

# Predicting the Accuracy of a Decision: A Neural Mechanism of Confidence

CHRISTOPHER R. FETSCH,<sup>1</sup> ROOZBEH KIANI,<sup>2</sup> AND MICHAEL N. SHADLEN<sup>1</sup>

<sup>1</sup>*Howard Hughes Medical Institute, Department of Neuroscience and Kavli Institute for Brain Science, Columbia University, New York, New York 10032*

<sup>2</sup>*Center for Neural Science, New York University, New York, New York 10003*

*Correspondence: crf2129@columbia.edu*

The quantitative study of decision-making has traditionally rested on three key behavioral measures: accuracy, response time, and confidence. Of these, confidence—defined as the degree of belief, prior to feedback, that a decision is correct—is least well understood at the level of neural mechanism, although recent years have seen a surge in interest in the topic among theoretical and systems neuroscientists. Here we review some of these developments and highlight a particular candidate mechanism for assigning confidence in a perceptual decision. The mechanism is appealing because it is rooted in the same decision-making framework—bounded accumulation of evidence—that successfully explains accuracy and reaction time in many tasks, and it is validated by neurophysiology and microstimulation experiments.

The merits of self-knowledge were well known to the ancients, as reflected in the Confucian saying: “When you know a thing, to hold that you know it; and when you do not know a thing, to allow that you do not know it—that is knowledge” (*Analects*, 2:17). A rudimentary but important form of self-knowledge is the degree of certainty, or confidence, one has in a decision (e.g., whether it is correct or will lead to a desired outcome). In many instances,<sup>1</sup> a judgment of confidence boils down to the question: How reliable was the evidence on which my decision was based? Here, evidence is meant to include external events transduced by the senses, as well as internal signals derived from memory. It seems likely that evidence from multiple sources, along with other factors such as biases and prior expectations, converges onto an internal representation that governs the choice. As we will show, the same process that converts evidence to a choice can additionally explain not only the time required to make the decision but also the confidence associated with it.

Confidence has obvious practical and social value in our daily lives. Any time you attempt to convince someone that something is true, or justify a course of action, your degree of certainty guides the interaction. Confidence serves to modulate learning rates and the allocation of attentional resources (Dayan et al. 2000), and it enables online regulation of the decision process itself (Vickers 1979). More generally, confidence is crucial for behavior in complex environments where sequences of decisions must be made without immediate feedback. This type of problem is common to most animals, as it includes deci-

sions about foraging (explore or exploit?) and predator avoidance (hide or flee?). Yet in common parlance it also carries an air of subjectivity that would seem to preclude its study in animal models. This is a challenge if we wish to study confidence at a mechanistic level using the full toolkit of neuroscience. Fortunately, recent advances have begun to break through this barrier. Researchers interested in the neurobiology of decision-making have taken on the problem in earnest, buoyed by the propagation of new behavioral paradigms that allow an assessment of confidence in nonverbal animals. One might say that, in neuroscience at least, confidence no longer plays a “Cinderella role. . . , overlooked as an interesting variable in its own right” (Vickers 2001).

Psychologists have long recognized two main sources of information for generating confidence judgments: evidence strength and decision time. The relative emphasis on the latter has waxed and waned over the decades, and recent studies in both humans (e.g., Rahnev et al. 2011; Ko and Lau 2012) and animals (Kepecs et al. 2008; Komura et al. 2013) have employed signal-detection models that do not involve the temporal dimension at all. However, most decisions are based on streams of evidence rather than single observations and generally incorporate a termination rule that implements the trade-off between speed and accuracy. Thus, the decision process is naturally conceived as the accumulation of evidence over time, formalized with a variety of models based on sequential analysis (Stone 1960; Laming 1968; Link and Heath 1975; Ratcliff 1978; Vickers 1979). These models have been highly successful in explaining the speed and accuracy of decisions in multiple settings, and they find support in neurophysiological observations (Gold and Shadlen 2007; Shadlen and Kiani 2013). We will describe recent work (Kiani and Shadlen 2009; Fetsch et al. 2014; Kiani et al. 2014a) that

---

<sup>1</sup>Naturally, we are unable to cover every sense of the word confidence (but see Conclusions).

extends the bounded accumulation framework in a way that—building on insights by Vickers and others—parsimoniously explains confidence and is consistent with neurophysiology. We will also touch on a few lesser known implications of these advances—the nature of noise, biases, and other forms of confidence—and propose some future directions.

### EVIDENCE AND TIME: ORIGINS OF THE PSYCHOPHYSICAL STUDY OF CONFIDENCE

Confidence was first approached in a quantitative manner by a second generation of experimental psychologists (Peirce and Jastrow 1884; Fullerton and Cattell 1892) who began to question the assumption of their forebears—Fechner, Weber, and Wundt—of a “just-noticeable” difference or increment. The strong version of this idea, known as absolute- or high-threshold theory, asserts that variations in the internal representation of a stimulus (or difference between stimuli) cannot influence a percept unless they result in the crossing of an absolute threshold. Below the threshold, or within a just-noticeable increment, changes in stimulus intensity or difference magnitude affect only the probability of threshold crossing, not the quality of the sensation. By the end of the 19th century, however, it became evident that the nervous system permits smooth gradations of “clearness” in discrimination (Fullerton and Cattell 1892), an intuitive concept that opened the door to the rigorous measurement and modeling of subjective confidence judgments.<sup>2</sup>

For confidence to be useful, it ought to predict accuracy on average, at least weakly.<sup>3</sup> Indeed, the empirical relationship between confidence and accuracy inspired the first attempt to describe confidence using an equation, namely as the scaled logit of probability correct (Peirce and Jastrow 1884). Later experiments (Garrett 1922; Johnson 1939; Festinger 1943; Pierrel and Murray 1963) found that in fact confidence is more tightly coupled to evidence strength or discriminability than to accuracy per se. In parallel, another school of thought was emerging based on an observed inverse relationship between confidence and the time taken to reach a decision (reaction time, RT; Henmon 1911; Seward 1928; Kellogg 1931; Volkman 1934; Johnson 1939; Audley 1960). The consistency of the trend led some authors to conclude that the level of confidence might be dictated entirely by RT (Volkman 1934; Reed 1951) or, relatedly, the number of contradicting observations leading up to the choice (Audley 1960). According to this view, confidence bears a positive relationship with evidence strength and accuracy solely as a by-product of being inversely

related to time, because those other quantities are also inversely related to time.

To reconcile these findings, Vickers emphasized the distinction between the two ways a decision process may be terminated (Vickers 1979; Vickers and Packer 1982). Intuitively, if the experimenter controls the stimulus duration, then longer trials will tend to deliver more information and result in greater accuracy. In this case (“time-limited,” Fig. 1A,B), confidence should increase as a function of time. If instead the subject decides when to terminate the decision (“information-limited,” Fig. 1C), longer decisions will be associated with lower discriminability and hence lower confidence. This insight, and its agreement with data (e.g., Irwin et al. 1956; Audley 1960), led Vickers to posit the balance-of-evidence (BE) hypothesis (Vickers 1972), in which confidence is a function of the difference between accumulated evidence totals favoring the chosen and unchosen option. With BE, a single mechanism can be responsible for confidence judgments in both types of task yet produce the opposite relationship with trial duration.

Importantly, the BE hypothesis implies that the primary relationship is between evidence strength and confidence; the correlation between RT and confidence is incidental (Vickers 2001)—a property shared by more recent work as well (Rolls et al. 2010; Zylberberg et al. 2012; Ratcliff and Starns 2013). In contrast, our recent findings (Kiani and Shadlen 2009; Fetsch et al. 2014; Kiani et al. 2014a) support a role for time in and of itself, as explained below.

