At this stage no reward was given. After a short interval, monkeys chose between a high and a low bet target. If monkeys chose the low bet, they always received a small reward. If monkeys chose the high bet, they received a large reward but only if the answer was correct. If the answer was incorrect, they received nothing. Correct responses were more likely to be associated with high bets than incorrect responses, consistent with the hypothesis that monkeys were aware of not knowing and that they would make choices to indicate the strength of their belief. Neuronal activity in SEF but not the frontal eye field (FEF, area 8) nor the prefrontal cortex (area 46), correlated with monkeys' confidence about their decisions.

In an odor discrimination task, Kepecs and colleagues, (Kepecs et al., 2008) found a neural correlate of confidence in the rodent orbitofrontal cortex (OFC). To assess confidence, the authors gave a delayed reward, while the subjects were free to start another trial if they chose. Rats waited longer if they were confident about their decision. When less confident, they more often aborted a trial and started another one. The main finding was that the neuronal activity in the OFC discriminated between the trials when the rat waited longer for the reward versus those that were aborted quickly. In a following study, Lak et al., (2014) found that inactivation of the ventrolateral OFC during the same odor-discrimination task, disrupted confidence-dependent waiting time, without affecting decision, even in ambiguous situations, showing again a clear dissociation between decision-making and confidence

All the above studies suggest that confidence is implemented in regions that are not commonly considered as part of the decision-making circuitry, evoking the image of a looker inside the brain. Some of these studies, like Lau and Passingham (2006), Del Cul et al., (2009), Rounis et al., (2010) Fleming et al., (2010) McCurdy et al., (2013), Komura et al., (2013) and Lak et al., (2014) show a clear dissociation between performance and confidence, others like, Hebart et al., (2014) Middlebrooks and Sommer (2012) and Kepecs et al., (2008) show correlates of confidence in regions that are not traditionally considered to be involved in decision-making.

Together these results suggest that there are separate and perhaps multiple areas involved in confidence monitoring and reporting. Future studies could be aimed at elucidating how these areas work together to form the circuit involved in monitoring and reporting confidence.

A theoretical foundation of this dualistic view of confidence is provided by the two-stage dynamic signal detection theory (2DSD, Pleskac, 2010). This is a modification of the drift diffusion model popularized in the decision-making literature (Ratcliff, 1978; Ratcliff and McKoon, 2008). The diffusion model posits that a decision is determined by the position of a hypothetical diffusion particle. For two alternative forced-choice decisions, sensory evidence in favor of each choice is accumulated over time. Evidence in favor of the two possible decisions is added together to form a decision variable.

The decision variable is conceived of as a diffusion particle that drifts according to the evidence, toward one or the other boundary favoring one or the other decision (Figure 3). Boundaries are set arbitrarily by the observer as their decision criterion. Once the particle reaches one of the two boundaries, a choice is made (Figure 3A). In the standard model, the accumulation stops when evidence reaches one of the two boundaries (Figure 3A, Ratcliff, 1978). In the 2DSD (Pleskac and Busemeyer, 2010) model, the diffusion continues after a decision is reached and evidence is accumulated to give a confidence rating (Figure 3B).

Although the 2DSD model does not directly address the question of whether decision and confidence are implemented by the same or different brain structures or neurons, it explicitly addresses the dissociation between the perceptual decision (during the first stage) and the determination of confidence (during the second stage). One could in principle modulate this late stage process, thereby changing confidence without affecting the perceptual decision itself and this could explain how it would be possible in experimental paradigms, to dissociate confidence from performance.

Behavioral evidence for a dissociation between performance and metacognition

As mentioned earlier, our confidence in perceptual judgments can be affected by many factors including social pressure (Asch, 1951) and even excluding these factors, introspection is not guaranteed to be accurate (Maniscalco and Lau, 2012). A typical example of failure of both decision-making and metacognition is observed in pathological gamblers. These subjects show a recurrent and maladaptive gambling behavior that disrupts their personal and social life as well as their work (American Psychiatric Association, 2013). Pathological gamblers are impaired in performing the Iowa Gambling task, a test commonly used to study decision-making under uncertainty (Bechara et al., 1994) and in an artificial grammar learning task (see below). Moreover, when asked to evaluate their decisions, pathological gamblers are less accurate than control subjects (Brevers et al., 2013; Brevers et al., 2014).

In some cases, it is also possible to see specific dissociations in subjects who have normal or quasi-normal performance but fail in evaluating their confidence and vice versa. Blindsight and implicit learning are two examples of this. In blindsight, characterized by a lesion of V1 (Weiskrantz, 2009), patients perform above chance when forced to guess about a visual stimulus presented in the damaged visual field, but when asked if they saw something, they report being unaware (Cowey, 2010; Stoerig, 2006; Weiskrantz, 2009). This constitutes a clear dissociation between performance and metacognition, and shows that whereas V1 is not completely necessary for decision-making, the metacognitive system appears blind after its damage. A simple interpretation of these findings in that V1 is directly involved in

metacognition. However, this is unlikely (Crick and Koch, 1995; Ko and Lau, 2012). A more reasonable explanation is that the metacognitive system relies upon information elaborated in the primary visual cortex. When V1 is damaged, its signal is deteriorated and the metacognitive system becomes unable to read it out. According to signal detection theory, the lack of confidence in blindsight patients may be explained by the inability of the metacognitive system to adjust the criterion so that the degraded signal is confused with noise (Ko and Lau, 2012; Lau, 2008).

Implicit learning is learning without awareness and without intention (Reber, 1989). A common paradigm to study implicit learning is artificial grammar learning (AGL). In AGL participants are asked to memorize strings of letters. The sequence of the letters is determined by hidden rules that are not told to the subjects (Reber, 1967; Reber, 1989). In a second phase of the experiment, subjects judge if newly presented strings are grammatical or not. Usually, subjects perform above chance after the learning period; however, most of them are unaware of their acquired knowledge. The inverse dissociation has been recently shown by Scott et al., (2014) who, in an AGL task, selected the participants that did not perform above chance. When asked to rate their confidence, these subjects showed above-chance metacognition even if their performance was at chance. They called this phenomenon blind insight.

