Visual Confidence

Pascal Mamassian[1,2]

[1]Laboratoire des Systèmes Perceptifs, CNRS UMR 8248, 75005 Paris, France

[2]Institut d'Etude de la Cognition, Ecole Normale Supérieure, PSL Research University, 75005 Paris, France; email: pascal.mamassian@ens.fr

Annu. Rev. Vis. Sci. 2016. 2:459–81

First published online as a Review in Advance on August 3, 2016

The *Annual Review of Vision Science* is online at vision.annualreviews.org

This article's doi: 10.1146/annurev-vision-111815-114630

Keywords

visual confidence, metaperception, overconfidence, Type 2 ROC, signal detection theory, accumulation of evidence models, decision making

Abstract

Visual confidence refers to an observer's ability to judge the accuracy of her perceptual decisions. Even though confidence judgments have been recorded since the early days of psychophysics, only recently have they been recognized as essential for a deeper understanding of visual perception. The reluctance to study visual confidence may have come in part from obtaining convincing experimental evidence in favor of metacognitive abilities rather than just perceptual sensitivity. Some effort has thus been dedicated to offer different experimental paradigms to study visual confidence in humans and nonhuman animals. To understand the origins of confidence judgments, investigators have developed two competing frameworks. The approach based on signal decision theory is popular but fails to account for response times. In contrast, the approach based on accumulation of evidence models naturally includes the dynamics of perceptual decisions. These models can explain a range of results, including the apparently paradoxical dissociation between performance and confidence that is sometimes observed.

1. INTRODUCTION

Imagine that you want to cross a road at a busy intersection (Tiwari et al. 2007). Relevant sources of visual information include the speed of the incoming vehicles and the size of the gaps between vehicles. However, to decide whether it is safe to cross the road, one cannot just infer the values of these speed and distance variables. To properly judge whether it is safe to cross the road, one should also estimate how trustworthy one's inferences are about speed and distance.

Our ability to doubt our knowledge is fundamental for our very existence (Descartes 1637). In perception, this ability helps us evaluate the quality of our percepts. Rather than blindly believing that our perceptions are always veridical, we feel that we are able to critically judge when our perceptions are more likely to be correct. The ability to estimate the accuracy of our visual decisions is known as visual confidence. Being a judgment on a judgment, visual confidence belongs to the domain of metacognition (Metcalfe & Shimamura 1994) and is thus also known as metaperception.

1.1. Scope of the Review

Proper visual confidence is by no means a trivial feat because multiple sources of uncertainty interfere with the retinal information at multiple stages of perceptual processing. Therefore, to achieve good visual confidence, observers likely need access to an enriched representation of the decision variables underlying the perceptual performance, above and beyond the typical binary values of a perceptual task (e.g., stimulus present versus absent, motion to the right or left). If the observer does not have access to such an enriched representation of the perceptual decision, then she should have at least some partial knowledge of the various sources of uncertainty that constrain perceptual performance in order to predict the outcome of her perceptual decision. Two different potential sources of confidence evidence are discussed when reviewing the two main contemporary frameworks of confidence below.

What is confidence good for? In daily life, visual confidence is important for transport safety, medical diagnosis, and witness reliability. If confidence is available and well calibrated, low confidence is an important signal to allocate more resources to a task and hopefully improve task performance through perceptual learning. But, clearly, visual confidence is not necessary to survive evolutionary pressure, as not all animals seem to have it. For instance, it is hard to argue that small creatures such as the fruit fly (*Drosophila*) are able to monitor their own performance, notwithstanding their ability to distinguish whether changes in the environment are caused by their own actions (Paulk et al. 2015). It is thus important to remain critical of the criteria used to assess whether an organism is showing evidence of visual confidence, an issue that is central to the comparative cognition literature briefly introduced below.

Though correlated, accuracy and visual confidence can be dissociated. In the case of blindsight resulting from lesions in the primary visual cortex, patients perform perceptual decisions above chance while claiming not being able to do the task (Sahraie et al. 1998). This is an extreme case of underconfidence, in which confidence judgments seriously underestimate perceptual performance. There have been attempts to simulate the phenomenon of blindsight in normal observers (Baker & Cass 2013, Kolb & Braun 1995), but the experimental methods used in these studies have not convinced all researchers (Morgan et al. 1997). The phenomenon of blindsight itself is controversial (Cowey 2010), but it provides a logical case for separating confidence from accuracy. Discussions of visual confidence are inevitably relevant to the issue of visual awareness (Kunimoto et al. 2001, Persaud et al. 2007), but this topic is beyond the scope of the present review.

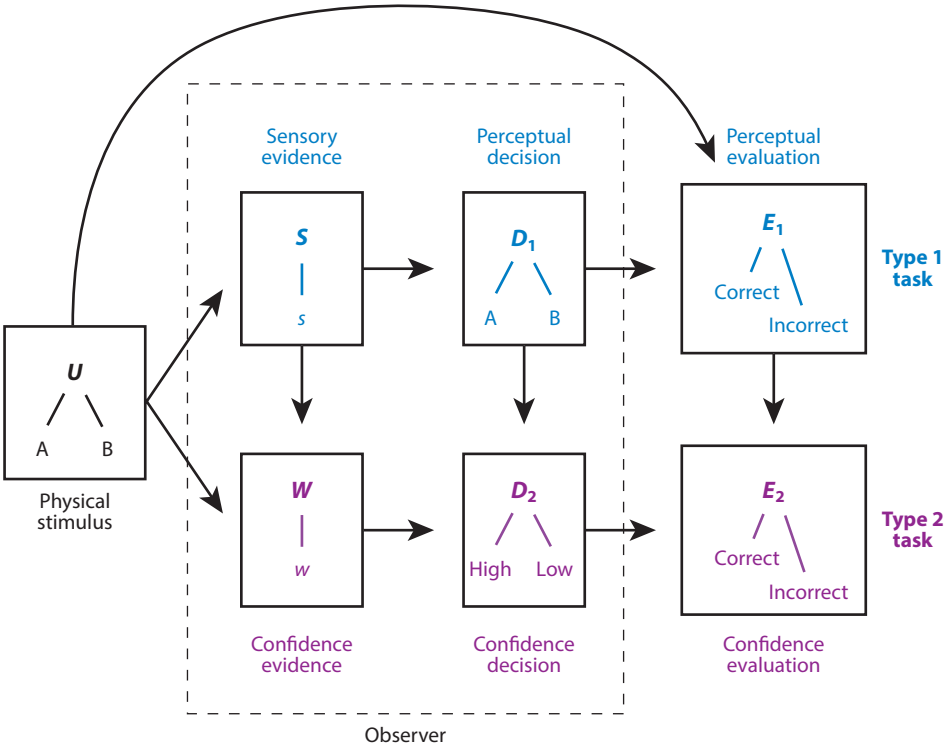


Figure 1

Difference between Type 1 and Type 2 tasks. In this example, the physical stimulus U belongs to either class A or class B. In Type 1 tasks, the stimulus gives rise to noisy sensory evidence s whose origin an observer has to decide is more likely attributable to either class A or B. Such a Type 1 perceptual decision D_1 can then be evaluated to be correct or incorrect by an external agent (E_1 process). In Type 2 tasks, confidence evidence w is generated from the sensory evidence, from the original physical stimulus, or both. Confidence is the observer's own estimate of the external perceptual evaluation E_1. This confidence is based on both perceptual decision and confidence evidence. A Type 2 confidence decision D_2 can then be generated from this confidence variable, for instance, by dividing it into high and low levels of confidence. Random variables are denoted by capital letters, and their realizations are in lowercase.

1.2. Defining Confidence Judgment

In an attempt to properly define visual confidence and differentiate it from accuracy, I examine below a typical visual decision and describe the current frameworks to study visual confidence. Briefly, the accuracy of a perceptual decision refers to the ability of the observer to make a good inference on the visual stimulus. In contrast, visual confidence refers to the ability of the observer to make a good inference on the validity of the response corresponding to this perceptual decision.

Let us illustrate the problem of estimating visual confidence with an observer who is engaged in a binary discrimination task, where physical stimulus U belongs to class A or class B (**Figure 1**). Examples of perceptual discriminations include various decisions associated with an introduced stimulus, for example, determining whether its direction of motion is to the left or to the right, whether its color is red or green, and whether its depth is far or near. By contrast, if the task were to detect a stimulus, the two classes could be "stimulus absent" and "stimulus present." Through

her sensory system, an observer has access to some sensory evidence s that is extracted from the physical stimulus but that is corrupted by internal noise (Juslin & Olsson 1997). The task of the observer is to infer the class of the original stimulus, in other words, to apply some decision rule D_1 to categorize the sensory evidence. Once the observer reports her perceptual decision, an outside agent (typically, an experimenter) can evaluate (E_1 process) the correctness of this decision by comparing it to the physical stimulus. Type 1 task is the term used to describe this perceptual chain of events, distinguishing this process from the one underlying confidence judgments (Type 2 task).