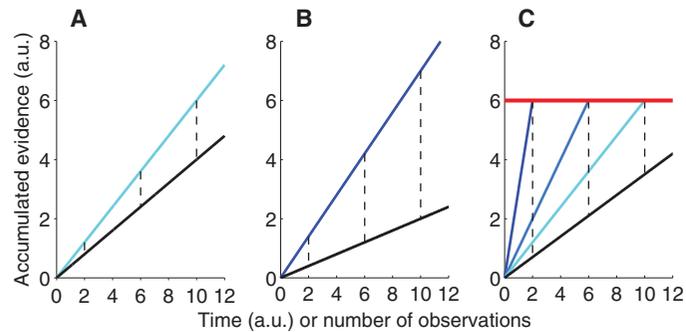
### A TALE OF TWO THEORIES

From the last century of psychophysical inquiry emerged two general frameworks that form the backbone of the modern study of confidence: signal-detection theory (SDT) and sequential analysis (SA). According to SDT (Peterson and Birdsall 1953; Tanner and Swets 1954; van Meter and Middleton 1954), an observation is drawn from one of two overlapping distributions, and the goal of the observer is to determine which distribution. In detection tasks, the two distributions represent “noise” and “signal plus noise,” whereas in discrimination tasks they might represent the two alternatives being discriminated. In either case, the decision is made by converting the noisy observation to a decision variable (DV) and comparing it with a criterion. Because the DV is monotonically related to the probability of making the observation given the state of the world (Swets et al. 1961; Gold and Shadlen 2001), its distance from the criterion is a reasonable proxy for confidence. The seeds of this idea actually predated formal SDT with the theory of Cartwright and Festinger (1943), but it matured in the ensuing decades among researchers across several disciplines (Norman and Wickelgren 1969; Kinchla and Collyer 1974; Ferrell and Mcgoey 1980; Treisman and Faulkner 1984).

SDT-based models of confidence remain highly influential in psychology and neuroscience (Higham 2007;

<sup>2</sup>It also presaged the more flexible criterion that underlies signal-detection theory (SDT; see below).

<sup>3</sup>If we define a decision as a commitment to a proposition or hypothesis, then confidence is the subjective (Bayesian) probability—for a single decision—that the hypothesis is true, given the data. This differs from accuracy, which is the objective (frequentist) probability—across multiple decisions—that the chosen hypothesis was true. The reason these quantities are not equivalent, except under certain conditions, has recently been clarified by Drugowitsch et al. (2014).



**Figure 1.** Confidence as the balance of evidence (BE) predicts different relationships with decision time depending on how the decision terminates. (*A,B*) Accumulated evidence totals favoring two options, as a function of time or number of observations, are depicted schematically as lines of different slope (blue = chosen option, black = unchosen). When the environment or experimenter dictates the time of commitment (*A,B*), (e.g., by soliciting a choice after 2, 6, or 10 observations), the BE (length of vertical dashed lines) increases with elapsed time. The pattern holds for both low (*A*) and high (*B*) discriminability. (*C*) If on the other hand the decision is terminated when accumulated evidence reaches a threshold (red line), BE decreases with elapsed decision time, because the latter is inversely related to discriminability (three different levels are shown as shades of blue). Note that this relationship breaks down when negative increments are allowed (i.e., if accumulators are anticorrelated; see description of Kiani et al. [2014a] below for details). (Adapted from Vickers and Packer 1982.)

Kepecs et al. 2008; Rounis et al. 2010; Rahnev et al. 2011, 2012; Ko and Lau 2012; Komura et al. 2013; Lak et al. 2014; Scott et al. 2014). However, these models are fundamentally limited: they cannot explain the time it takes to reach a decision nor the effects of decision time on confidence mentioned above. Of course, all models simplify to some degree, but in this case time is not just a detail that can safely be ignored for convenience. As we shall see, deliberation time is a key factor in the determination of confidence when there is a speed-accuracy trade-off at play—that is, when the decision-maker terminates a decision based on accumulated evidence.<sup>4</sup> In such settings, not only does time play a role in computing confidence, but also the key prediction of “distance from criterion” is reversed (see below). This is not to minimize the relevance of SDT, nor to rule out a distance-from-criterion mechanism in all cases (e.g., when the task involves little or no deliberation; Zariwala et al. 2013). But in line with the psychophysical tradition, it seems clear that a general neural theory of confidence must include the dimension of time.

A natural starting point for such a theory can be found within the family of models involving sequential analysis, a statistical decision process that can be viewed as an extension of SDT into the time domain (Fig. 2A). Enthusiasm for SA models—which include bounded evidence accumulation, drift-diffusion, and race models—derives from their effectiveness in explaining choice and RT distributions across many domains<sup>5</sup> (Busemeyer 1985; Luce 1986; Reddi et al. 2003; Ratcliff and Smith 2004; Krajbich and Rangel 2011; DasGupta et al. 2014), as well

as their robust neural correlates (Smith and Ratcliff 2004; Gold and Shadlen 2007; Heekeren et al. 2008; Ploran et al. 2011). When monkeys are trained to report the outcome of a decision with an eye movement, neurons related to oculomotor planning show persistent activity resembling the accumulation of noisy evidence toward a threshold or bound (Hanes and Schall 1996; Shadlen and Newsome 1996; Kim and Shadlen 1999; Horwitz and Newsome 2001; Roitman and Shadlen 2002; Ratcliff et al. 2003). In the lateral intraparietal area (LIP), this activity reflects not only the mean (Roitman and Shadlen 2002; Huk and Shadlen 2005; Kiani et al. 2008) but also the variance and autocorrelation (Churchland et al. 2011) of an evolving DV that explains choice and RT. As shown next, our recent work extends the bounded accumulation framework to explain how the brain might arrive at a prediction about the accuracy of a decision.

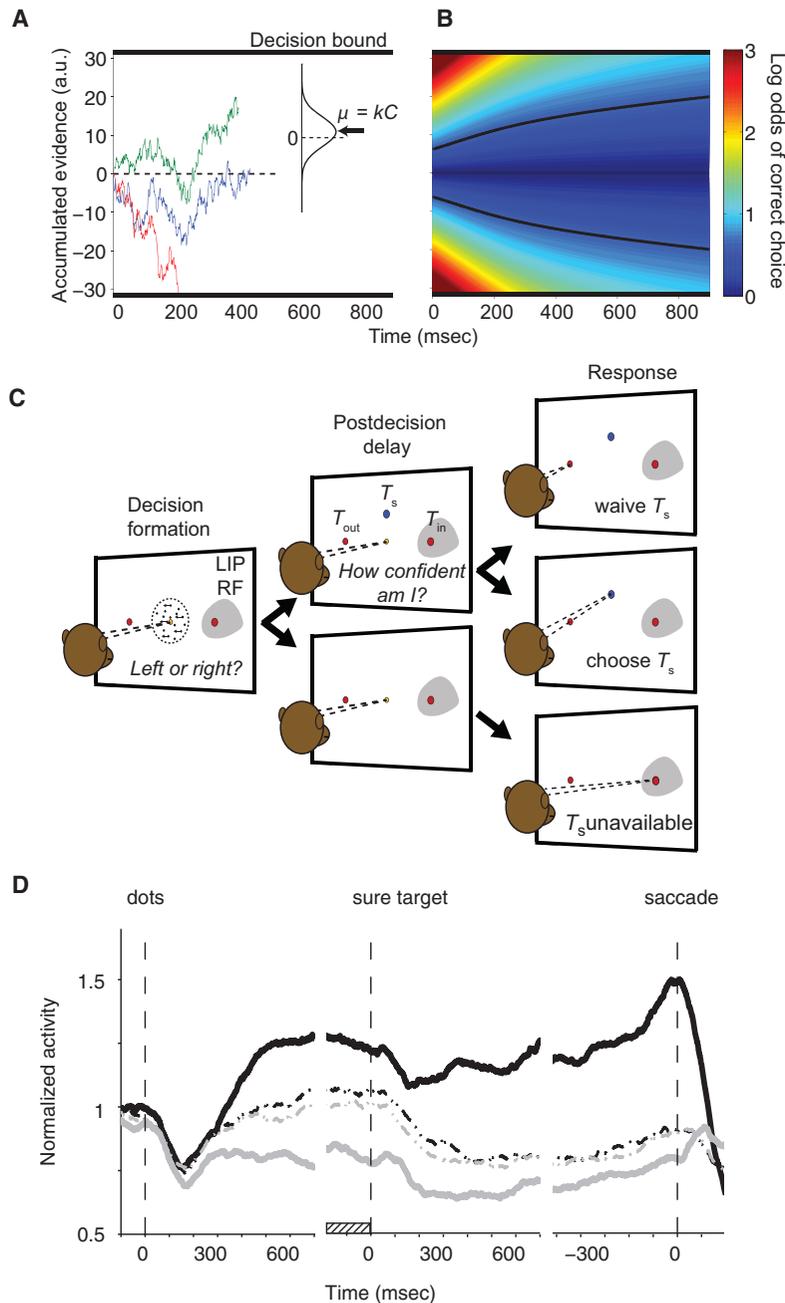
### BOUNDED EVIDENCE ACCUMULATION UNIFIES CHOICE, RT, AND CONFIDENCE IN A COMMON FRAMEWORK