Finally, Zylberberg and colleagues (Zylberberg et al., 2012) measured performance accuracy and confidence judgments in a visual motion discrimination task. Subjects reported the main direction of the perceived motion and, following the decision, indicated their degree of confidence. The authors observed that aspects of the stimulus influenced performance and confidence independently. Performance was determined by the number of dots going in the main direction (positive evidence) and in the opposite direction (negative evidence), whereas confidence was mainly influenced by the positive evidence.

These results show that confidence and perceptual performance can be dissociated and indicate that decision and confidence are generated by different specific rules, and therefore may have different neural substrates.

A computational framework for two separate systems: one for performing and one for monitoring, was recently proposed by Cleeramans et al. (2007) and Pasquali et al. (2010) who simulated two recurring networks: a performing network, that is trained to make a simple categorization task, and a monitoring network that judges the decisions of the performing one. The states of the performing network become input to the monitoring network, which may simply report them or place wagers. These networks have shown to be able to mimic the behavior of normal subjects and of patients with blindsight, subjects performing an artificial grammar learning or Iowa gambling task, who have good performance, but poor metacognition.

An alternative view: the shared encoding hypothesis

Although neurological, neuropsychological, fMRI and psychophysical data described above support the idea that confidence circuitry is separate from decision-making circuitry, recent electrophysiological experiments in monkeys suggest that these circuits are shared. Kiani

and Shadlen (2009) recorded from LIP neurons, while monkeys performed a motion-direction decision task. Animals reported their decision about the direction of motion by making an eye movement to one of two possible choice targets, one in the receptive field of the recorded neuron and one outside of the receptive field. LIP neurons are known to increase their discharge rate for saccades toward the receptive field and tend to decrease it for saccades away from the receptive field (Britten et al., 1992). To evaluate the degree of confidence, the authors included a third target, a sure choice located between the two directional choice targets, which yielded a small reward for both correct and incorrect responses. They found that when the monkey chose the sure choice, the activity of the neurons in LIP was reduced, compared to the activity when monkeys chose the target in the response field, but higher than that recorded for choices outside of the response field. The authors concluded that the scaling of LIP neuronal activity with correct, highly confident choices, to less confident choices, indicated that LIP neurons multiplex decision and confidence signals.

In a recent study, Fetsch et al., (2014) studied the effect of electrical microstimulation of area MT on confidence during a motion discrimination task. Monkeys were trained to report the direction of random moving dots while a current was injected in area MT and MST. Consistently with previous findings (Ditterich et al., 2003; Salzman et al., 1990; Salzman et al., 1992), the authors found that the monkey was more likely to choose the preferred direction of stimulated neurons. Like in Kiani and Shadlen (2009) , the monkey had the option to opt out if uncertain. The authors found that electrical stimulation increased confidence in the preferred direction and reduced it when dots were moving in the opposite direction. Although these findings suggest a common encoding of perceptual and confidence signals, the fact that performance and confidence were affected in the same direction makes these findings difficult to interpret. In other words, if the monkey perceived a stronger movement during stimulation, then the increase of confidence may have been a mere consequence of the altered perception.

Like area LIP, the superior colliculus, in the midbrain, contains neurons that discharge in relationship to the generation of eye movements and have preferred eye movement response fields (see Wurtz and Goldberg, 1972 for a review). In a recent study in the SC, trained monkeys performed an odd-ball selection task in which one stimulus was differently colored from three others. Trained monkeys made saccades to the stimulus that was differently colored. Neuronal recordings were made from four superior colliculus neurons simultaneously, each neuron encoding one of the locations containing the four stimuli. Using signal detection theory, Kim and Basso (2008), found that neuronal activity across the population was statistically distinguishable when monkeys' performance was accurate. When choice performance was poor, the discriminability of the neuronal activity was also reduced. Interestingly, the more separable neural activity patterns were in the superior colliculus, the more accurate the performance was. Although the authors did not explicitly measure confidence, this result indicates that activity in the superior colliculus may be encoding confidence together with decision signals.

A theoretical explanation of how confidence is encoded by the same neurons involved in decision-making is supported by the currently popular, Bayesian views of the brain (Friston,

2012; Lau, 2008; Pouget et al., 2013). Bayes theorem is a way to quantify uncertainty and is formally stated as:

$$P(a|b) = P(b|a)\,P(b)\,/\,P(a)$$

where *P(a|b)* is the conditional probability of event a occurring given the occurrence of event b, also called the *posterior*. *P(b|a)* is the conditional probability of observing event b given event a. This is also known as the *likelihood*. *P(b)* is the probability of event b also referred to as the prior. *P(a)* is a normalization term and for explanatory purposes can be ignored. Thus, Bayes theorem simplifies to:

$$P(a|b) \sim P(b|a)\,P(b)$$

and means simply, that the measure of uncertainty or the posterior, is proportional to the product of the likelihood and the prior.

Thinking about the brain in Bayesian terms is somewhat intuitive. Neurons, particularly those in sensory and motor areas have tuning curves; that is they can be described as radially symmetric functions of a stimulus parameter. Neurons show maximal discharge for optimal stimuli or movements and discharge less with stimulus or movement parameters that are less than optimal (Chalupa, 2003). Tuning curves are essentially, likelihood functions. They are a measures of the probability of a particular outcome given a particular discharge rate (Foldiak, 1993; Jazayeri and Movshon, 2006; Sanger, 2002, 2003). Recent theoretical work (Ma et al., 2006), supported by the previously described experimental studies (Beck et al., 2008; Kim and Basso, 2010) shows that populations of neurons representing the likelihood and the prior, can be combined linearly in much the same way as Bayes' theorem combines two probability distributions, to provide a read-out of a decision in the form a posterior distribution. Specifically, as shown in Figure 4, the x coordinate of the peak of the posterior is determined by the x value of the peak of both the prior and of the likelihood; the width of the distribution of the posterior is determined by the width of by the prior and the likelihood. When the influence of the prior is reduced, the influence of the likelihood is dominant.