The aim of the Type 2 task that underlies the confidence judgment is to estimate the validity of the perceptual decision, that is, to infer the value of the E_1 process. To be as general as possible, we allow the Type 2 task to follow a mechanism that differs from the Type 1 task. It starts with some confidence evidence w that presumably is strongly related to the sensory evidence, but not necessarily identical. Section 2.4 gives an example of such a relationship between sensory and confidence evidence. The confidence evidence may also contain some information about the stimulus that is not part of the sensory evidence. For instance, while judging the color of an object, whether the stimulus is small or big may be largely ignored for the sensory evidence, but it may be regarded as a relevant dimension for the confidence evidence. In addition to the stimulus size, the observer might be tempted to extract its contrast, apparent amount of noise, or perceived duration because large size, high contrast, low noise, and long duration are usually associated with good performance. Here confidence is defined as the probability that the perceptual evaluation is correct given the perceptual decision and the confidence evidence:

$$\text{Confidence} = \mathrm{P}\left(E_1 = \text{correct} \mid \ d_1,\ w\right). \tag{1}$$

Confidence decision D_2 is based on this confidence variable, typically by dividing it into high and low levels of confidence. For instance, the observer will aim to assign high confidence judgments only if she feels that her perceptual decision is likely to be correct with a probability larger than 0.75. An observer that is able to divide her confidence in agreement with her performance is said to be well calibrated (Fleming & Lau 2014).

Clarke et al. (1959) are credited for the useful distinction between Type 1 and Type 2 tasks. In the above example of the Type 1 task, an observer decides whether stimulus A or B was presented. Which event was actually presented has been chosen independently of the observer. When an observer decides between A and B, another event occurs: The observer is either correct or incorrect in her choice. The Type 2 task prompts an observer to judge her confidence in the accuracy of her Type 1 decision. Therefore, in contrast to Type 1, the Type 2 task refers to a judgment dependent on the observer. In **Figure 1**, the distinction between these two different decisions on the correctness of the perceptual judgment is the difference between E_1 and D_2 processes.

Here, visual confidence is defined as an observer's evaluation of her performance on a basic visual task based on some internal confidence evidence. Two consequences result from this definition. First, explicit reference to Type 1 performance is important. As such, confidence defined here is distinct from the ability to evaluate the quality of the internal representation of a sensory stimulation (Pouget et al. 2016). In particular, confidence is not just the opposite of the uncertainty of an internal representation (Meyniel et al. 2015). Second, even if sensory and confidence evidence turn out to be coded similarly in the brain, they are fundamentally different variables. This distinction is important for designing experimental paradigms, which should avoid conflating Type 1 and Type 2 decisions. In particular, all Type 2 decisions that can be reduced to (more complex) Type 1 decisions should be evaluated with caution.

Section 2 begins with a description of the main theoretical frameworks that have been proposed to investigate confidence. Sections 3 and 4 then review the experimental paradigms that have been used in humans and nonhuman animals as well as some results associated with these paradigms.

2. THE SIGNAL DETECTION THEORY FRAMEWORK

Since the early days of signal decision theory (SDT), confidence judgments have been recorded alongside discrimination decisions. In the pioneering work by Peirce & Jastrow (1885), participants were asked not only to make a perceptual decision, but also to report their confidence in this decision. In a task to discriminate between two pressures applied to their finger, participants rated their confidence on a four-point scale. At one end of the scale, "0 denoted absence of any preference for one answer over its opposite, so that it seemed nonsensical to answer at all"; at the other end of the scale, "3 denoted as strong a confidence as one would have about such sensations" (Peirce & Jastrow 1885, p. 77). Confidence ratings correlated so well with pressure discriminations that the authors proposed the following formula relating m the degree of confidence on the scale and p the probability of the answer being right:

$$m = c \log \frac{p}{1-p}, \tag{2}$$

where c is a constant called the index of confidence. More recent efforts to directly relate accuracy and confidence have relied on correlation measures, such as the standard Pearson r correlation. However, these measures are prone to undesirable effects, in particular due to confidence response biases (Fleming & Lau 2014). This is a serious source of concern because rating scales are often interpreted differently by different participants, some too often using one end of the scale and thus displaying over- or underconfident behaviors (Morgan et al. 1997). For this reason, contemporary studies of visual confidence avoid correlation measures and rely on a more detailed analysis, which are reviewed now.

Before describing the variables representing the confidence (Type 2) judgment, I first briefly review the perceptual (Type 1) task. For both perceptual and confidence judgments, decisions are taken by looking at the evidence in favor of one or the other hypothesis along a decision axis. The reader less familiar with elements of SDT is referred to classical textbooks on this topic (Green & Swets 1966, Macmillan & Creelman 2005).

2.1. Sensory Evidence

For Type 1 tasks, sensory evidence may be the information provided by a sensory organ, and the sensory decision axis S may indicate the range of outputs this sensory organ can generate. Uncertainty in obtaining sensory evidence s given that stimulus U was displayed is represented by a probability density function on the sensory decision axis. For instance, when stimulus A is displayed, Type 1 evidence s has a likelihood of occurrence equal to

$$\mathrm{P}(S = s \,|\, U = \mathrm{A}). \tag{3}$$

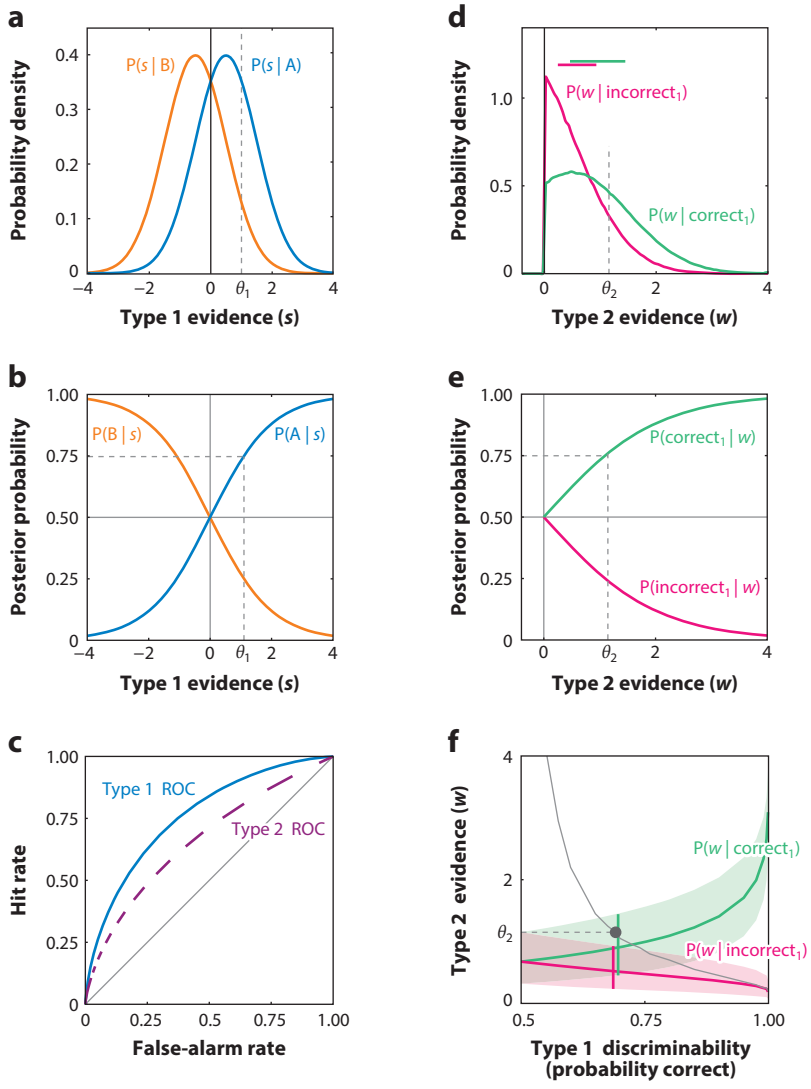
The concepts defined in this section are better understood via an explicit example. We assume that the likelihoods of stimuli A and B follow normal distributions with opposite means ($\mu_\mathrm{A} = 0.5; \mu_\mathrm{B} = -0.5$) but equal unit variance ($\sigma^2_\mathrm{A} = \sigma^2_\mathrm{B} = 1$). These likelihood functions are shown in **Figure 2*a***, where the separation between these two distributions indicates the sensitivity to discriminate A from B (here, $d' = 1$).

2.2. Perceptual Decision

The task of the observer is to infer which stimulus was displayed given some sensory evidence, that is, to infer the posterior probability $P(U = A|S = s)$. If the prior probabilities with which A and B are displayed, $P(U = A)$ and $P(U = B)$, are known, then Bayes' rule may be used to estimate posterior probability from the likelihood in Equation 3:

$$P(U = A|\ S = s) = k \cdot P(S = s|U = A) \cdot P(U = A), \tag{4}$$

where k is a normalizing constant such that $P(A|s) + P(B|s) = 1$ (omitting equal signs inside the equations when there is no ambiguity). If we assume that A and B are equally likely to be displayed [$P(U = A) = P(U = B) = 0.5$], the posterior probability functions are the ones shown in **Figure 2*b***. We can see that the posterior probability to decide between A and B deviates from chance level (0.5) as Type 1 evidence s is more distant from the origin.



To indicate whether A or B was displayed, an observer places a Type 1 criterion θ_1 along the sensory decision axis. The Type 1 decision D_1 will then be in favor of hypothesis A if evidence s exceeds this criterion, that is,

$$D_1(s) = \text{A if } (s > \theta_1); \text{ otherwise, } D_1(s) = \text{B}. \tag{5}$$

In this illustrative example, as it is clear from looking at the posterior probabilities in **Figure 2*b*** the optimal Type 1 criterion is located at the origin, exactly in between the two likelihoods, that is, $\theta_1 = 0$. However, the observer may place her criterion wherever she wants (see, e.g., **Figure 2*a***).