A feature of SA models is their explicit representation of both time and evidence, the principal ingredients for establishing a level of confidence. A key insight is that the amount of accumulated evidence (i.e., the state of the DV) at the time of the choice is associated with a particular probability that the decision will be correct (Fig. 2B). Kiani and Shadlen (2009) hypothesized that the brain uses implicit knowledge of this association, or mapping, to establish a degree of confidence in the choice, harnessing the same decision-making machinery that gives rise to the choice itself and the time it takes to decide.

Coming up with a way to test the hypothesis at the level of neurophysiology was not trivial: How does one ask a laboratory animal how confident it is? Fortunately, some promising answers to this question had recently been

<sup>4</sup>Note that the brain may apply such a termination rule even when the environment or experimenter controls stimulus duration (Kiani et al. 2008).

<sup>5</sup>Why do evidence accumulation/drift-diffusion models provide such good fits to choice and RT data even when the task does not involve a continuous stream of evidence to integrate? This is an important question for future work, but we speculate the common feature is the sequential nature of updating (e.g., repeated queries of a mnemonic representation).



**Figure 2.** Model, behavioral task, and neural representation of a DV that explains confidence. (A) Depiction of bounded accumulation model (i.e., one-dimensional drift-diffusion) in which momentary evidence is drawn from a normal distribution with mean ( $\mu$ ) proportional to motion strength (coherence, C). (B) The model assumes the brain has implicit knowledge of the probability of being correct, plotted as the logarithm of the posterior odds (color), as a function of accumulated evidence and decision time. Black contours indicate a criterion on log-odds that partitions the space into low- and high-confidence regions. (C) Postdecision wagering (PDW) task (see text). Circle in *left* panel represents the random-dot motion stimulus. Red spots indicate direction targets, one of which was aligned with the response field (RF, gray region) of a neuron in area LIP. Blue spot is the “sure-bet” target ( $T_s$ ). (A–C, Reproduced from Fetsch et al. 2014, with permission from Elsevier.) (D) LIP responses predict direction choices and PDW. Solid black and gray traces show averaged activity for trials ending in a saccadic choice into ( $T_{in}$ ) or out of ( $T_{out}$ ) the RF, respectively. Dashed traces represent trials in which  $T_s$  was chosen (black and gray, motion toward  $T_{in}$  or  $T_{out}$ , respectively). Figure shows only the random half of trials in which  $T_s$  was presented during the delay period. The monkey did not know whether the  $T_s$  option would be available until this time. (D, Reproduced from Kiani and Shadlen 2009, with permission from AAAS.)

furnished by a number of innovative studies in comparative psychology (for reviews, see Hampton 2009; Smith 2009; Terrace and Son 2009; Kepecs and Mainen 2012). Inspired by these advances in behavioral methodology, Kiani and Shadlen (2009) trained monkeys to report their confidence in a decision about the direction of visual motion (left or right) using a form of a postdecision wagering (PDW). Following the perceptual decision phase, on a random half of trials, the monkeys were allowed to refrain from reporting the direction choice and instead select a “sure bet” target ( $T_s$ ; Fig. 2C). Choosing the sure bet yielded a guaranteed reward but one that was smaller than for a correct direction choice. The monkey could not predict whether he would have the opportunity to choose  $T_s$  until well after the visual stimulus was turned off. Indeed, the task was explicitly designed so that the monkey had to form a binary decision about the motion on each trial, even if on a particular trial they might have hoped that  $T_s$  would be shown. Crucially, the monkey’s accuracy was greater on trials when the sure bet was offered but waived compared with when it was not offered. The improvement in accuracy held within each level of difficulty and for a subset of trials with identical motion stimuli (Kiani and Shadlen 2009). This strongly suggests that the decision to opt out was based on an assessment of the reliability of internal sensory evidence, not simply a classification of trial difficulty or the presence of lapses of attention.<sup>6</sup>

If the DV that underlies the choice also governs confidence, then a putative neural representation of the DV should predict postdecision wagering behavior. This was found to be the case. To appreciate the result, recall that LIP neurons represent the accumulated evidence favoring the saccadic choice target in their response field (RF). This activity ramps upward preceding choices into the RF and decreases or remains flat preceding choices out of the RF. Prior to sure-bet choices (Kiani and Shadlen 2009), the firing rates were intermediate (Fig. 2D), analogous to the intermediate levels of evidence associated with lower confidence in the model (Fig. 2B). This trend remained after accounting for stimulus difficulty (motion strength and viewing duration) and was apparent in both the rate of rise of activity during motion viewing and the sustained level of activity during the delay period before the sure-bet target was presented. In control experiments with the eventual sure-bet target location aligned with the RF, firing rates did not deviate from baseline during the stimulus or delay periods, regardless of trial outcome. In other words, the neurons also did not behave as though it were a three-choice task, suggesting that the intermediate activity was truly a signature of lower confidence in the binary perceptual decision.

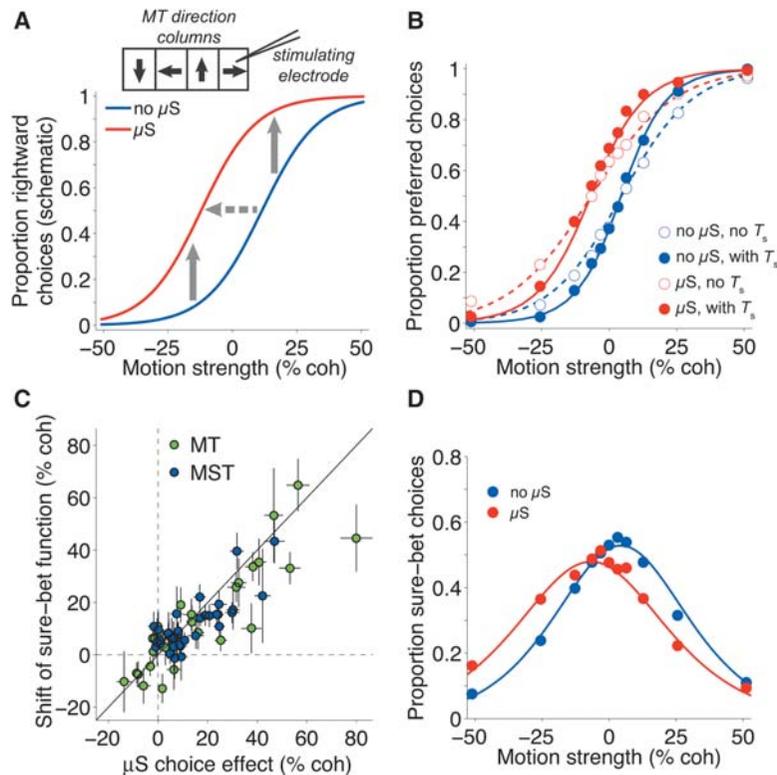
<sup>6</sup>Despite these reassurances, the task has a drawback that choice and confidence are not overtly reported on the same trials. This was rectified in subsequent experiments in humans (Kiani et al. 2014a; see below) and monkeys (Kiani et al. 2011). For discussions of the pros and cons of different methods for assessing confidence and metacognition in animals, see Smith et al. (2012), Kepecs and Mainen (2012), Hampton (2009), Middlebrooks and Sommer (2011), and Terrace and Son (2009).