A critical feature of this kind of an approach to understanding decision-making is that confidence or uncertainty is encoded implicitly across the population response or the posterior. Let us assume that the value on the x coordinate in Figure 4 corresponds to a saccade in a certain direction. In this case, the x position of the peak of the posterior will determine the direction of the saccade, and the width of the posterior distribution will determine the confidence in that decision. When the posterior distribution across all possible choices is narrow, the confidence will be high because most of the influence will be given to one choice and little influence will be given to the other choices. In contrast, when the posterior distribution is wide, a more even distribution of influence will be given to more of the choices, resulting in less confidence in any one choice. Thus, in this scheme, confidence is encoded implicitly in the width of the population distribution of activity (Beck et al., 2008; Kim and Basso, 2010).

This kind of view, that perceptual decisions are computed in terms of Bayesian probability distributions in the brain, is one motivation for believing that confidence does not depend on specialized circuitry. If perceptual decisions are already computed in such probabilistic terms, confidence information should already be present in the circuits for decision-making. However, one concern is whether such information in the superior colliculus or LIP for example, can be read-out by those structures themselves. Even if the information is there, it is still possible that a monitoring module reads out the width of the distribution of the posterior.

Conclusions and future directions

In this review, we described experimental evidence supporting the hypothesis that confidence is encoded in regions of the brain that also encode decision information such as area LIP, and the superior colliculus. We also reviewed evidence supporting the opposite conclusion, namely, that confidence is performed by specialized monitoring circuits including the prefrontal cortex, specifically the polar regions of the prefrontal cortex and the orbitofrontal cortex, the SEF, the ventral striatum and the pulvinar. In terms of computational models, it is fair to say that the current dominant view supports the latter scheme. Many authors believe - or at least consider it an exciting possibility - that perceptual decisions are determined based on probabilistic information. If explicit representations of probabilities are already present in the circuits for perceptual decision-making, it is intuitive to assume that confidence is also determined within the same circuit. Yet, how do we explain the suggestion from lesion, fMRI and psychophysical evidence, that there may be separate structures uniquely supporting confidence?

Behavioral and neurological dissociations are strong indicators that the two functions are performed by separate systems. If confidence and perceptual decision-making rely on completely identical circuits, it would be hard to explain the results of studies that successfully demonstrate behavioral dissociations between confidence and decision performance. In addition, importantly, the fact that circuits for decision-making carry information about confidence does not mean that these circuits are the sites where confidence is ultimately generated. Confidence information may be generated in the prefrontal cortex and then fed back to area LIP, the superior colliculus, the pulvinar or more widely. Alternatively, and more plausibly, there may be a shared early signal source for both perceptual decisions and confidence judgments, but as soon as these signals reach upstream areas in prefrontal cortex, confidence and decision information may diverge.

How then, can we explain the apparent differences in conclusions by the different studies? Why do some physiological studies suggest there are shared circuits for confidence and decision-making? First, as noted above, the findings are compatible with feedback signals from specialized regions. Second, since the data suggesting a shared circuit for confidence and decision-making come from studies in which decision-related areas were assessed, it is not surprising that confidence signals were found. Electrophysiological studies in behaving animals in particular, are subject to sampling biases for a number of reasons. For example, only neurons that have evoked responses to the stimuli used in the experiment are studied simply because they are the ones that can most easily be recorded. If confidence signals are

encoded in neurons that do not show a simple relationship to visual or movement parameters (Mante et al., 2013), they might be easily overlooked.

A third explanation for the discrepant results is that in most tasks, confidence and decision performance accuracy are confounded. Given that the two variables are correlated behaviorally, we should expect to find shared neuronal signatures.

In the endeavor to understand the neuronal mechanism of confidence, we consider important to manipulate, with specific tasks, the confidence system independently from decision-making (Lau and Passingham, 2006; Zylberberg et al., 2012). This involves more careful design of the experimental stimuli, and application of more rigorous tools for psychophysical modeling.

Effort should also be directed to identifying brain regions in animals that are involved in confidence but not in decision, as has been done in humans (Del Cul et al., 2009; Fleming et al., 2010; Lau and Passingham, 2006). Valuable help may come from functional imaging, which is becoming available for use in non-human animals (Logothetis et al., 1999; Vanduffel et al., 2001). Such an approach might constrain the search space for areas related to confidence monitoring and reporting. Imaging studies in humans as well as non-human animals can guide electrophysiological studies in animal models. Translating these approaches to electrophysiological experiments we believe will be at the root of understanding how the brain encodes confidence.

Aknowledgements

The work performed in the laboratories if Lau and Basso. This work is supported by funding to HL from the Templeton Foundation (grant nos. 21569 and 15462), and NIH R01 NS088628-01 and to MAB, NIH EY13692

References

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th Edn.. American Psychiatric Publishing; Arlington,VA.: 2013.

Aristotle. De Alma. Penguin books; London: 1987.

Asch, S. Groups, leadership and men. Carnegie Press; Pittsburgh, PA: 1951. Effects of group pressure on the modification and distortion of judgments.; p. 177-190.

Bechara A, Damasio AR, Damasio H, Anderson SW. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition. 1994; 50:7–15. [PubMed: 8039375]

Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, Pouget A. Probabilistic population codes for Bayesian decision making. Neuron. 2008; 60:1142–1152. [PubMed: 19109917]

Bransford, J,D.; Brown, A,L.; Cocking, R,R. How People Learn: Brain, Mind, Experience, and School. National Academies Press; Washington, DC.: 2005.