2.3. Perceptual Evaluation

Through evaluation process E_1, an agent other than the observer (e.g., the experimenter) can determine the validity of the perceptual decision (see **Figure 1**). To evaluate the performance of the observer who adopts decision rule D_1, we can compute the Type 1 hit and false-alarm (FA) rates as

$$\text{Hit}_1 = \text{P}(D_1 = A | U = A) \quad \text{and} \quad \text{FA}_1 = \text{P}(D_1 = A | U = B). \tag{6}$$

Correct_1 (correct Type 1 decision) occurs when A (or B) is displayed and A (or B) is chosen; the remaining two cases contribute to Incorrect_1 (incorrect Type 1 decision). Therefore, the proportion of correct Type 1 decisions is

$$\begin{aligned} \text{PC}_1 &= \text{P}(\text{Correct}_1) = \text{P}(D_1 = \text{A}, U = \text{A}) + \text{P}(D_1 = \text{B}, U = \text{B}) \\ &= \text{Hit}_1 \cdot \text{P}(U = \text{A}) + \text{FA}_1 \cdot \text{P}(U = \text{B}). \end{aligned} \tag{7}$$

Importantly, the proportion of correct responses is based on a decision rule that in turn depends on the Type 1 criterion. When the criterion is placed at its optimal location ($\theta_1 = 0$), performance is

Figure 2

Computing confidence within a signal detection theory framework. (*a*) Two classes of signals, A (*blue*) and B (*orange*), have to be discriminated. The observer chooses A whenever the sensory (Type 1) evidence on one trial exceeds a criterion represented here by the dashed vertical bar (the optimal criterion is at the origin). The distance between the two distributions determines how easily A can be discriminated from B. (*b*) If we know a priori how often A and B occur, we can compute the posterior probability that A or B was displayed given a particular sensory evidence. If the sensory evidence exceeds the criterion shown in the plot, the probability that A was displayed exceeds 0.75. (*c*) Receiver operating characteristic (ROC) curves are obtained by varying the location of the criterion and plotting the corresponding hit and false-alarm rates. The Type 1 ROC based on sensory evidence is shown as a blue solid line, and the Type 2 ROC based on confidence evidence appears as a purple dashed line. (*d*) The computation of confidence relies on some measure of confidence (Type 2) evidence. Here, the Type 2 evidence is chosen to be the distance of the sensory evidence to the Type 1 criterion. One can then determine the probability of obtaining a value of confidence evidence given that the Type 1 decision was correct (*green*) or incorrect (*red*). Here and in the following plots, the Type 1 criterion was placed at its optimal location. The horizontal green and red lines at the top indicate the 25th and 75th percentiles of the confidence evidence distributions. (*e*) If we know a priori the probability of Type 1 correct decisions, we can compute the posterior probability of being correct or incorrect given a particular confidence evidence. If the confidence evidence exceeds the Type 2 criterion shown in the plot (*vertical dotted line*), the probability of being correct in the sensory decision exceeds 0.75. (*f*) The distributions of confidence evidence for correct and incorrect responses are more separated for better sensory discriminability. All the previous plots were obtained for a particular sensory discriminability, a Type 1 probability correct equal to 0.69 and shown here by the vertical lines. When sensory (Type 1) discriminability increases, the separability of the confidence evidence for correct and incorrect responses improves, and the Type 2 criterion corresponding to 75% of correct responses decreases (*gray solid line*).

maximal (here, the proportion correct is 0.69). All other criteria reduce performance all the way to chance level for criteria far away from the optimal location. To obtain a measure of discriminability that is criterion-free, one often refers to the receiver operating characteristic (ROC). The ROC curve is obtained by plotting the Type 1 hit rate against the false-alarm rate for all possible criteria (**Figure 2*c***). The area under the ROC curve (AUC) is a criterion-free measure of sensitivity just like the more familiar d'. The AUC has a maximal value of 1 and in the illustrative example AUC = 0.76.

2.4. Confidence Evidence

For Type 2 tasks, an observer's evidence for being correct in her perceptual decision varies along a confidence decision axis W. For instance, the distance between sensory evidence and the Type 1 criterion can serve as confidence decision axis W:

$$W = |S - \theta_1|. \tag{8}$$

In the above illustrative example where the Type 1 criterion was placed at its optimal location ($\theta_1 = 0$), confidence evidence reduces to the absolute value of sensory evidence.

It is then important to distinguish confidence evidence when the observer was correct or incorrect. For instance, the likelihood of obtaining confidence evidence w when the observer was correct is

$$\mathrm{P}(W = w|\mathrm{Correct}_1). \tag{9}$$

The likelihoods of confidence evidence for correct and incorrect decisions are shown in **Figure 2*d*** and were first detailed in Clarke et al. (1959). These functions are far from being normally distributed, which prevents use of common sensitivity measures such as d' to characterize Type 2 sensitivity. However, we can still run a Type 2 ROC analysis (Clarke et al. 1959, Fleming et al. 2010).

In this review, we assume that the confidence evidence is the distance of the sensory evidence to the Type 1 criterion (Equation 8). Even though this is a popular choice, there are other possibilities. For instance, Galvin et al. (2003) derived the posterior probabilities of confidence when the Type 2 evidence is the likelihood ratio of the sensory evidence given that the perceptual evidence was correct or incorrect (see also de Gardelle & Mamassian 2014).

2.5. Confidence Decision

From the point of view of the observer, confidence evidence w is generated on each trial, and the observer has to use this information to infer the probability that her perceptual decision was correct. In other words, the observer is attempting to compute the posterior probability of confidence $\mathrm{P}(\mathrm{Correct}_1|W = w)$. As with Type 1, this posterior probability can be computed from the likelihood in Equation 9:

$$\begin{aligned} \mathrm{P}(\mathrm{Correct}_1|W = w) &= k \cdot \mathrm{P}(W = w|\mathrm{Correct}_1) \cdot PC_1, \\ \mathrm{P}(\mathrm{Incorrect}_1|W = w) &= k \cdot \mathrm{P}(W = w|\mathrm{Incorrect}_1) \cdot (1 - PC_1), \end{aligned} \tag{10}$$

where k is a normalizing constant such that $\mathrm{P}(\mathrm{Correct}_1|w) + \mathrm{P}(\mathrm{Incorrect}_1|w) = 1$. Assuming for now that the proportion of correct decisions PC_1 is known, then the posterior probability of confidence equals that shown in **Figure 2*e***. Of course, this assumption is not insignificant, because knowing PC_1 is nothing less than knowing one's confidence on average (i.e., not specifically on the current trial, but on average if one was allowed to run the same trial multiple times). Although

this assumption is unrealistic in real life, it is still somewhat reasonable in a typical laboratory experiment where the experimenter has arranged that performance falls in-between chance and ceiling levels.

To decide between high confidence (HC) and low confidence (LC) that the Type 1 decision was correct, the observer places a Type 2 criterion θ_2 along the confidence decision axis (**Figure 2*d***). The Type 2 decision D_2 will then favor the high-confidence hypothesis if evidence w exceeds this criterion, that is,

$$D_2(w) = \text{HC if } (w > \theta_2), \text{ and } D_2(w) = \text{LC otherwise.} \tag{11}$$

Where should one place the Type 2 criterion? To answer this question, we can refer back to the posterior probability of confidence (**Figure 2*e***). Placing the criterion too close to the origin increases the risk of assigning high confidence for incorrect Type 1 decisions, that is, to be overconfident. In contrast, large criterion values correspond to risk-aversive behaviors in which high confidence judgments are assigned only when the expected probability of being correct is high.

2.6. Confidence Evaluation

To estimate the Type 2 performance of the observer who adopts decision rule D_2, we can compute the Type 2 hit and false-alarm rates as

$$\text{Hit}_2 = \text{P}(D_2 = \text{HC}|\text{Correct}_1) \quad \text{and} \quad \text{FA}_2 = \text{P}(D_2 = \text{HC}|\text{Incorrect}_1). \tag{12}$$

Discriminability between correct and incorrect Type 1 decisions can be characterized by the Type 2 ROC obtained by plotting the Type 2 hit rate against false-alarm rate for all possible criteria (**Figure 2*c***). Importantly, this Type 2 ROC is not unique; it varies when the Type 1 criterion varies (Galvin et al. 2003). In addition, the Type 2 AUC is typically less than that of Type 1 (here, the Type 2 AUC = 0.66), so it is hazardous to directly compare Type 1 and Type 2 ROCs.

2.7. Definition of Confidence

As defined in Section 1, confidence is an observer's estimate of the probability that her perceptual decision for the current trial is correct given her perceptual decision and current confidence evidence (Equation 1). In the example provided, the Type 1 decision was assumed to be taken deterministically given sensory evidence (see Equation 5). Likewise, confidence evidence was deterministically determined from sensory evidence (see Equation 8). Finally, per the illustrative example provided, scenarios in which A and B are displayed are completely symmetric to each other, thus leading to a simplified definition of confidence:

$$\text{Confidence} = \text{P}(\text{Correct}_1|w). \tag{13}$$

Equation 13 actually matches the posterior probability of confidence that we saw in Equation 10 (see **Figure 2*e***). As expected, the posterior probability of being correct increases as the Type 2 evidence w increases, thereby validating the choice of the variable for confidence evidence in Equation 8.

2.8. Confidence and Perceptual Discriminability

Confidence performance is constrained by perceptual performance. Intuitively, if the perceptual sensitivity is very poor, the observer cannot be expected to be good at discriminating high- and

low-confidence trials. The illustrative example considers a single difficulty level for sensory discrimination. The separation between the two distributions of sensory evidence (in units of the standard deviation of the distribution) determines the Type 1 accuracy (here a d' of 1.0 or equivalently 69% correct; **Figure 2*a***). For this difficulty level, the distributions of confidence evidence overlap strongly (**Figure 2*d***), thereby forcing the observer to set a large Type 2 criterion to secure that high confidence ratings are associated to high sensory performance (**Figure 2*e***). For easier sensory discriminations, obtained, for instance, by increasing the separation between A and B signals, the overlap between the confidence evidence distribution decreases, and the critical Type 2 criterion also decreases (**Figure 2*f***). This observation has an important consequence when the observer does not know a priori the difficulty of the discrimination she is facing, a situation that occurs, for instance, when different difficulty levels are intermixed within a block of trials. If the observer sets a unique Type 2 criterion for all difficulty levels, then easy trials are often rated with low confidence and hard trials with high confidence. This strategy might be responsible for the hard-easy effect discussed below (e.g., Zylberberg et al. 2014; see Section 4.5).