## A CAUSAL TEST OF THE THEORY

These findings were consistent with the hypothesis that a common mechanism of bounded evidence accumulation underlies all three pillars of choice behavior in decision theory: choice (i.e., accuracy), decision time, and confidence. Conveniently, our understanding of the neural representation of sensory evidence in the motion discrimination task allowed a powerful causal test of the hypothesis. The momentary evidence for this task is known to reside in motion-sensitive regions of extrastriate visual cortex, especially the middle temporal (MT) and medial superior temporal (MST) areas. Electrical microstimulation ( $\mu$ S) of MT/MST biases monkeys’ choices (Salzman et al. 1990; Celebrini and Newsome 1995) and RTs (Ditterich et al. 2003) in a manner consistent with a change in the strength of momentary evidence (Fig. 3A). Based on these studies, we reasoned that  $\mu$ S combined with PDW would provide a sensitive test of the idea that confidence arises from the same neural mechanism as choice and RT.

Monkeys performed the direction discrimination task with PDW as in the previous study (Kiani and Shadlen 2009), but during motion viewing, on half the trials, we applied  $\mu$ S to a cluster of neurons selective for one of the two alternatives. As previously shown (Salzman et al. 1990, 1992; Celebrini and Newsome 1995),  $\mu$ S caused the monkey to choose the preferred direction of the stimulated neurons more frequently, manifesting as a horizontal shift of the choice function (Fig. 3B). The key finding was that  $\mu$ S altered confidence in a similar way, mimicking a change in the visual motion itself (Fetsch et al. 2014). Specifically, the bell-shaped function relating the proportion of sure-bet choices to motion strength (Fig. 3D) shifted along the motion axis in a similar fashion as the choice function, and the effects on choice and confidence were well-correlated across sessions (Fig. 3C). Moreover,  $\mu$ S did not prevent the improvement in sensitivity when the sure bet was offered but waived (Fig. 3B). This implies that the brain’s ability to assess the reliability of evidence and opt out accordingly was not disrupted, despite the artificial pattern of activity induced by microstimulation. In terms of the model, this means  $\mu$ S did not alter the internal mapping between accumulated evidence and confidence (Fig. 2B), nor the criteria applied to that mapping for opting out.

Taken together with previous work (Salzman et al. 1990; Ditterich et al. 2003; Hanks et al. 2006), this study provides a quantitative, causal link between the activity of neurons in visual cortex and the three core behavioral indices in decision-making. The results support the idea that integration of noisy sensory signals gives rise to a common quantity that underlies choice, decision time, and confidence. That one can perturb all three is not surprising, but it is remarkable that the perturbation can be quantitatively reconciled such that the amount of change in choice is commensurate with specific changes in PDW (Fetsch et al. 2014) and RT (Ditterich et al. 2003; Hanks et al. 2006). Because  $\mu$ S alters neural activity in a manner that is unlike natural vision, one might have expected it to generate an anomalous experience



**Figure 3.** MT/MST microstimulation ( $\mu\text{S}$ ) affects confidence in a manner similar to a change in visual motion. (A) (Top) Illustration of experimental approach (Salzman et al. 1990) that exploits the columnar architecture of MT to inject a relatively pure directional signal (in this case, rightward). Red and blue curves show schematically the proportion of rightward choices by the monkey as a function of motion strength (percent coherence; positive = rightward), for  $\mu\text{S}$  and no- $\mu\text{S}$  trials, respectively. Electrical  $\mu\text{S}$  in this example causes the monkey to choose rightward more often (upward arrows). The effect on choice and RT (Ditterich et al. 2003, not shown) mimics a horizontal shift along the motion strength axis (dashed arrow). (B) Proportion of preferred-direction choices as a function of motion strength (positive coh = preferred direction of neurons at the stimulation site) for 63 sessions in two monkeys (Fetsch et al. 2014), plotted separately for the four conditions of the  $2 \times 2$  design:  $\mu\text{S}$  present (red) or absent (blue), and with (solid) or without (dashed) the sure bet available. (C) Magnitude of  $\mu\text{S}$  effects on choice and confidence (horizontal “shifts”)—plotted separately for stimulation sites in area MT (green) and MST (blue)—were correlated across sessions ( $r = 0.87$ ). Error bars represent SE of the fitted parameters quantifying the shift. (D) Proportion of sure bet ( $T_s$ ) choices as a function of motion strength for all sessions. Smooth curves in B and D represent fits/predictions of the bounded evidence-accumulation model (see text). (B–D, Reproduced from Fetsch et al. 2014, with permission from Elsevier.)

leading to a net decrease in confidence. It is also conceivable that, having been compelled to choose the preferred direction of the stimulated neurons more often, the monkey would have experienced an unusually high degree of confidence in those choices. Instead, the model—which implements  $\mu\text{S}$  essentially as an offset to motion strength—was able to *predict* the shift of the choice function and the degree of improvement on waived-sure-bet trials (Fig. 3B, red curves) based on a fit to the remaining observations.

One may notice in Figure 3D that the sure-bet curve on  $\mu\text{S}$  trials (red) is not a translated replica of the no- $\mu\text{S}$  curve (blue). There is a slight decrease in the probability of a sure-bet choice, most prominently at the weaker motion strengths, indicating an increase in confidence with  $\mu\text{S}$ . The most plausible explanation for this feature of the data is a counterintuitive one: It suggests  $\mu\text{S}$  adds a small amount of noise to the sensory representation. Of course an increase in noise should reduce the animal’s accuracy (sensitivity), and indeed this is consistent with

the subtle flattening<sup>7</sup> of the choice function in Figure 3B (red vs. blue solid curves). But why should this cause an increase in confidence? In the context of evidence accumulation, increasing the variability of momentary evidence leads the DV to “diffuse” further away from the starting point, which in our model (Fig. 2B) implies greater confidence and fewer sure-bet choices. Notably, this explanation requires that the brain does not “know” that the variance of the evidence has changed: It applies the same mapping to the accumulated evidence on all trials, hence greater confidence with  $\mu\text{S}$  despite lower sensitivity. This paradoxical result is also predicted

<sup>7</sup>This effect is not obvious by eye but is statistically reliable. A similar flattening was observed by Salzman et al. (1992), an effect they attributed to the spread of stimulation current to neighboring columns with different preferred directions of motion. Although they referred to this as adding noise, it is better construed as a reduction of the signal strength. This phenomenon may have accounted for part of the decrease in sensitivity we observed, but not the effect on confidence, which suggests an increase in noise itself.

by SDT-based models and at least one predecessor (Cartwright and Festinger 1943; Vickers 1979), and similar observations have been made in human subjects undergoing transcranial magnetic stimulation (Rahnev et al. 2012). We are currently exploring this phenomenon further by measuring MT/MST activity and PDW behavior in response to a visual manipulation designed to mimic the effect of adding noise to the DV (Zylberberg et al. 2014).