Breitmeyer, BG. Visual Masking: An Integrative Approach. Oxford University Press; Oxford: 1984.

Brevers D, Bechara A, Cleeremans A, Noel X. Iowa Gambling Task (IGT): twenty years after - gambling disorder and IGT. Front. Psychol. 2013; 4:665. [PubMed: 24137138]

Brevers D, Cleeremans A, Bechara A, Greisen M, Kornreich C, Verbanck P, Noel X. Impaired metacognitive capacities in individuals with problem gambling. J. Gambl. Stud. 2014; 30:141–152. [PubMed: 23149513]

Britten KH, Shadlen MN, Newsome WT, Movshon JA. The analysis of visual motion: a comparison of neuronal and psychophysical performance. J. Neurosci. 1992; 12:4745–4765. [PubMed: 1464765]

Chalupa, L,M.; Werner, J,S. Vis. Neurosci. MIT Press; Cambridge: 2003.

Cheesman J, Merikle PM. Distinguishing conscious from unconscious perceptual processes. Can J. Psychol. 1986; 40:343–367. [PubMed: 3502878]

Cleeremans A, Timmermans B, Pasquali A. Consciousness and metarepresentation: a computational sketch. Neural Netw. 2007; 20:1032–1039. [PubMed: 17904799]

Cowey A. The blindsight saga. Exp. Brain. Res. 2010; 200:3–24. [PubMed: 19568736]

Crick F, Koch C. Are we aware of neural activity in primary visual cortex? Nature. 1995; 375:121–123. [PubMed: 7753166]

Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A. Causal role of prefrontal cortex in the threshold for access to consciousness. Brain. 2009; 132:2531–2540. [PubMed: 19433438]

Descartes, R. Discourse on method. Penguin Books; London: 1999.

Dienes Z, Perner J. A theory of implicit and explicit knowledge. Behav. Brain. Sci. 1999; 22:735–755. discussion 755-808. [PubMed: 11301570]

Ding L, Gold JI. Caudate encodes multiple computations for perceptual decisions. J. Neurosci. 2010; 30:15747–15759. [PubMed: 21106814]

Ditterich J, Mazurek ME, Shadlen MN. Microstimulation of visual cortex affects the speed of perceptual decisions. Nat. Neurosci. 2003; 6:891–898. [PubMed: 12858179]

Fetsch CR, Kiani R, Newsome WT, Shadlen MN. Effects of cortical microstimulation on confidence in a perceptual decision. Neuron. 2014; 83:797–804. [PubMed: 25123306]

Fleming SM, Dolan RJ. Effects of loss aversion on post-decision wagering: implications for measures of awareness. Conscious. Cogn. 2010; 19:352–363. [PubMed: 20005133]

Fleming SM, Dolan RJ. The neural basis of metacognitive ability. Phil. Trans. R. Soc. B. 2012; 367:1338–1349. [PubMed: 22492751]

Fleming SM, Lau HC. How to measure metacognition. Front. Hum. Neurosci. 2014; 8:443. [PubMed: 25076880]

Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. Relating introspective accuracy to individual differences in brain structure. Science. 2010; 329:1541–1543. [PubMed: 20847276]

Fodor, J. The Modularity of Mind: An Essay on Faculty Psychology. MIT Press; Cambridge, MA.: 1983.

Foldiak, P. The “ideal homonculus”: statistical inference from neural population responses. In: Eeckman, F.; B.J., editors. Computation and neural systems. Kluwer Academic; Norwell, MA: 1993. p. 55-60.

Foote AL, Crystal JD. Metacognition in the rat. Curr. Biol. 2007; 17:551–555. [PubMed: 17346969]

Friston K. The history of the future of the Bayesian brain. NeuroImage. 2012; 62:1230–1233. [PubMed: 22023743]

Galvin SJ, Podd JV, Drga V, Whitmore J. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. Psychon. Bull. Rev. 2003; 10:843–876. [PubMed: 15000533]

Gold JI, Shadlen MN. Representation of a perceptual decision in developing oculomotor commands. Nature. 2000; 404:390–394. [PubMed: 10746726]

Hampton RR. Rhesus monkeys know when they remember. Proc. Natl. Acad. Sci. USA. 2001; 98:5359–5362. [PubMed: 11274360]

Hart JT. Memory and the feeling-of-knowing experience. J. Educ. Psychol. 1965; 56:208–216. [PubMed: 5825050]

Hebart MN, Schriever Y, Donner TH, Haynes JD. The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. Cereb. cortex. 2014 Epub ahead of print.

Horwitz GD, Newsome WT. Separate signals for target selection and movement specification in the superior colliculus. Science. 1999; 284:1158–1161. [PubMed: 10325224]

Horwitz GD, Newsome WT. Target selection for saccadic eye movements: direction-selective visual responses in the superior colliculus. J. Neurophysiol. 2001; 86:2527–2542. [PubMed: 11698540]

Humphrey, NK. The mind made flesh: Essays from the frontiers of psychology and evolution, The uses of consciousness. Oxford University Press; New York: 2003. p. 65-85.

James, W. The Principles of Psychology. Harvard University Press; Cambridge, MA.: 1983.

Jazayeri M, Movshon JA. Optimal representation of sensory information by neural populations. Nat. Neurosci. 2006; 9:690–696. [PubMed: 16617339]

Kepecs A, Uchida N, Zariwala HA, Mainen ZF. Neural correlates, computation and behavioural impact of decision confidence. Nature. 2008; 455:227–231. [PubMed: 18690210]

Kiani R, Shadlen MN. Representation of confidence associated with a decision by neurons in the parietal cortex. Science. 2009; 324:759–764. [PubMed: 19423820]

Kim B, Basso MA. Saccade target selection in the superior colliculus: a signal detection theory approach. J. Neurosci. 2008; 28:2991–3007. [PubMed: 18354003]

Kim B, Basso MA. A probabilistic strategy for understanding action selection. J. Neurosci. 2010; 30:2340–2355. [PubMed: 20147560]

Kim JN, Shadlen MN. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. Nat. Neurosci. 1999; 2:176–185. [PubMed: 10195203]

Ko Y, Lau H. A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. Proc. R. Soc. B. 2012; 367:1401–1411.