2.9. Confidence and Perceptual Bias

As noted above, different Type 2 ROC curves are obtained for different Type 1 criteria (Galvin et al. 2003). It is problematic if researchers are tempted to use Type 2 AUC as a measure of an observer's sensitivity to differentiate between high- and low-confidence trials. A similar issue arises if one is tempted to compute a Type 2 d' (Barrett et al. 2013). To circumvent this issue, some authors have proposed the use of an alternative measure (Maniscalco & Lau 2012, Barrett et al. 2013). Even though the Type 2 ROC varies with the Type 1 criterion, the Type 2 AUC presents a single maximum value for a given perceptual sensitivity d'. One can then search for an equivalent perceptual sensitivity that matches the measured Type 2 performance. This equivalent perceptual sensitivity is called meta-d'. If meta-d' equals the actual perceptual sensitivity d', then an observer is fully efficient at making Type 2 judgments and is limited only by her perceptual sensitivity to discriminate high- and low-confidence trials. Even though some researchers find it intuitively obvious that confidence should be independent of the Type 1 criterion, such a strong postulate for a theory of confidence probably demands further investigation.

2.10. Confidence and Response Time

Because SDT models assume sensory evidence arises from a single sample s of the stimulus, the perceptual decision is reached within a fixed amount of time. In particular, these models predict that stimulus difficulty has no impact on response time. Therefore, an inverse relationship of the one typically found between stimulus difficulty and confidence (e.g., Baranski & Petrusic 1994; see Section 4.7) is not straightforward to achieve.

3. THE ACCUMULATION OF EVIDENCE FRAMEWORK

In contrast to models inspired by SDT, accumulation of evidence models assume that new sensory evidence becomes available until a decision is reached (Yeung & Summerfield 2012). New evidence reinforces or sometimes challenges past evidence so that waiting longer helps an observer reach a better decision. Two main stopping rules can be used to trigger a decision. In a diffusion-to-bound model, a decision is made when the accumulation process crosses a predefined bound. In contrast, in a time-limited model, a decision is made at a particular time or after a certain number of accumulation steps have taken place.

Accumulation of evidence models embrace multiple variations that can be likened to two main classes. In drift diffusion models (DDMs), a single variable encodes the accumulation of evidence in favor of one or the other possible stimulus categories (Bogacz et al. 2006, Ratcliff & McKoon 2008). For instance, if two stimulus categories are equally likely to occur, a diffusion process will aim to reach a positive bound for one category and a negative bound of equal absolute value for the other category. Which bound is crossed determines the perceptual decision.

The other main class of accumulation models is the independent race model (IRM), which have an equal number of accumulation variables and stimulus categories (Raab 1962). In an IRM, a single bound applies to all diffusion processes (or races), and the race that first crosses the bound determines the perceptual decision. Even though an assumption of independence for the different diffusion processes is very common, some amount of correlation is possible (Moreno-Bote 2010). In fact, when two races are perfectly anticorrelated, or when evidence in favor of one stimulus category is also evidence against the other category, the model reduces to a DDM. The DDM and IRM can thus be seen as two extremes on a continuum of partially correlated accumulator models. For this reason, the IRM is emphasized in the following discussion because it is a bit more general than the DDM, although at least some partial correlation in the real world is likely between the races.

In seminal works using an IRM, Vickers and colleagues proposed a model of confidence based on a balance of evidence (BoE) at the time of perceptual decision (Vickers 1979, Vickers & Packer 1982, Smith & Vickers 1988). The critical feature of their model is that confidence compares the decision made with the alternative option declined.

3.1. Sensory Evidence and Decision

Using the binary discrimination task employed in Section 2 for SDT, I here illustrate the IRM for confidence. Presenting stimulus U to an observer initiates two simultaneous races representing evidence in favor of hypothesis A and hypothesis B (**Figure 3*a***). For instance, if stimulus A is presented, then evidence accumulates at each time step by a small amount that is drawn from a normal distribution with larger mean for A than for B (here $\mu_A = 0.32$ and $\mu_B = 0.25$) and equal unit variance ($\sigma_A^2 = \sigma_B^2 = 1$). Following a diffusion-to-bound model, the first race that crosses a predetermined bound wins (to match Type 1 performance with the performance obtained in the SDT example above, we set bound $= 30$). Even though the winning race determines the Type 1 decision, the losing race is also important in the IRM. At the time of the decision, the distance of the evidence of the losing race to the bound defines the so-called BoE (Vickers 1979).

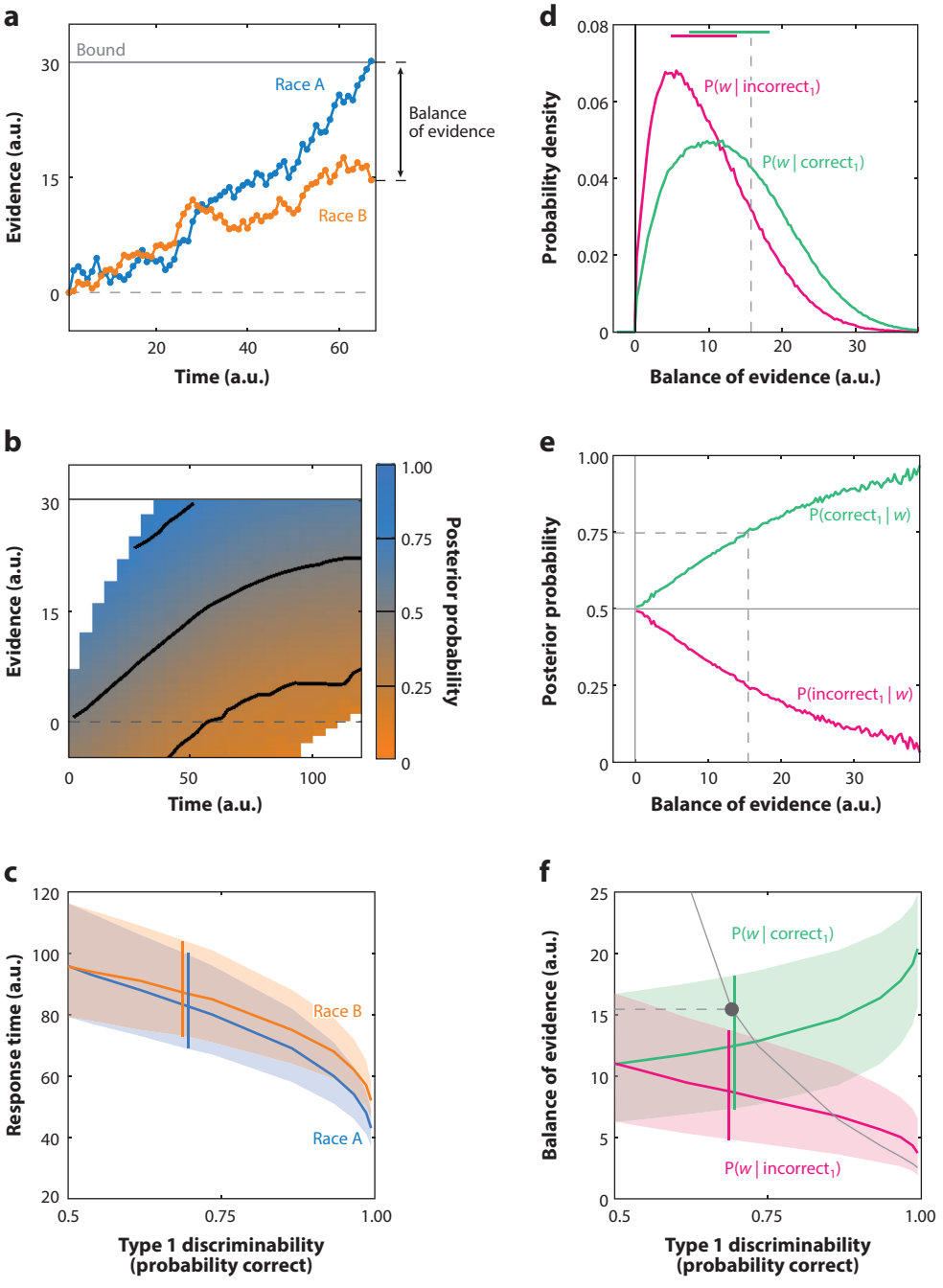
The robustness of the perceptual decision can be evaluated across multiple trials. While simulating multiple trials where stimulus A is presented, the number of times that evidence value e is reached at time t may be registered separately for A and B races. Upon renormalization of each pair (e, t) (similarly to Equation 4), posterior probability $P(U = A|E = e, T = t)$ is obtained to infer that stimulus A was presented given current evidence and time (**Figure 3*b***).