### ELAPSED DECISION TIME INFORMS THE SENSE OF CONFIDENCE

The success of the modeling exercise in the microstimulation study (Fetsch et al. 2014) and earlier LIP recording study (Kiani and Shadlen 2009) was due to a critical feature of the model: The mapping between evidence and degree of belief depends on elapsed deliberation time. The time dependence comes from the mixture of motion strengths in the experiment: The subject does not know the reliability of evidence a priori, and so the passage of time itself becomes indicative of weaker evidence. However, owing to a lack of RT measurements in these studies, a more direct test of the role of time was warranted.

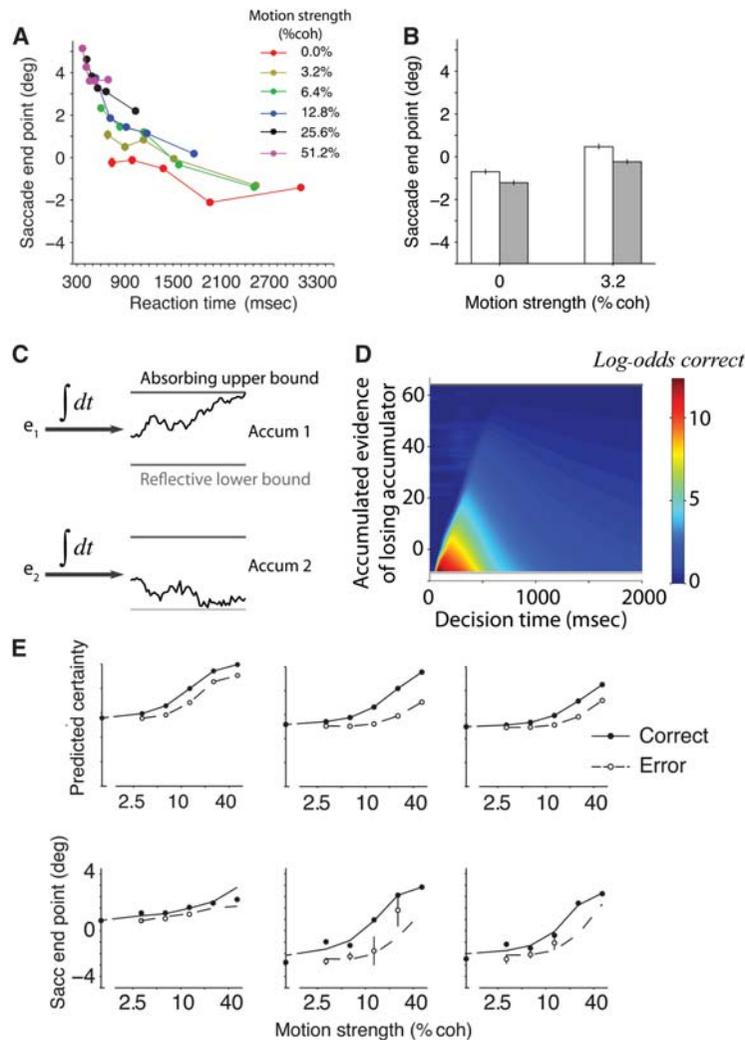
Kiani et al. (2014a), therefore, conducted a psychophysical experiment in which all three components of the decision were measured simultaneously. In this study, human subjects performed a direction discrimination task and could report their decision (up or down) as soon as they were ready. They did so with an upward or downward eye movement, but were instructed to vary the horizontal position of their saccadic end point based on how confident they felt in the decision. The pattern of saccadic end points indicated that, for all six subjects, motion strength and RT had independent leverage on confidence ratings (Fig. 4A). Because these three measures are correlated with accuracy and with each other, establishing an independent role for time required a stimulus manipulation that could prolong the decision without affecting accuracy. This was achieved by presenting a reversal of the animation sequence, which effectively undid evidence that had been supplied in the preceding ~160 msec. The manipulation had the effect of prolonging the decision without a significant change in accuracy, and on those trials the subjects reported lower confidence (Fig. 4B). The result suggests that the brain uses elapsed time as a source of information about signal reliability, even when controlling for the amount of accumulated evidence (Kiani et al. 2014a).

Explaining these findings required a more complicated model than the one-dimensional bounded accumulation process depicted in Figure 2A. This is because for a single accumulator with symmetric bounds, decisions are always terminated at a particular level of evidence, leaving nothing to covary with accuracy (and confidence) except RT. For example, if the bounds were flat (i.e., stationary as a function of time) the only possible arbiter of confidence would be decision time. Confidence would not

depend on motion strength,<sup>8</sup> but this is not supported by the data. A solution (which happens to be more biologically plausible) is to posit separate accumulators for different options—a race between two diffusion processes (Fig. 4C; Mazurek et al. 2003; Churchland et al. 2008; Niwa and Ditterich 2008)—and a calculation of certainty based on the status of the losing accumulator (Fig. 4D). Intuitively, if the race between accumulators was a close call, it means the net evidence supporting the chosen option was weak and we should be less confident. In some ways the model resembles Vickers' BE hypothesis (Fig. 1C), except that negative increments are allowed, and the races are anticorrelated, because stimulus evidence for left is negative evidence for right. The degree of this anticorrelation effectively determines the degree to which time and motion strength account for confidence.

A critical feature of this study (Kiani et al. 2014a) was the simultaneous report of choice and confidence by a single ballistic (saccadic) eye movement, ensuring that both were guided by the same information. In this case the model predicts—and the data show—that confidence is an increasing function of evidence strength on both correct and error trials (Fig. 4E; Kiani et al. 2014a). This contradicts a key prediction of SDT-based models, that confidence on error trials should decrease with evidence strength or discriminability (Fig. 5A,B). The logic of the SDT prediction is that for strong stimuli the distributions representing the two alternatives are only slightly overlapping, such that any error must have arisen from an observation very close to the criterion, hence with low confidence (see also Vickers et al. 1985). We think this pattern has been oversold as a defining feature of confidence and its neural correlates (Insabato et al. 2010; Rolls et al. 2010; Komura et al. 2013). Consider that most experiments solicit a confidence judgment after the decision has been reported (Baranski and Petrusic 1998), allowing information “in the pipeline” to contribute to confidence but not the choice (Pleskac and Busemeyer 2010). The use of such information has been demonstrated in the study of changes of mind (Fig. 6A; Rabbitt 1966; Resulaj et al. 2009; Song and Nakayama 2009; Kiani et al. 2014b). This would allow the rare error associated with an easy stimulus to undergo correction, leading to low confidence on errors with stronger stimuli. In fact, we can reproduce the pattern predicted by SDT in a simple simulation of bounded accumulation in which the DV is allowed to continue evolving and affect confidence after the decision bound is reached (Fig. 6B). Thus, the superficial similarity between SDT and some published experimental results does not indicate validity of SDT-based models over SA-based alternatives. When confidence and choice are reported at different times, such results must be interpreted with caution to avoid incorrect conclusions about the computation of certainty and its neural mechanisms.

<sup>8</sup>This prediction is surprisingly robust even if the bounds collapse over time, which they do, typically.