Komura Y, Nikkuni A, Hirashima N, Uetake T, Miyamoto A. Responses of pulvinar neurons reflect a subject's confidence in visual categorization. Nat. Neurosci. 2013; 16:749–755. [PubMed: 23666179]

Kornell N, Son LK, Terrace HS. Transfer of metacognitive skills and hint seeking in monkeys. Psychol. Sci. 2007; 18:64–71. [PubMed: 17362380]

Lak A, Costa GM, Romberg E, Koulakov AA, Mainen ZF, Kepecs A. Orbitofrontal cortex is required for optimal waiting based on decision confidence. Neuron. 2014; 84:190–201. [PubMed: 25242219]

Lau H, Rosenthal D. Empirical support for higher-order theories of conscious awareness. Trends Cogn. Sci. 2011; 15:365–373. [PubMed: 21737339]

Lau HC. A higher order Bayesian decision theory of consciousness. Prog. Brain Res. 2008; 168:35–48. [PubMed: 18166384]

Lau HC, Passingham RE. Relative blindsight in normal observers and the neural correlate of visual consciousness. Proc. Natl. Acad. Sci. USA. 2006; 103:18763–18768. [PubMed: 17124173]

Logothetis NK, Guggenberger H, Peled S, Pauls J. Functional imaging of the monkey brain. Nat. Neurosci. 1999; 2:555–562. [PubMed: 10448221]

Luttrell AB,P, Pettya R,E, Cunningham W, Díazd D. Metacognitive confidence: A neuroscience approach. Int. J. Soc. Psychol. 2013; 28:317–332.

Ma WJ, Beck JM, Latham PE, Pouget A. Bayesian inference with probabilistic population codes. Nat. Neurosci. 2006; 9:1432–1438. [PubMed: 17057707]

Maniscalco B, Lau H. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious. Cogn. 2012; 21:422–430. [PubMed: 22071269]

Mante V, Sussillo D, Shenoy KV, Newsome WT. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature. 2013; 503:78–84. [PubMed: 24201281]

McCurdy LY, Maniscalco B, Metcalfe J, Liu KY, de Lange FP, Lau H. Anatomical coupling between distinct metacognitive systems for memory and visual perception. J. Neurosci. 2013; 33:1897–1906. [PubMed: 23365229]

Metcalfe J. Feeling of knowing in memory and problem solving. J. Exp. Psychol.-Learn. Mem. Cogn. 1986; 12:288–294.

Middlebrooks PG, Sommer MA. Neuronal correlates of metacognition in primate frontal cortex. Neuron. 2012; 75:517–530. [PubMed: 22884334]

Pasquali A, Timmermans B, Cleeremans A. Know thyself: metacognitive networks and measures of consciousness. Cognition. 2010; 117:182–190. [PubMed: 20825936]

Persaud N, McLeod P, Cowey A. Post-decision wagering objectively measures awareness. Nat. Neurosci. 2007; 10:257–261. [PubMed: 17237774]

Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. Nature. 1999; 400:233–238. [PubMed: 10421364]

Pleskac T,J, Busemeyer J,R. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. Psychol. Rev. 2010; 117:864–901. [PubMed: 20658856]

Pouget A, Beck JM, Ma WJ, Latham PE. Probabilistic brains: knowns and unknowns. Nat. Neurosci. 2013; 16:1170–1178. [PubMed: 23955561]

Ramsøy TZ, Overgaard M. Introspection and subliminal perception. Penomenol. Cogn. Sci. 2004; 3(1):1–23., 1-23.

Ratcliff R. A theory of memory retrieval. Psychol. Rev. 1978; 85:59–108.

Ratcliff R, McKoon G. The diffusion decision model: theory and data for two-choice decision tasks. Neural. Comput. 2008; 20:873–922. [PubMed: 18085991]

Reber A. Implicit learning of artificial grammars. J. Verb. Learn. Verb. Beh. 1967; 6:855–863.

Reber AS. Implicit learning and tactic knowledge. J. Exp. Psychol. Gen. 1989; 118:219–235.

Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. 2002; 22:9475–9489. [PubMed: 12417672]

Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. Cogn. Neurosci. 2010; 1:165–175. [PubMed: 24168333]

Ruffman T, Garnham W, Import A, Connolly D. Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. J. Exp. Child. Psychol. 2001; 80:201–224. [PubMed: 11583523]

Salzman CD, Britten KH, Newsome WT. Cortical microstimulation influences perceptual judgements of motion direction. Nature. 1990; 346:174–177. [PubMed: 2366872]

Salzman CD, Murasugi CM, Britten KH, Newsome WT. Microstimulation in visual area MT: effects on direction discrimination performance. J. Neurosci. 1992; 12:2331–2355. [PubMed: 1607944]

Sandberg K, Timmermans B, Overgaard M, Cleeremans A. Measuring consciousness: is one measure better than the other? Conscious. Cogn. 2010; 19:1069–1078. [PubMed: 20133167]

Sanger TD. Decoding neural spike trains: calculating the probability that a spike train and an external signal are related. J. Neurophysiol. 2002; 87:1659–1663. [PubMed: 11877538]

Sanger TD. Neural population codes. Curr. Opin. Neurobiol. 2003; 13:238–249. [PubMed: 12744980]

Scherberger H, Andersen RA. Target selection signals for arm reaching in the posterior parietal cortex. J. Neurosci. 2007; 27:2001–2012. [PubMed: 17314296]

Scott RB, Dienes Z, Barrett AB, Bor D, Seth AK. Blind insight: metacognitive discrimination despite chance task performance. Psychol. Sci. 2014; 25:2199–2208. [PubMed: 25384551]

Sergent C, Dehaene S. Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. Psychol. Sci. 2004; 15:720–728. [PubMed: 15482443]

Shields WE, Smith JD, Washburn DA. Uncertain responses by humans and rhesus monkeys (Macaca mulatta) in a psychophysical same-different task. J. Exp. Psychol.-Gen. 1997; 126:147–164. [PubMed: 9163934]

Shimamura AP, Squire LR. Memory and metamemory: a study of the feeling-of-knowing phenomenon in amnesic patients. J. Exp. Psychol.-Learn. Mem. Cogn. 1986; 12:452–460.