At the bound (focusing on the top border of **Figure 3*b***), the posterior probability is

$$P(U = A|E = \text{bound}, T = rt), \tag{14}$$

and it represents the probability that the stimulus A was presented given that the response time was rt. The posterior probability of inferring that A was presented when A was actually presented reflects the accuracy of Type 1 decisions. In our simulations, this accuracy varies little with the response time. In contrast, if we had adopted a time-limited stopping rule (focusing at the right-hand border of **Figure 3*b***), the evidence at the time set by the stopping rule correlates strongly

with the sensory posterior: A larger value of the evidence corresponds to a larger probability that stimulus A was presented. The isoposterior probability curves in **Figure 3*b*** are monotonically increasing, illustrating a trade-off between speed and evidence. Identical perceptual accuracy can be achieved by small evidence accumulated quickly or larger evidence accumulated more slowly.



3.2. Accuracy and Response Time

Distributions of response times when A or B wins the race are illustrated in **Figure 3*c***. An important property of response times in accumulation models is that the mean response time decreases with accuracy, a phenomenon that is almost universally observed (e.g. Baranski & Petrusic 1994, Kiani et al. 2014). Examining one particular accuracy level more finely (taking a vertical slice of **Figure 3*c***), we can see that response time distributions are very similar for correct and incorrect responses. This property runs against behavioral evidence in which response times for incorrect responses are often longer than those for correct responses. However, this lack of difference in response times can easily be corrected by adding a variability of the start of the accumulation process (Laming 1968).

3.3. Confidence Evidence and Decision

Following the proposal of Vickers (1979), the evidence accumulated by the losing race at the time of the perceptual decision appears relevant when estimating the quality of the decision. If the losing race is almost at the level of the bound, then it could well have been the winning race if the last noise increment were more favorable. In contrast, if the losing race is very far from the bound, it is safe to accept the perceptual decision. The BoE can thus characterize Type 2 confidence evidence:

$$W = \text{BoE}. \tag{15}$$

As done within the SDT framework, it is useful to distinguish confidence evidence when an observer is correct or incorrect. For instance, the likelihood of obtaining confidence evidence w when an observer is correct is again provided by Equation 9. The likelihoods of confidence evidence for correct and incorrect decisions are shown in **Figure 3*d***.

An observer is interested in inferring whether her perceptual decision is correct given confidence evidence w. In other words, the observer is attempting to compute the posterior probability of confidence $\text{P}(\text{Correct}_1|W = w)$, which can be done using Equation 10. The posterior probability of confidence shown in **Figure 3*e*** assume the proportion of correct decisions PC_1 is known. As

Figure 3

Computing confidence within an accumulation of evidence model. (*a*) Two races for hypotheses A and B accumulate sensory evidence over time. The first race to cross the bound wins and sets the perceptual decision (here A). The distance of the losing race to the bound defines the balance of evidence (BoE). (*b*) If we know a priori how often A and B occur, we can compute the posterior probability that A or B was displayed for each pair of sensory evidence and time. One million trials were simulated when A was always the stimulus, interrupting the simulation if one of the races crossed the bound. Isoposterior probability curves (*black contour lines*) are monotonically increasing, illustrating the trade-off between speed and evidence to achieve the same level of perceptual accuracy. Cells in this plot containing less than 0.1% of simulated trials are omitted. (*c*) Response times decrease with increasing sensory discriminability. Here A was always the displayed stimulus. Bold lines show the mean, and the area the range between 25th and 75th percentiles, of the response time distributions for A decisions (*blue*, correct) and B decisions (*orange*, incorrect). (*d*) The BoE can serve as confidence (Type 2) evidence, and one can determine the probability of responding with a certain BoE level, given that the sensory (Type 1) decision was correct (*green*) or incorrect (*red*). All response times were pooled to obtain this plot. (*e*) Taking into account the probability of correct and incorrect decisions, one can compute the posterior probability of being correct or incorrect given a particular BoE. If the confidence evidence exceeds the Type 2 criterion shown in the plot (*vertical dotted line*), the probability of being correct in the sensory decision exceeds 0.75. (*f*) The distributions of confidence evidence for correct and incorrect responses differ more when the sensory discriminability is higher. When sensory (Type 1) discriminability increases, the separability of the confidence evidence for correct and incorrect responses improves, and the Type 2 criterion corresponding to 75% of correct responses decreases (*gray solid line*). Abbreviation: a.u., arbitrary units.

indicated, the posterior probability of being correct increases as Type 2 evidence w also increases, thereby validating the choice of BoE for confidence evidence (Equation 15).

3.4. Confidence and Perceptual Discriminability

Within the signal detection theory framework, as discussed above, the distributions of confidence for correct and incorrect responses became more separable for better Type 1 discriminability (**Figure 2*f***). A similar phenomenon can be observed for accumulation of evidence models (**Figure 3*f***). Consequently, when sensory discriminability increases, smaller confidence (Type 2) criteria are needed to secure that at least 75% of perceptual decisions are correct.

Behavioral studies typically find a negative relationship between the response time and confidence (e.g., Baranski & Petrusic 1994, Kiani et al. 2014). Longer response times are associated with lower confidence. This negative relationship is also found between the response time and BoE, although this relationship is not very reliable (Drugowitsch et al. 2014).

3.5. Variations of the Definition of Confidence Evidence

In DDMs, the BoE cannot be computed because there is a single accumulation process, and thus there is no losing race. However, similar reasoning can be applied if confidence evidence is taken to be the current position of the accumulated evidence. In a time-limited task, one can thus define confidence evidence as the proximity of the accumulated evidence to the bound at the time of the decision (Kiani & Shadlen 2009).

Other proposals have been offered to compute confidence in DDMs. For instance, the accumulation process can continue after a perceptual decision has been made, and the ultimate location of accumulated evidence serves as a proxy for confidence evidence (Pleskac & Busemeyer 2010). If confidence evidence keeps accumulating after the perceptual decision has been taken, the confidence may end up reaching a different bound than the perceptual one, and if given the opportunity, the observer may change her mind (Rabbitt 1966, Resulaj et al. 2009, van den Berg et al. 2016). In yet another model, multiple DDMs are started for different confidence levels, and the DDM that wins determines confidence (Ratcliff & Starns 2013).

The extent to which both races in an accumulation are independent is still a debated issue. For instance, anticorrelation of neural activity in and out of the receptive field of neurons in cortical area LIP (Roitman & Shadlen 2002, Churchland et al. 2008) provides an argument against IRMs. On the other hand, the negative correlation between response times and the discriminability of both the selected and unselected stimuli in a two-alternative forced choice task is an argument against DDMs (Liston & Stone 2013). Irrespective of the conclusion of this issue, all accumulation of evidence models attempt to characterize all three basic components of a perceptual task—accuracy, decision time, and confidence. As such, these models are good competitors of models based on SDT, which say nothing about decision time.

4. HUMAN STUDIES

Before reviewing some results related to meta-perception in humans, it is important to first discuss the methods used to obtain these results.

4.1. Confidence Ratings

Following the early work of Peirce & Jastrow (1885), researchers commonly measure confidence by asking participants to use a rating scale. Confidence ratings have since been used regularly in

SDT experiments to allow an experimenter to build a Type 1 ROC curve (Green & Swets 1966) (see **Figure 2*c***). As such, it is pertinent to ask whether confidence ratings really tap into Type 2 judgments or instead are just Type 1 judgments in disguise. Indeed, confidence ratings prompt observers to place multiple criteria along their sensory decision axis (e.g., Aitchison et al. 2015), thereby showing that they can perform a fine-grained Type 1 discrimination, but not that they can judge the correctness of their perceptual decision. Another source of concern is the idiosyncratic way participants may use the rating scale (Morgan et al. 1997). For instance, some participants may be optimistic and use only the high end, whereas others may try to use all confidence levels uniformly.

In spite of these concerns, confidence ratings have several advantages, not least of which are how easy they are to collect after a perceptual judgment and the lack of difficulty participants usually have in understanding their task (e.g., Zizlsperger et al. 2014). Confidence ratings also appear easy to analyze, as experimenters can simply compute the correlation between Type 1 and Type 2 performance (Nelson 1984). However, this analysis is not immune to biases (Masson & Rotello 2009, Fleming & Lau 2014). Confidence ratings can be restricted to just two levels, low and high confidence, and the ability to use these two levels properly is sometimes used as a bias-free measure of awareness (Kunimoto et al. 2001). Unfortunately, later work also showed that this measure is itself not immune to biases (Evans & Azzopardi 2007), unless Type 2 performance in near chance and the experimenter is only interested in thresholds for awareness (Galvin et al. 2003).

A variation of the confidence ratings paradigm is to combine perceptual decision with rating task (e.g., Kiani et al. 2014, Aitchison et al. 2015). For instance, in a binary discrimination between stimuli A and B, an observer can be given four choices: (1) The stimulus is clearly A, (2) the stimulus is probably A, (3) the stimulus is probably B, and (4) the stimulus is clearly B. The advantage of this paradigm is a gain of experimental time. However, the task can now be interpreted as a four-alternative forced choice, where the internal space for sensory evidence is divided into four regions (i.e., the observer places three Type 1 criteria). In other words, this paradigm exacerbates the issue raised for confidence ratings, namely that the task can be seen to measure only Type 1 performance and not properly measure the observer's estimate of her Type 1 performance. For this reason, it is sometimes called a pseudo–Type 1 ratings paradigm (Galvin et al. 2003).

4.2. Opt-Out Paradigms

Another paradigm to measure confidence is to allow participants to opt-out of performing the perceptual decision if they feel too uncertain (Gherman & Philiastides 2015, Kiani & Shadlen 2009). In other words, an observer is engaged in a perceptual decision, but she can decide not to commit to a perceptual choice on some trials. This opt-out paradigm is especially popular in animal studies (see Section 5). Similarly to combined perceptual and confidence judgments, this task may be interpreted in terms of a three-alternative forced choice. Instead of deciding between percept A and percept B, the observer decides between percept "clearly A" and percept "clearly B" or chooses an intermediate percept between A and B (i.e., the observer places two Type 1 criteria on the sensory decision axis).