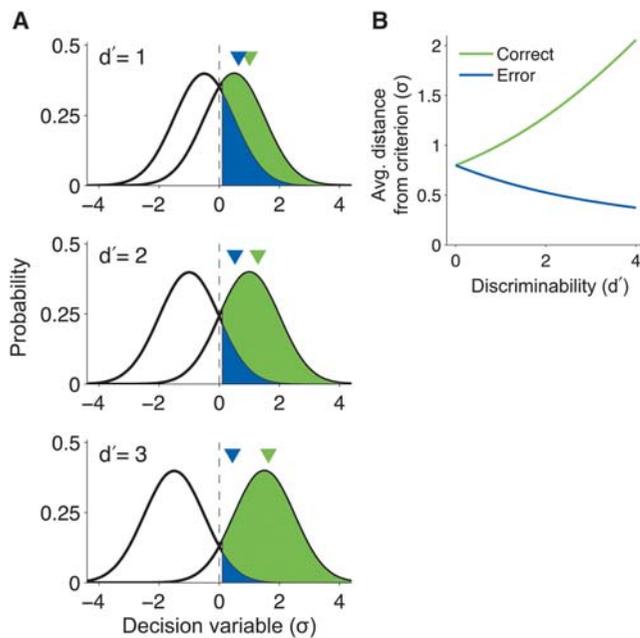
**Figure 4.** Confidence is informed by evidence and decision time. (A) Across-subject average ( $N = 6$ ) of saccade end points, a psychophysical report of confidence, plotted as a function of RT and color-coded by motion strength. Positive end points indicate greater confidence. (B) Average saccade end point on trials with (gray) and without (white) a “reverse pulse” that increased RT without a significant change in accuracy (not shown). Error bars represent SEM. (C) In an extension of the model in Figure 2A, two accumulators compete by integrating noisy momentary evidence in favor of the two choices. The accumulator that first reaches the absorbing bound dictates the choice and decision time. (D) Confidence in the model is governed by the status of the losing accumulator, which at any given time is associated with a particular log-odds of being correct (color). Each level of evidence is associated with less confidence as time elapses. (E) Theoretical predictions (*top row*) and model fits (*bottom row*) of the confidence, or certainty, for three representative subjects, plotted as a function of motion strength and separated by trial outcome (correct vs. error). Model parameters were first estimated by fitting each subject’s RT distributions irrespective of choice and confidence, and then used to predict the general form of the certainty functions and to fit their observed saccade end points. (Reproduced with permission from Kiani et al. 2014a.)

## CONCLUSIONS AND FUTURE DIRECTIONS

It seems neuroscience may soon be in a position to help resolve decades of debate in mathematical psychology about the mechanism of confidence judgments. This optimism should be tempered somewhat by the reality that the brain is flexible enough to use different mechanisms depending on context or past experience. Nevertheless, a few core principles are emerging that should help frame the discussion and guide future investigations. From our perspective, these principles include the use of time as a proxy for evidence reliability (see also Hanks

et al. 2011) and the idea that a common neural mechanism may be responsible for choice, RT, and confidence.

Some key unanswered questions remain. We think confidence involves the conversion of a DV to degree of belief based on both the magnitude of the DV and deliberation time. But how does this happen? Along with accumulated evidence, neurons in LIP represent information about elapsed time (Leon and Shadlen 2003; Janssen and Shadlen 2005; Maimon and Assad 2006) and thus could participate in computations involving both quantities. On the other hand, temporal signals related to urgency (Churchland et al. 2008; Cisek et al. 2009) or dynamic

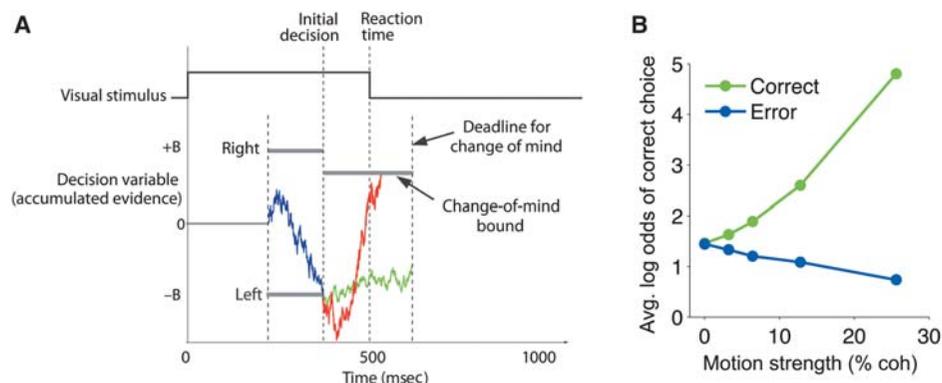


**Figure 5.** Confidence as the distance from criterion in signal-detection theory (SDT). (A) The overlapping Gaussians represent probability distributions of a noisy observation under two competing hypotheses (i.e., a leftward or rightward stimulus was present). The panels depict different levels of discriminability defined by the difference in means of the distributions in units of their standard deviation ( $d'$ ). An observation to the right of the decision criterion (dashed line) will be correct if arising from the rightward distribution (green) or will be an error if from the leftward distribution (blue). Confidence in these choices could be computed as the distance from the criterion, the means of which are shown by the inverted triangles. Depiction is consistent with a hypothetical two-interval task, but the same logic applies in the single-interval case. (B) The average distance from criterion increases with discriminability for correct choices (green), but decreases for errors (blue). (See also Fig. 8g in Kim and Shadlem [1999].)

bias (Hanks et al. 2011) may need to be discounted or subtracted at later stages in order to recover the true level of evidence that informs confidence. Moreover, LIP represents a DV for decisions about where to move the eyes, but for other effectors or types of decisions, the readout mechanism that establishes a sense of confidence must be targeted elsewhere.

Interestingly, our microstimulation results (Fetsch et al. 2014) suggest there is at least one signal that is not subtracted prior to establishing a level of certainty. When we cause monkeys to choose the preferred direction of the stimulated neurons more frequently, they adopt a

compensatory “null choice” bias to respond in the opposite direction (Fig. 5B, lower; Salzman et al. 1992). This is believed to reflect a stubborn prior: The brain does not adjust its expectation that the two choice targets are equally likely to be rewarded. Thus, when faced with a preponderance of, say, rightward choices induced by  $\mu S$ , it adds a leftward bias so that the proportion of each choice remains  $\sim 50\%$ . The compensatory bias may therefore be akin to a bias induced by asymmetry in reward expectation (Feng et al. 2009). For example, on trials without  $\mu S$  in which the monkey viewed 0% coherent motion, he tended to choose the null direction (Fig.



**Figure 6.** Confidence on errors could in some cases be explained by a process akin to change-of-mind. (A) Depiction of a hypothetical “rightward” trial. Because of noise, the accumulated evidence first reaches the bound signaling “left,” and movement is initiated toward the wrong choice target. However, evidence in the processing pipeline continues to accrue favoring rightward, and if the decision is reported with a nonballistic movement (i.e., a reach of the arm), this additional evidence can cause a revision of the initial movement trajectory toward the correct choice. (B) If the initial choice in A is not revised, but the experiment solicits a confidence report following the choice, the extra evidence in the pipeline would lead to lower confidence on errors when evidence is strong, as observed in some experiments and asserted as a signature of confidence within an SDT framework. The graph (compare with Fig. 5B) shows results of a simulated one-dimensional bounded accumulation process where evidence was allowed to accumulate for 250 msec after bound crossing. Symbols show average log-odds associated with the final level of evidence as a function of motion strength on correct trials (green) and errors (blue). (A, Modified, with permission, from Resulaj et al. 2009.)

3B, blue curves shifted to the right), yet we do not believe he perceives motion in this direction. He is simply biased. It turns out this bias is also reflected in PDW: the monkey opts out most frequently for stimuli supporting an equal probability of choosing either direction (Fig. 3D, blue curve shifted to the right). In terms of the model, this says that the mapping (Fig. 2B) does not shift to account for the bias, just as it does not shift in response to  $\mu$ S. Thus, a distinction to keep in mind is whether a behavioral assay measures confidence about the state of the world (is the motion leftward?) or about whether the decision is appropriate (should I choose leftward?). PDW can indicate confidence in the state of the world only to the extent that subjects' expectations reflect it. For most of our experiments the connection is tight, but there are many ways it can be loosened.