Son, L.; Kornell, N. Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In: Terrace, H,S.; Metcalfe, J., editors. The missing link in cognition: Origins of self-reflective consciousness. Oxford University Press; New York: 2005.

Stoerig P. Blindsight, conscious vision, and the role of primary visual cortex. Prog. Brain. Res. 2006; 155:217–234. [PubMed: 17027389]

Terrace, HS.; Metcalfe, J. The Missing Link in Cognition: Origins of Self-Reflective Consciousness. OXFORD UNIVERSITY PRESS; Oxford: 2005.

Tulving E, Madigan SA. Memory and verbal learning. Annu. Rev. Psychol. 1970; 21:437–484.

Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RB, Orban GA. Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. Neuron. 2001; 32:565–577. [PubMed: 11719199]

Weiskrantz, L. Blindsight. Oxford University Press; Oxford: 2009.

Wierzchon M, Asanowicz D, Paulewicz B, Cleeremans A. Subjective measures of consciousness in artificial grammar learning task. Conscious. Cogn. 2012; 21:1141–1153. [PubMed: 22728143]

Wilke M, Turchi J, Smith K, Mishkin M, Leopold DA. Pulvinar inactivation disrupts selection of movement plans. J. Neurosci. 2010; 30:8650–8659. [PubMed: 20573910]

Wurtz RH, Goldberg ME. The role of the superior colliculus in visually-evoked eye movements. Bibliotheca ophthalmologica : supplementa ad ophthalmologica. 1972; 82:149–158. [PubMed: 4631287]

Zylberberg A, Barttfeld P, Sigman M. The construction of confidence in a perceptual decision. Front. Integr. Neurosci. 2012; 6:79. [PubMed: 23049504]

Highlights

- Here we review the literature about visual confidence in humans and animals
- The traditional view is that confidence is encoded by the same areas for decision
- Recent findings show that confidence may be achieved by dedicated structures
- In light new evidence, we provide an explanation for these conflicting results

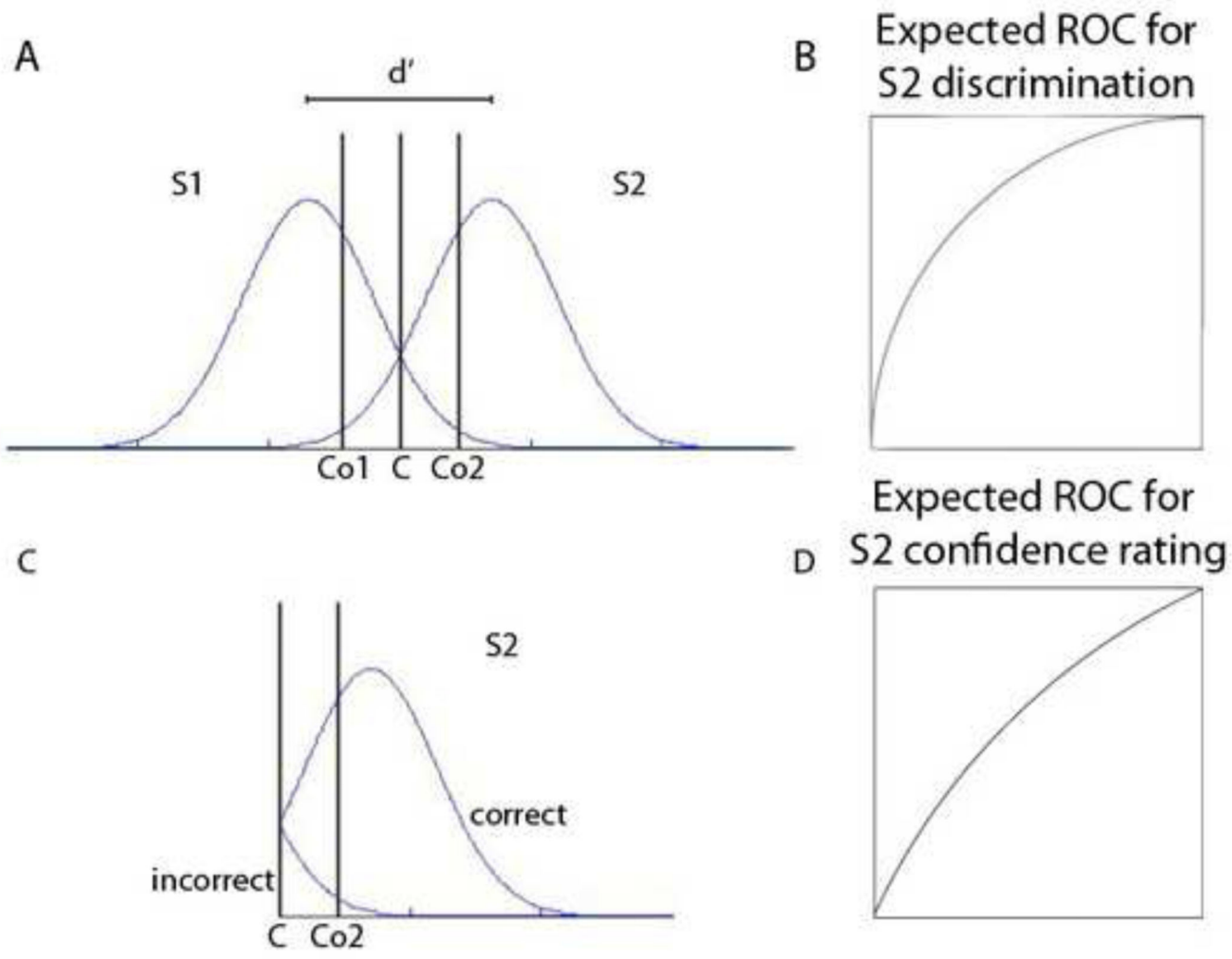

Figure1.