4.3. Postdecision Wagering

One way to improve Type 2 confidence ratings and opt-out paradigms is to impose postdecision wagering (Persaud et al. 2007). In this scenario, participants are asked to bet on the outcome of their perceptual decisions. If an observer is confident that her perceptual decision is correct, then

she will be willing to bet more. In this case, it is critical to pay careful attention to the payoff matrix that rewards good bets and penalizes bad ones (Clifford et al. 2008). In a slightly different format, an observer is asked to place a region of confidence over a perceptual space (e.g., Graf et al. 2005), and one can compute the efficiency with which the observer can use her perceptual knowledge (Landy et al. 2007). Postdecision wagering has been used in animal studies where the opt-out option is associated with a sure but small reward (Kiani & Shadlen 2009). Another form of this paradigm used in animal studies is to measure the waiting time to receive a reward (Kepecs et al. 2008). The longer the animal is willing to wait for its reward, the larger the confidence that the decision was correct.

4.4. Confidence Forced Choice

A final class of confidence paradigms is based on a confidence forced-choice task (Barthelmé & Mamassian 2009, 2010). In this paradigm, participants are presented with two stimuli on each trial, and they are asked to choose the stimulus for which they feel they are more likely to be correct on a perceptual decision. Metaperception is observed if participants consistently choose the stimulus that actually leads to a better performance. Given that on each trial two stimuli are presented, the experimenter can then split the analysis between stimuli that were chosen with confidence and those that were not chosen, and check whether there is a difference in Type 1 performance between these two sets (de Gardelle & Mamassian 2014). Similarly to other forced-choice paradigms in psychophysics (Green & Swets 1966), the confidence forced-choice method offers a bias-free measure of confidence because it is not affected by whether participants are over- or underconfident.

4.5. Over- and Underconfidence

When confidence is measured on a rating scale, biases in assigning confidence ratings to expected performance are typically found. If the (subjective) reported probability exceeds the (objective) probability of being correct, the observer is overconfident. If the subjective probability is less than the objective, the observer is underconfident. As with other metacognition studies, easy perceptual decisions are typically associated with underconfidence and hard decisions with overconfidence (Baranski & Petrusic 1994). This interaction between perceptual decision difficulty and over- or underconfidence is sometimes called the hard-easy effect (Gigerenzer et al. 1991, Harvey 1997). Overconfidence also arises when uncertainty is increased within the stimulus display (Baldassi et al. 2006) or across stimuli within a block of trials (Zylberberg et al. 2014; see comment in Section 2.8). Conversely, underconfidence arises when the task is made easier by cueing the location of the target stimulus (Rahnev et al. 2011, Wilimzig et al. 2008). Determining the ideal to match perceptual decision performance to confidence is known as the calibration problem. Indeed, given the infinite number of tasks faced by a visual system, it would be surprising to find that the system is well calibrated. Calibration can be accomplished, though not always, by practicing a task with feedback (Baranski & Petrusic 1994).

4.6. Proxies for Confidence

Do observers rely on substitutes to make confidence judgments? Intuitively, when a stimulus is highly contrasted, when most dots in a motion display move coherently, or when a stimulus is presented for a long time, it is normal to expect that perceptual accuracy should be higher than when the stimulus has low contrast, when dots move with low coherence, or when the stimulus

has brief duration, respectively. Variables that affect the visibility of a stimulus are therefore reasonable proxies for confidence. In a direction of motion discrimination task, the direction variability of individual dots affects estimated confidence, above and beyond its effect on accuracy (Spence et al. 2016). However, variability as a good proxy for confidence has limitations. First, a stimulus typically varies along more than one dimension, so monitoring uncertainty along a single dimension is not sufficient to evaluate performance. Barthelmé & Mamassian (2010) found that human observers were able to make confidence judgments when performance was influenced by more than one visual attribute, suggesting that humans can make complex confidence judgments that do not rely only on how visible a stimulus appears. Second, stimulus variability should always be combined with stimulus signal strength to predict performance accurately. De Gardelle & Mamassian (2015) found large interindividual differences when more stimulus variability or more stimulus signal strength was used as a proxy for confidence. Zylberberg et al. (2014) even found an inverse relationship between stimulus variability and confidence. Incidentally, large interindividual differences are typically found in confidence judgments, and this variability can be beneficial in searches for the neural bases of confidence computation (Fleming et al. 2010).

4.7. Confidence and Response Time

Another potential substitute for confidence may be found in the response. In particular, hard perceptual decisions typically lead to long response times (see **Figure 3*c***), so intuitively, it should be sufficient to monitor response times to obtain a sensible estimate of confidence. An inverse relationship between response time and confidence exists and is particularly strong when an observer is cautious and slow or when the range of difficulty levels is wide (Baranski & Petrusic 1994). In favor of a role for the motor system in perceptual confidence, confidence decreases when the response time is artificially lengthened without affecting perceptual accuracy (Kiani et al. 2014). There is also evidence that confidence judgments can be disrupted when transcranial magnetic stimulation (TMS) is applied over the premotor cortex (Fleming et al. 2015), although confidence can also be disrupted when TMS is applied over the occipital cortex (Rahnev et al. 2012) or over the prefrontal cortex (Rounis et al. 2010). Overall, it is difficult to argue that confidence is mainly an estimate of response time. In particular, one should keep in mind the possibility that perceptual and confidence decisions can be made at different times. Confidence could follow perceptual decision because the accumulation process continues after the perceptual decision (Pleskac & Busemeyer 2010). Alternatively, confidence could precede perceptual decision if the temporal window over which evidence is accumulated is longer for the perceptual decision than for the confidence decision (Zylberberg et al. 2012). Disentangling perceptual response time and confidence decision time is critical to testing the extent to which the motor response contributes to the computation of perceptual confidence.

4.8. Abstracting Confidence

Irrespective of the way confidence is computed, its use would be stronger if it can be compared across tasks and across observers. In a study using a confidence forced-choice paradigm (see Section 4.4), de Gardelle & Mamassian (2014) measured confidence preferences for two perceptual decisions that were either along the same stimulus attribute (e.g., stimulus orientation) or along two different attributes (e.g., orientation and spatial frequency). They found that observers were equally sensitive to making confidence choices within a single task or across tasks, suggesting that observers were able to extract an estimate of confidence that was registered in a common currency across the two tasks. Similar results were obtained when the two tasks spanned two different

sensory modalities (de Gardelle et al. 2016). Across observers, Bahrami et al. (2010) found that two observers were collectively better than one in a perceptual decision but only if the observers could communicate their confidence of being correct on every trial (see also Bang et al. 2014). However, an alternative explanation is based on the consensuality principle according to which participants are more often correct than they are incorrect. As such, the correct decision is the one consensually chosen (Koriat 2011). More work, helped by more thorough models, is expected on these topics.

5. ANIMAL STUDIES

The study of metacognition in nonhuman animals is of growing interest in comparative cognition to check whether, and if so which, animal species have the ability to judge their own judgments (Smith 2009, Kepecs & Mainen 2012). Of critical importance here is whether alternative explanations can be found to account for evidence yielding claims of metacognition in various species. In particular, it is important to check that a model based on associative learning principles whereby the animal learns specific stimulus-response contingencies cannot account for the data presented as evidence for metacognitive abilities (Smith et al. 2008).

5.1. Confidence in Rats

In a now seminal study, Kepecs et al. (2008) investigated perceptual confidence judgments in rats. The main behavioral evidence for metacognition in this study came from measuring how long the animals were willing to wait for a reward (see also Lak et al. 2014). As the authors found, when rats were confident in the correctness of their perceptual judgment, they were willing to wait a long time for an expected reward. The animals were trained to categorize two odors A and B as well as a range of mixtures of these two odors. Accuracy in this categorization task increased as the distance of the odor mixture to the stimulus category boundary increased. To characterize an animal's confidence in its categorization judgment, the authors chose the distance to the internal category boundary of the sensed odor mixture as their confidence variable (see W in Equation 8) (to be precise, the authors ran their variable through a nonlinearity, the hyperbolic tangent, to limit its maximum value to 1). The authors showed that several neurons in the rats' orbitofrontal cortex displayed activity that correlated well with this confidence variable: Low confidence was correlated with high activity. This pattern naturally emerges from an integrate-and-fire model of decision making (Insabato et al. 2010). Although it can be perilous to draw comparisons across species (Wallis 2012), human studies also emphasize the role of the prefrontal cortex in confidence judgments. Anatomically, confidence sensitivity correlated with gray matter volume and white matter microstructure in the anterior prefrontal cortex across individuals (Fleming et al. 2010). Functionally, several prefrontal areas present a larger blood-oxygen-level-dependent (BOLD) signal change for sensory discriminations that were reported easier than others, including the dorsolateral prefrontal cortex (Lau & Passingham 2006) and the medial prefrontal cortex (Rolls et al. 2010). Similarly, the ventromedial prefrontal cortex presents a larger BOLD signal change for more extreme stimulus values (De Martino et al. 2013, Lebreton et al. 2015).

A dominant paradigm in animal studies is to train an animal to categorize two stimuli and then to give it the option not to commit to either of these categories in trials when it is unsure (see Section 4.2). If an animal makes use of this third opt-out option in reasonable cases, for instance, when stimuli are presented near the boundary between the two categories, then researchers assume the animal has the ability to monitor its own uncertainty in its perceptual decision. On the basis of this paradigm, a study of stimulus duration discrimination argued that rats have metacognitive

abilities (Foote & Crystal 2007), but the results in this study could be simulated by a model that does not rely on metacognitive principles (Smith et al. 2008). Other studies where rats were given the possibility of replaying a stimulus to increase their chances of being correct also led to inconclusive results (Foote & Crystal 2012). In summary, whether rats have metacognitive abilities remains a debated issue.