Beyond this distinction, other forms of confidence are involved in foraging (Hayden et al. 2011), monitoring performance and conflict (Stuphorn et al. 2000; Botvinick et al. 2001; Ito et al. 2003), and ranking unchosen options (Hayden et al. 2009; Rushworth et al. 2012). These processes may involve neural representations of confidence-like quantities that are disentangled from the decision process (Lak et al. 2014) and therefore independent of modality and effector. We emphasize that the mechanism we propose for computing confidence in a perceptual decision is not incompatible with such a task-independent representation (de Gardelle and Mamassian 2014; Heereman and Heekeren 2014). Rather our experiments point toward a stage of processing in which certainty emerges directly from a decision process—consisting of operations on evidence and (in many cases) elapsed time—which may then support elaboration to a broader array of cognitive functions such as those listed above. Exactly where and how these and other forms of (un)certainly (e.g., Bromberg-Martin and Hikosaka 2011; Monosov and Hikosaka 2013) are transformed and combined in the brain will be an area of great future interest for experimenters.

José Saramago once wrote that “strictly speaking, we do not make decisions, decisions make us.” Indeed, one might say that it is through our decisions that we differentiate ourselves as moral agents (Kane 2002), and it is through this window that neuroscience illuminates the philosopher's study of ethics (Shadlen and Roskies 2012). More generally, we feel that decision-making offers a window on the neurobiological underpinnings of a variety of aspects of cognition that have little to do with decision-making. In particular, in the study of confidence we encounter the topic of metacognition. When the monkey chooses or waives the sure bet, it is now indicating something about the decision process—something beyond the choice itself. One might be tempted to dismiss the metacognitive designation in light of the nuts-and-bolts account of the process emerging from our experiments. We would regard that as mistaken, however, for explaining a phenomenon, especially in cognitive neuroscience, is not explaining that phenomenon away. Just as vision arises via neural mechanisms, so too do many of our treasured cognitive functions.

Turning to vision itself, it is tempting to ask if the monkeys experience electrical microstimulation as vision. Pioneers of this experimental approach (Salzman et al. 1992; Romo et al. 1998) speculated that the subjective experience during  $\mu$ S resembles that of natural sensory input, an intriguing proposal given that the manipulation directly affects only a small portion of the circuit and does so in an artificial manner. In our experiment (Fetsch et al. 2014), this question is really whether the experience of  $\mu$ S plus random-dot motion is like seeing another random-dot stimulus (e.g., one that appears more “rightward”-ish). We would like to ask a monkey what it feels like on the trials with  $\mu$ S. When we displayed a stimulus with no net motion accompanied by  $\mu$ S of rightward-preferring neurons, did this feel like seeing motion to the right? We cannot pose these questions to a monkey, but the PDW paradigm might come as close as possible. Consider that the monkeys had an opportunity to use the sure bet to tell us something is peculiar on stimulation trials. They did not do that, yet we know that they will indicate a decrease in confidence if we increase the current (Fetsch et al. 2014). More importantly,  $\mu$ S does not appear to affect the way that the brain consults the mapping of DV (and elapsed time) to render a metacognitive decision to choose or forgo the sure bet. This is a strong test of what it means to experience visual motion plus artificial  $\mu$ S as equivalent to visual motion plus a variant of visual motion. It does not rise to the subjective, personal aspect of “feels like,” but it comes close.

## ACKNOWLEDGMENTS

The authors' research highlighted in this essay was supported by the Howard Hughes Medical Institute (HHMI), the Human Frontiers Research Program (HFSP), National Eye Institute Grant EY11378 to M.N.S., National Eye Institute Core Grant P30EY01730, National Center for Research Resources Grant RR00166, a Sloan Research Fellowship to R.K., and a Whitehall Foundation Research Grant to R.K. We thank Ariel Zylberberg, Bill Newsome, and Daniel Wolpert for helpful discussions.

## REFERENCES

- Audley RJ. 1960. A stochastic model for individual choice behavior. *Psychol Rev* **67**: 1–15.
- Baranski JV, Petrusic WM. 1998. Probing the locus of confidence judgments: Experiments on the time to determine confidence. *J Exp Psychol Hum Percept Perform* **24**: 929–945.
- Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. *Psychol Rev* **108**: 624–652.
- Bromberg-Martin ES, Hikosaka O. 2011. Lateral habenula neurons signal errors in the prediction of reward information. *Nat Neurosci* **14**: 1209–1216.
- Busemeyer JR. 1985. Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *J Exp Psychol Learn Mem Cogn* **11**: 538–564.
- Cartwright D, Festinger L. 1943. A quantitative theory of decision. *Psychol Rev* **50**: 595–621.