Signal detection theory for performance (A, B) and confidence (C, D). A. Distribution of internal responses for two stimuli, S1 and S2. The observer sets a criterion C such as all responses that are higher than C are considered as belonging to S2, all those that are lower than C are considered S1. Figure B shows the ROC curve generated by distribution in A. The observer also sets additional criteria Co1 and Co2, such that the responses that are lower than Co1 and higher than Co2 are endorsed with high confidence. Figure C shows only the part of the graph that surpasses C. Figure D: ROC curve generated by swiping Co2 in Figure C. Modified from (Maniscalco and Lau, 2012).

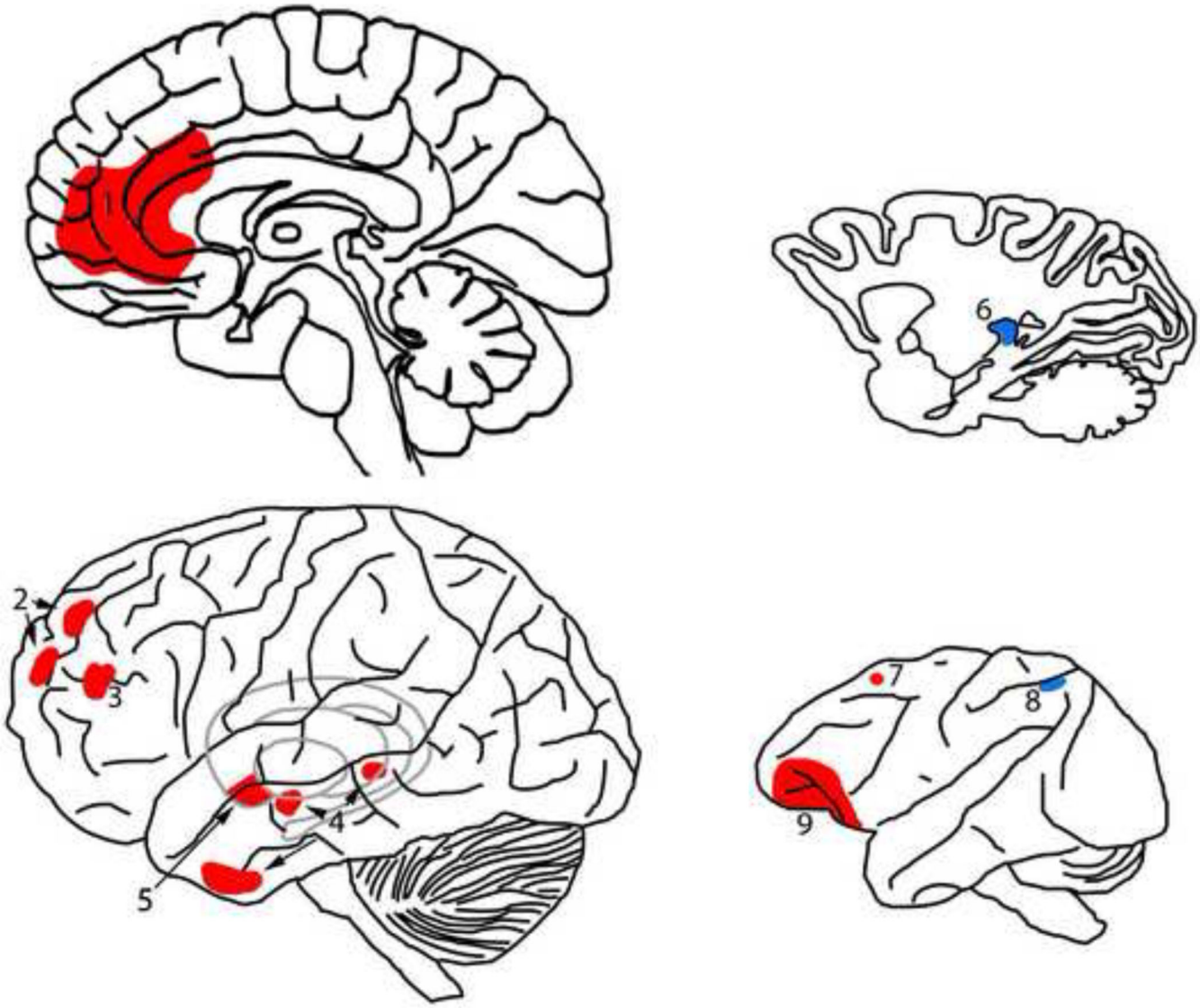

Figure 2.

Map of the areas involved in confidence in the human (left; basal ganglia in grey) and monkey brain (right). 1: antero-medial prefrontal cortex (Del Cul et al., 2009), 2, 4: anterior prefrontal cortex and temporal lobe (Fleming et al., 2010), 3: Brodmann area 46 (Lau and Passingham, 2006), 5: ventral striatum (Hebart et al., 2014), 6: pulvinar (Komura et al., 2013), 7: SEF (Middlebrooks and Sommer, 2012), 8: LIP (Kiani and Shadlen, 2009). 9: OFC. Areas in red are those that are solely involved in confidence. Areas in blue are those involved in both confidence and decision-making.