5.2. Confidence in Primates

In the opt-out paradigm, it is critical to determine how an animal should be rewarded when the available opt-out option is chosen. In a study with New World monkeys, Beran et al. (2009) trained capuchin monkeys to discriminate between sparse and dense sets of dots. They found that capuchin monkeys did not choose the opt-out option when doing so was not rewarded, but animals did choose a middle category between the two main categories if this middle option was rewarded. From this study, the authors concluded that capuchin monkeys did not present metacognitive abilities. Other studies with Old World monkeys showed different results. In a seminal study, Kiani & Shadlen (2009) found that rhesus macaque monkeys consistently chose an opt-out option for difficult trials where the motion direction of a stimulus was highly corrupted by noise. Critically, the opt-out option was presented in only half the trials, but when available, it was associated with a small but sure reward. In trials where macaque monkeys had this opt-out option, their performance was better than when it was not available. This result held for all stimulus noise levels, suggesting that the monkeys were not simply choosing the opt-out option as a middle category that corresponded to high-noise stimuli (i.e., splitting the stimulus space into three categories rather than two). Interestingly, and in contrast to capuchin monkeys, macaque monkeys chose the opt-out option even when this option was not rewarded (Smith et al. 1997), thus leading researchers to conclude that macaque monkeys have metacognitive abilities. Further studies have shown macaque monkeys can also use a rating scale to report their evaluation of the degree of accuracy of their perceptual judgments (Shields et al. 2005). However, in contrast to humans who can use a rating scale with an arbitrary number of levels of confidence, macaque monkeys seemed limited to only two levels.

The neural structures involved in metacognition have been extensively studied in primates. In macaque monkeys, Kiani & Shadlen (2009) found that neurons in the lateral intraparietal (LIP) cortex displayed activity that reflected the monkey's choice certainty. Neurons in area LIP increased their discharge when the stimulus was more salient, that is, when stimulus noise was low or when stimulus duration was long. When a monkey was less likely to select the opt-out choice, this discharge also increased. Because LIP neurons are involved in accumulation of evidence models for perceptual decision making (Roitman & Shadlen 2002), these results suggest that the same cortical area is involved in perceptual choice and confidence judgment. Further evidence that LIP neuron activity is causally related to confidence came from microstimulation studies. Microstimulating LIP neurons during stimulus presentation had the same effect as changing evidence in favor of one perceptual decision (Fetsch et al. 2014). Other studies have shown that the pulvinar neurons decrease their activity when monkeys decide to opt out, suggesting that this visual thalamic structure contributes to confidence judgments (Komura et al. 2013).

The conclusions from these studies in primates, arguing that confidence is computed in perceptual brain structures, are in stark contrast with the studies summarized above in rats and humans, where confidence was predominantly associated with the prefrontal cortex. Clearly, a better understanding of the mechanisms underlying the computation of confidence will come from the combination of advanced theoretical principles, careful methodologies, and precise physiological results.

SUMMARY POINTS

1. Visual confidence is defined as the observer's estimated probability of being correct in a perceptual task given her decision in this task and some internal confidence evidence.
2. In SDT models of confidence, the confidence evidence is often the distance of the sensory evidence to the criterion.
3. Accumulation of evidence models can account for accuracy, confidence, and response times.
4. The use of rating scales is a popular method to measure confidence, but one should be cautious not to confuse confidence decision and finer perceptual decision.
5. Overconfidence is often found for hard perceptual decisions.
6. There is evidence that nonhuman primates and even rats have some metaperception abilities.

FUTURE ISSUES

1. Can the two frameworks based on SDT and accumulation models be combined into a single one?
2. Are perceptual and confidence decisions computed separately?
3. Do observers use proxies of confidence such as stimulus variability?
4. Is there a link between metaperception ability and perceptual learning efficiency?
5. Will there ever be a consensual paradigm to study confidence in nonhuman animals?

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by French ANR grant "Visual Confidence" (ANR-10-BLAN-1910) and grant ANR-NSF-CRCNS-1430262. The author wishes to acknowledge fruitful discussions with Simon Barthelmé, Baptiste Caziot, Vincent de Gardelle, and Alan Lee.

LITERATURE CITED

Aitchison L, Bang D, Bahrami B, Latham PE. 2015. Doubly Bayesian analysis of confidence in perceptual decision-making. *PLOS Comput. Biol.* 11:e1004519

Bahrami B, Olsen K, Latham PE, Roepstorff A, Rees G, Frith CD. 2010. Optimally interacting minds. *Science* 329:1081–85

Baker DH, Cass JR. 2013. A dissociation of performance and awareness during binocular rivalry. *Psychol. Sci.* 24:2563–68

Baldassi S, Megna N, Burr DC. 2006. Visual clutter causes high-magnitude errors. *PLOS Biol.* 4:e56

Bang D, Fusaroli R, Tylén K, Olsen K, Latham PE, et al. 2014. Does interaction matter? Testing whether a confidence heuristic can replace interaction in collective decision-making. *Conscious. Cogn.* 26:13–23
Baranski JV, Petrusic WM. 1994. The calibration and resolution of confidence in perceptual judgments. *Percept. Psychophys.* 55:412–28
Barrett AB, Dienes Z, Seth AK. 2013. Measures of metacognition on signal-detection theoretic models. *Psychol. Methods* 18:535–52
Barthelmé S, Mamassian P. 2009. Evaluation of objective uncertainty in the visual system. *PLOS Comput. Biol.* 5:e1000504
Barthelmé S, Mamassian P. 2010. Flexible mechanisms underlie the evaluation of visual confidence. *PNAS* 107:20834–39
Beran MJ, Smith JD, Coutinho MVC, Couchman JJ, Boomer J. 2009. The psychological organization of "uncertainty" responses and "middle" responses: a dissociation in capuchin monkeys (*Cebus apella*). *J. Exp. Psychol.* 35:371–81
Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. 2006. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* 113:700–65
Churchland AK, Kiani R, Shadlen MN. 2008. Decision-making with multiple alternatives. *Nat. Neurosci.* 11:693–702
Clarke FR, Birdsall TG, Tanner WP Jr. 1959. Two types of ROC curves and definitions of parameters. *J. Acoust. Soc. Am.* 31:629–30
Clifford CWG, Arabzadeh E, Harris JA. 2008. Getting technical about awareness. *Trends Cogn. Sci.* 12:54–58
Cowey A. 2010. The blindsight saga. *Exp. Brain Res.* 200:3–24
de Gardelle V, Le Corre F, Mamassian P. 2016. Confidence as a common currency between vision and audition. *PLOS ONE* 11:e0147901
de Gardelle V, Mamassian P. 2014. Does confidence use a common currency across two visual tasks? *Psychol. Sci.* 25:1286–88
de Gardelle V, Mamassian P. 2015. Weighting mean and variability during confidence judgments. *PLOS ONE* 10:e0120870
De Martino B, Fleming SM, Garrett N, Dolan RJ. 2013. Confidence in value-based choice. *Nat. Neurosci.* 16:105–10
Descartes R. 1637. *Discours de la Méthode: Pour bien conduire sa raison, et chercher la vérité dans les sciences*. The Hague: Ian Maire
Drugowitsch J, Moreno-Bote R, Pouget A. 2014. Relation between belief and performance in perceptual decision making. *PLOS ONE* 9:e96511
Evans S, Azzopardi P. 2007. Evaluation of a "bias-free" measure of awareness. *Spat. Vis.* 20:61–77
Fetsch CR, Kiani R, Newsome WT, Shadlen MN. 2014. Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83:797–804
Fleming SM, Lau H. 2014. How to measure metacognition. *Front. Hum. Neurosci.* 8:443
Fleming SM, Maniscalco B, Ko Y, Amendi N, Ro T, Lau H. 2015. Action-specific disruption of perceptual confidence. *Psychol. Sci.* 26:89–98
Fleming SM, Weil RS, Nagy Z, Dolan RJ, Rees G. 2010. Relating introspective accuracy to individual differences in brain structure. *Science* 329:1541–43
Foote AL, Crystal JD. 2007. Metacognition in the rat. *Curr. Biol.* 17:551–55
Foote AL, Crystal JD. 2012. "Play it again": a new method for testing metacognition in animals. *Anim. Cogn.* 15(2):187–99
Galvin SJ, Podd JV, Drga V, Whitmore J. 2003. Type 2 tasks in the theory of signal detectability: discrimination between correct and incorrect decisions. *Psychon. Bull. Rev.* 10:843–76
Gherman S, Philiastides MG. 2015. Neural representations of confidence emerge from the process of decision formation during perceptual choices. *NeuroImage* 106:134–43
Gigerenzer G, Hoffrage U, Kleinbolting H. 1991. Probabilistic mental models: a Brunswikian theory of confidence. *Psychol. Rev.* 98:506–28
Graf E, Warren PA, Maloney LT. 2005. Explicit estimation of visual uncertainty in human motion processing. *Vis. Res.* 45:3050–59