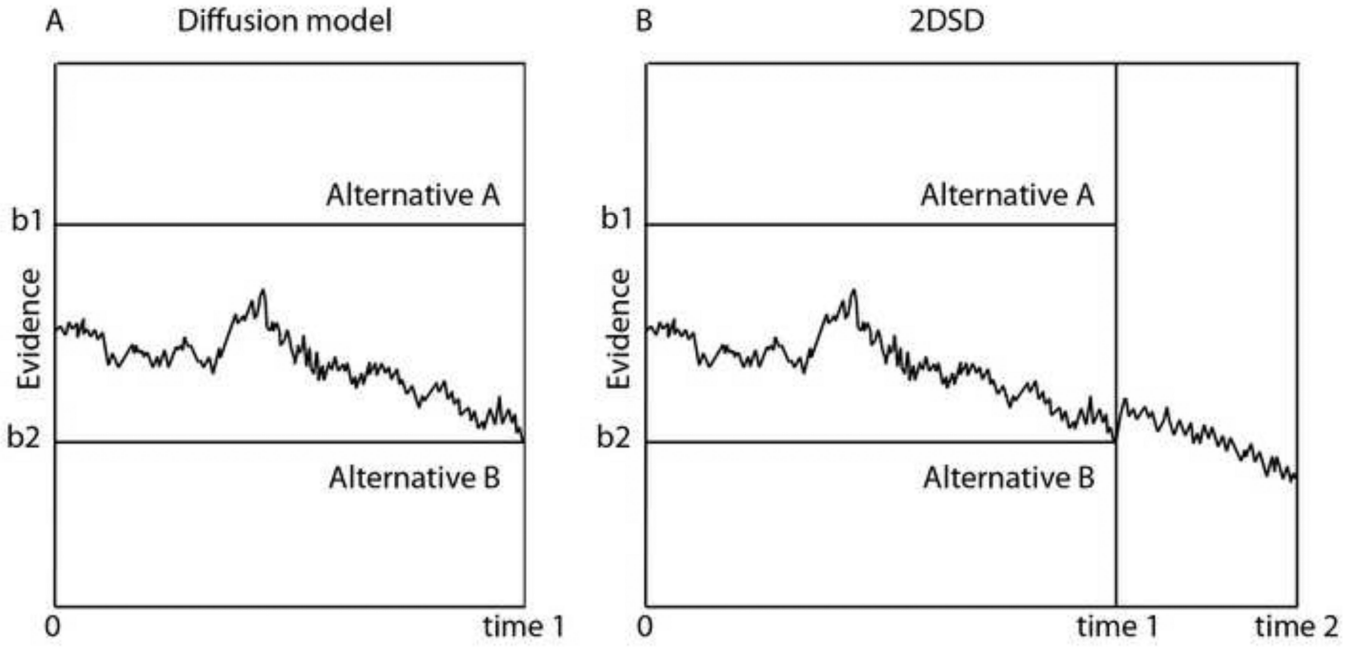

Figure 3.

Schematic representation of the diffusion (A) and 2DSD model (B). A. Diffusion model: The black jagged line depicts the accumulation process when the observer chooses alternative B. The boundaries (b1 and b2) are arbitrarily set by the observer as a decision criterion. B. 2DSD model. In this model the accumulation continues after the decision is taken to give a confidence judgment (time 2). Modified from (Pleskac and Busemeyer, 2010).

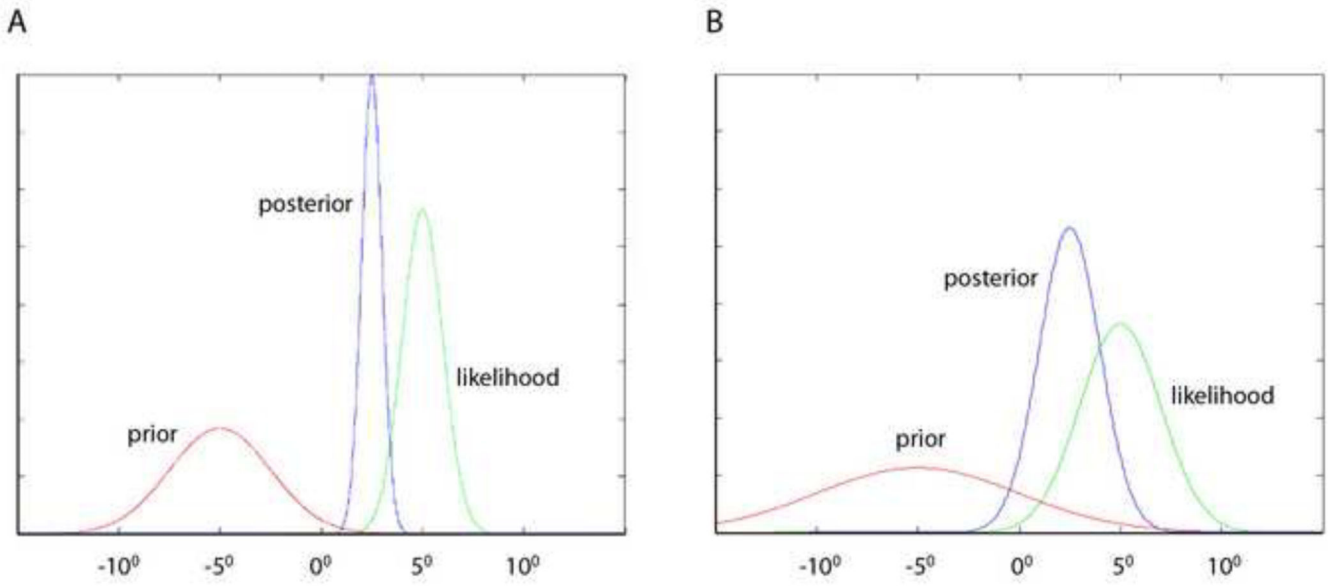

Figure 4.

Two examples of how the distribution of the prior and the likelihood can influence the posterior in Bayesian statistics.