Green DM, Swets JA. 1966. *Signal Detection Theory and Psychophysics*. New York: Wiley
Harvey N. 1997. Confidence in judgment. *Trends Cogn. Sci.* 1:78–82
Insabato A, Pannunzi M, Rolls ET, Deco G. 2010. Confidence-related decision making. *J. Neurophysiol.* 104:539–47
Juslin P, Olsson H. 1997. Thurstonian and Brunswikian origins of uncertainty in judgment: a sampling model of confidence in sensory discrimination. *Psychol. Rev.* 104:344–66
Kepecs A, Mainen ZF. 2012. A computational framework for the study of confidence in humans and animals. *Philos. Trans. R. Soc. Lond. B* 367:1322–37
Kepecs A, Uchida N, Zariwala HA, Mainen ZF. 2008. Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455:227–31
Kiani R, Corthell L, Shadlen MN. 2014. Choice certainty is informed by both evidence and decision time. *Neuron* 84:1329–42
Kiani R, Shadlen MN. 2009. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* 324:759–64
Kolb FC, Braun J. 1995. Blindsight in normal observers. *Nature* 377:336–38
Komura Y, Nikkuni A, Hirashima N, Uetake T, Miyamoto A. 2013. Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat. Neurosci.* 16:749–55
Koriat A. 2011. Subjective confidence in perceptual judgments: a test of the self-consistency model. *J. Exp. Psychol. Gen.* 140:117–39
Kunimoto C, Miller J, Pashler H. 2001. Confidence and accuracy of near-threshold discrimination responses. *Conscious. Cogn.* 10:294–340
Lak A, Costa GM, Romberg E, Koulakov AA, Mainen ZF, Kepecs A. 2014. Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron* 84:190–201
Laming DRJ. 1968. *Information Theory of Choice Reaction Time*. New York: Wiley
Landy MS, Goutcher R, Trommershauser J, Mamassian P. 2007. Visual estimation under risk. *J. Vis.* 7(6):4
Lau HC, Passingham RE. 2006. Relative blindsight in normal observers and the neural correlate of visual consciousness. *PNAS* 103:18763–68
Lebreton M, Abitbol R, Daunizeau J, Pessiglione M. 2015. Automatic integration of confidence in the brain valuation signal. *Nat. Neurosci.* 18:1159–67
Liston DB, Stone LS. 2013. Saccadic brightness decisions do not use a difference model. *J. Vis.* 13(8):1
Macmillan NA, Creelman CD. 2005. *Detection Theory: A User's Guide*. Mahwah, NJ: Lawrence Erlbaum Assoc. 2nd ed.
Maniscalco B, Lau H. 2012. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Conscious. Cogn.* 21:422–30
Masson MEJ, Rotello CM. 2009. Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: implications for studies of metacognitive processes. *J. Exp. Psychol. Learn. Mem. Cogn.* 35:509–27
Metcalfe J, Shimamura AP. 1994. *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press
Meyniel F, Sigman M, Mainen ZF. 2015. Confidence as Bayesian probability: from neural origins to behavior. *Neuron* 88:78–92
Moreno-Bote R. 2010. Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators. *Neural Comput.* 22:1786–11
Morgan MJ, Mason AJ, Solomon JA. 1997. Blindsight in normal subjects? *Nature* 385:401–2
Nelson TO. 1984. A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychol. Bull.* 95:109–33
Paulk AC, Kirszenblat L, Zhou Y, van Swinderen B. 2015. Closed-loop behavioral control increases coherence in the fly brain. *J. Neurosci.* 35:10304–15
Peirce CS, Jastrow J. 1885. On small differences of sensation. *Mem. Natl. Acad. Sci.* 3:73–83
Persaud N, Mcleod P, Cowey A. 2007. Post-decision wagering objectively measures awareness. *Nat. Neurosci.* 10:257–61
Pleskac TJ, Busemeyer JR. 2010. Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychol. Rev.* 117:864–901

Pouget A, Drugowitsch J, Kepecs A. 2016. Confidence and certainty: distinct probabilistic quantities for different goals. *Nat. Neurosci.* 19:366–74
Raab DH. 1962. Statistical facilitation of simple reaction times. *Trans. N.Y. Acad. Sci.* 24:574–90
Rabbitt PMA. 1966. Error correction time without external error signals. *Nature* 212:438
Rahnev DA, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H. 2011. Attention induces conservative subjective biases in visual perception. *Nat. Neurosci.* 14:1513–15
Rahnev DA, Maniscalco B, Luber B, Lau HC, Lisanby SH. 2012. Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J. Neurophysiol.* 107:1556–63
Ratcliff R, McKoon G. 2008. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput*. 20:873–922
Ratcliff R, Starns JJ. 2013. Modeling confidence judgments, response times, and multiple choices in decision making: recognition memory and motion discrimination. *Psychol. Rev.* 120:697–719
Resulaj A, Kiani R, Wolpert DM, Shadlen MN. 2009. Changes of mind in decision-making. *Nature* 461:263–66
Roitman JD, Shadlen MN. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J. Neurosci.* 22:9475–89
Rolls ET, Grabenhorst F, Deco G. 2010. Decision-making, errors, and confidence in the brain. *J. Neurophysiol.* 104:2359–74
Rounis E, Maniscalco B, Rothwell JC, Passingham RE, Lau H. 2010. Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn. Neurosci*. 1:165–75
Sahraie A, Weiskrantz L, Barbur JL. 1998. Awareness and confidence ratings in motion perception without geniculo-striate projection. *Behav. Brain Res.* 96:71–77
Shields WE, Smith JD, Guttmannova K, Washburn DA 2005 Confidence judgments by humans and rhesus monkeys. *J. Gen. Psychol.* 132:165–86
Smith JD. 2009. The study of animal metacognition. *Trends Cogn. Sci.* 13:389–96
Smith JD, Beran MJ, Couchman JJ, Coutinho MVC. 2008. The comparative study of metacognition: sharper paradigms, safer inferences. *Psychon. Bull. Rev.* 15:679–91
Smith JD, Shields WE, Schull J, Washburn DA. 1997. The uncertain response in humans and animals. *Cognition* 62:75–97
Smith PL, Vickers D. 1988. The accumulator model of two-choice discrimination. *J. Math. Psychol.* 32:135–68
Spence ML, Dux PE, Arnold DH. 2016. Computations underlying confidence in visual perception. *J. Exp. Psychol. Hum. Percept. Perform.* 42:671–82
Tiwari G, Bangdiwala S, Saraswat A, Gaurav S. 2007. Survival analysis: pedestrian risk exposure at signalized intersections. *Transport. Res. F* 10:77–89
van den Berg R, Anandalingam K, Zylberberg A, Kiani R, Shadlen MN, Wolpert DM. 2016. A common mechanism underlies changes of mind about decisions and confidence. *eLife* 5:e12192
Vickers D. 1979. *Decision Processes in Visual Perception*. New York: Academic
Vickers D, Packer J. 1982. Effects of alternating set for speed or accuracy on response-time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychol.* 50:179–97
Wallis JD. 2012. Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat. Neurosci.* 15:13–19
Wilimzig C, Tsuchiya N, Fahle M, Einhäuser W, Koch C. 2008. Spatial attention increases performance but not subjective confidence in a discrimination task. *J. Vis.* 8(5):7
Yeung N, Summerfield C. 2012. Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. Lond. B* 367:1310–21
Zizlsperger L, Sauvigny T, Händel B, Haarmeier T. 2014. Cortical representations of confidence in a visual perceptual decision. *Nat. Commun.* 5:3940
Zylberberg A, Barttfeld P, Sigman M. 2012. The construction of confidence in a perceptual decision. *Front. Integr. Neurosci.* 6:79
Zylberberg A, Roelfsema PR, Sigman M. 2014. Variance misperception explains illusions of confidence in simple perceptual decisions. *Conscious. Cogn.* 27:246–53

Annual Review of
Vision Science

Volume 2, 2016

Contents

The Road to Certainty and Back
Gerald Westheimer 1

Experience-Dependent Structural Plasticity in the Visual System
Kalen P. Berry and Elly Nedivi 17

Strabismus and the Oculomotor System: Insights from Macaque Models
Vallabh E. Das 37

Corollary Discharge and Oculomotor Proprioception: Cortical Mechanisms for Spatially Accurate Vision
Linus D. Sun and Michael E. Goldberg 61

Mechanisms of Orientation Selectivity in the Primary Visual Cortex
Nicholas J. Priebe 85

Perceptual Learning: Use-Dependent Cortical Plasticity
Wu Li 109

Early Visual Cortex as a Multiscale Cognitive Blackboard
Pieter R. Roelfsema and Floris P. de Lange 131

Ocular Photoreception for Circadian Rhythm Entrainment in Mammals
Russell N. Van Gelder and Ethan D. Buhr 153

Probing Human Visual Deficits with Functional Magnetic Resonance Imaging
Stelios M. Smirnakis 171

Retinoids and Retinal Diseases
Philip D. Kiser and Krzysztof Palczewski 197

Understanding Glaucomatous Optic Neuropathy: The Synergy Between Clinical Observation and Investigation
Harry A. Quigley 235

Vision and Aging
Cynthia Owsley 255

Electrical Stimulation of the Retina to Produce Artificial Vision
James D. Weiland, Steven T. Walston, and Mark S. Humayun 273

Evolution of Concepts and Technologies in Ophthalmic Laser Therapy
Daniel Palanker 295

Low Vision and Plasticity: Implications for Rehabilitation
Gordon E. Legge and Susana T.L. Chung 321

The Human Brain in Depth: How We See in 3D
Andrew E. Welchman 345

Visual Object Recognition: Do We (Finally) Know More Now Than We Did?
Isabel Gauthier and Michael J. Tarr 377

3D Displays
Martin S. Banks, David M. Hoffman, Joohwan Kim, and Gordon Wetzstein 397

Capabilities and Limitations of Peripheral Vision
Ruth Rosenholtz 437

Visual Confidence
Pascal Mamassian 459

Errata

An online log of corrections to *Annual Review of Vision Science* articles may be found at http://www.annualreviews.org/errata/vision