- Celebrini S, Newsome WT. 1995. Microstimulation of extrastriate area MST influences performance on a direction discrimination task. *J Neurophysiol* **73**: 437–448.
- Churchland AK, Kiani R, Shadlen MN. 2008. Decision-making with multiple alternatives. *Nat Neurosci* **11**: 693–702.
- Churchland AK, Kiani R, Chaudhuri R, Wang XJ, Pouget A, Shadlen MN. 2011. Variance as a signature of neural computations during decision making. *Neuron* **69**: 818–831.
- Cisek P, Puskas GA, El-Murr S. 2009. Decisions in changing conditions: The urgency-gating model. *J Neurosci* **29**: 11560–11571.
- DasGupta S, Ferreira CH, Miesenböck G. 2014. FoxP influences the speed and accuracy of a perceptual decision in *Drosophila*. *Science* **344**: 901–904.
- Dayan P, Kakade S, Montague PR. 2000. Learning and selective attention. *Nat Neurosci* **3**: 1218–1223.
- de Gardelle V, Mamassian P. 2014. Does confidence use a common currency across two visual tasks? *Psychol Sci* **25**: 1286–1288.
- Ditterich J, Mazurek ME, Shadlen MN. 2003. Microstimulation of visual cortex affects the speed of perceptual decisions. *Nat Neurosci* **6**: 891–898.
- Drugowitsch J, Moreno-Bote R, Pouget A. 2014. Relation between belief and performance in perceptual decision making. *PLoS ONE* **9**: e96511.
- Feng S, Holmes P, Rorie A, Newsome WT. 2009. Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Comput Biol* **5**: e1000284.
- Ferrell WR, McGoey PJ. 1980. A model of calibration for subjective probabilities. *Organ Behav Hum Perf* **26**: 32–53.
- Festinger L. 1943. Studies in decision: I. Decision-time, relative frequency of judgment and subjective confidence as related to physical stimulus difference. *J Exp Psychol* **32**: 291–306.
- Fetsch CR, Kiani R, Newsome WT, Shadlen MN. 2014. Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* **83**: 797–804.
- Fullerton GS, Cattell JM. 1892. *On the perception of small differences: With special reference to the extent, force and time of movement*. Philosophy Series No 2. University of Pennsylvania, Philadelphia.
- Garrett HE. 1922. A study of the relation of accuracy to speed. *Arch Psychol* **8**: 1–105.
- Gold JJ, Shadlen MN. 2001. Neural computations that underlie decisions about sensory stimuli. *Trends Cogn Sci* **5**: 10–16.
- Gold JJ, Shadlen MN. 2007. The neural basis of decision making. *Annu Rev Neurosci* **30**: 535–574.
- Hampton RR. 2009. Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comp Cogn Behav Rev* **4**: 17–28.
- Hanes DP, Schall JD. 1996. Neural control of voluntary movement initiation. *Science* **274**: 427–430.
- Hanks TD, Ditterich J, Shadlen MN. 2006. Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nat Neurosci* **9**: 682–689.
- Hanks TD, Mazurek ME, Kiani R, Hopp E, Shadlen MN. 2011. Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J Neurosci* **31**: 6339–6352.
- Hayden BY, Pearson JM, Platt ML. 2009. Fictive reward signals in the anterior cingulate cortex. *Science* **324**: 948–950.
- Hayden BY, Pearson JM, Platt ML. 2011. Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* **14**: 933–939.
- Heekeren HR, Marrett S, Ungerleider LG. 2008. The neural systems that mediate human perceptual decision making. *Nat Rev Neurosci* **9**: 467–479.
- Heereman J, Heekeren H. 2014. A task-independent neural representation of subjective certainty. *Program No. 587.10. 2014 Neuroscience Meeting Planner*. Society for Neuroscience, Washington, DC.
- Henmon VAC. 1911. The relation of the time of a judgment to its accuracy. *Psychol Rev* **18**: 186.
- Higham PA. 2007. No special K! A signal detection framework for the strategic regulation of memory accuracy. *J Exp Psychol Gen* **136**: 1–22.
- Horowitz GD, Newsome WT. 2001. Target selection for saccadic eye movements: Prelude activity in the superior colliculus during a direction-discrimination task. *J Neurophysiol* **86**: 2543–2558.
- Huk AC, Shadlen MN. 2005. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J Neurosci* **25**: 10420–10436.
- Insabato A, Pannunzi M, Rolls ET, Deco G. 2010. Confidence-related decision making. *J Neurophysiol* **104**: 539–547.
- Irwin FW, Smith WAS, Mayfield JF. 1956. Tests of two theories of decision in an “expanded judgment” situation. *J Exp Psychol* **51**: 261–268.
- Ito S, Stuphorn V, Brown JW, Schall JD. 2003. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* **302**: 120–122.
- Janssen P, Shadlen MN. 2005. A representation of the hazard rate of elapsed time in macaque area LIP. *Nat Neurosci* **8**: 234–241.
- Johnson DM. 1939. Confidence and speed in the two-category judgement. *Arch Psychol* **34**: 1–53.
- Kane R. 2002. Free will: New directions for an ancient problem. In *Free will* (ed. Kane R), pp. 222–246. Blackwell, Oxford.
- Kellogg WN. 1931. The time of judgment in psychometric measures. *Am J Psychol* **43**: 65–86.
- Kepecs A, Mainen ZF. 2012. A computational framework for the study of confidence in humans and animals. *Philos Trans R Soc Lond B Biol Sci* **367**: 1322–1337.
- Kepecs A, Uchida N, Zariwala HA, Mainen ZF. 2008. Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**: 227–231.
- Kiani R, Shadlen MN. 2009. Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* **324**: 759–764.
- Kiani R, Hanks T, Shadlen M. 2008. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci* **28**: 3017–3029.
- Kiani R, Shadlen MN, Newsome WT. 2011. Belief and time: Response dynamics of lateral intraparietal neurons represent choice certainty. *Program No. 17.06. 2011 Neuroscience Meeting Planner*. Society for Neuroscience, Washington, DC.
- Kiani R, Corthell L, Shadlen MN. 2014a. Choice certainty is informed by both evidence and decision time. *Neuron* **84**: 1329–1342.
- Kiani R, Cueva CJ, Reppas JB, Newsome WT. 2014b. Dynamics of neural population responses in prefrontal cortex indicate changes of mind on single trials. *Curr Biol* **24**: 1542–1547.
- Kim J-N, Shadlen MN. 1999. Neural correlates of a decision in the dorsolateral prefrontal cortex of the macaque. *Nat Neurosci* **2**: 176–185.
- Kinchla RA, Collyer CE. 1974. Detecting a target letter in briefly presented arrays: A confidence rating analysis in terms of a weighted additive effects model. *Percept Psychophys* **16**: 117–122.
- Ko Y, Lau H. 2012. A detection theoretic explanation of blindsight suggests a link between conscious perception and metacognition. *Philos Trans R Soc Lond B Biol Sci* **367**: 1401–1411.
- Komura Y, Nikkuni A, Hirashima N, Uetake T, Miyamoto A. 2013. Responses of pulvinar neurons reflect a subject’s confidence in visual categorization. *Nat Neurosci* **16**: 749–755.
- Krajbich I, Rangel A. 2011. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc Natl Acad Sci* **108**: 13852–13857.
- 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 times*. Academic, Oxford.

- Leon MI, Shadlen MN. 2003. Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron* **38**: 317–327.
- Link SW, Heath RA. 1975. A sequential theory of psychological discrimination. *Psychometrika* **40**: 77–105.
- Luce RD. 1986. *Response times: Their role in inferring elementary organization*. Oxford University Press, Oxford, UK.
- Maimon G, Assad JA. 2006. A cognitive signal for the proactive timing of action in macaque LIP. *Nat Neurosci* **9**: 948–955.
- Mazurek ME, Roitman JD, Ditterich J, Shadlen MN. 2003. A role for neural integrators in perceptual decision making. *Cereb Cortex* **13**: 1257–1269.
- Middlebrooks PG, Sommer MA. 2011. Metacognition in monkeys during an oculomotor task. *J Exp Psychol Learn Mem Cogn* **37**: 325–337.
- Monosov IE, Hikosaka O. 2013. Selective and graded coding of reward uncertainty by neurons in the primate anterodorsal septal region. *Nat Neurosci* **16**: 756–762.
- Niwa M, Ditterich J. 2008. Perceptual decisions between multiple directions of visual motion. *J Neurosci* **28**: 4435–4445.
- Norman DA, Wickelgren WA. 1969. Strength theory of decision rules and latency in retrieval from short-term memory. *J Math Psychol* **6**: 192–208.
- Peirce CS, Jastrow J. 1884. On small differences in sensation. *Mem Natl Acad Sci* **3**: 73–83.
- Peterson WW, Birdsall TG. 1953. *The theory of signal detectability*. Technical Report No. 13, Electronic Defense Group, University of Michigan, Ann Arbor, MI.
- Pierrel R, Murray CS. 1963. Some relationships between comparative judgment, confidence, and decision-time in weight-lifting. *Am J Psychol* **76**: 28–38.
- Pleskac TJ, Busemeyer JR. 2010. Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychol Rev* **117**: 864–901.
- Ploran EJ, Tremel JJ, Nelson SM, Wheeler ME. 2011. High quality but limited quantity perceptual evidence produces neural accumulation in frontal and parietal cortex. *Cereb Cortex* **21**: 2650–2662.
- Rabbitt PM. 1966. Error correction time without external error signals. *Nature* **212**: 438.
- Rahnev D, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H. 2011. Attention induces conservative subjective biases in visual perception. *Nat Neurosci* **14**: 1513–1515.
- Rahnev DA, Maniscalco B, Luber B, Lau H, Lisanby SH. 2012. Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J Neurophysiol* **107**: 1556–1563.
- Ratcliff R. 1978. A theory of memory retrieval. *Psychol Rev* **85**: 59–108.
- Ratcliff R, Smith PL. 2004. A comparison of sequential sampling models for two-choice reaction time. *Psychol Rev* **111**: 333–367.
- 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.
- Ratcliff R, Cherian A, Segraves M. 2003. A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *J Neurophysiol* **90**: 1392–1407.
- Reddi BA, Asrress KN, Carpenter RH. 2003. Accuracy, information, and response time in a saccadic decision task. *J Neurophysiol* **90**: 3538–3546.
- Reed JB. 1951. *The speed and accuracy of discriminating differences in hue, brilliance, area, and shape*. Technical Report, Special Devices Center, US Navy.
- Resulaj A, Kiani R, Wolpert DM, Shadlen MN. 2009. Changes of mind in decision-making. *Nature* **461**: 263–266.
- Roitman JD, Shadlen MN. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* **22**: 9475–9489.
- Rolls ET, Grabenhorst F, Deco G. 2010. Decision-making, errors, and confidence in the brain. *J Neurophysiol* **104**: 2359–2374.
- Romo R, Hernández A, Zainos A, Salinas E. 1998. Somatosensory discrimination based on cortical microstimulation. *Nature* **392**: 387–390.
- 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**: 165–175.
- Rushworth MF, Kolling N, Sallet J, Mars RB. 2012. Valuation and decision-making in frontal cortex: One or many serial or parallel systems? *Curr Opin Neurobiol* **22**: 946–955.
- Salzman CD, Britten KH, Newsome WT. 1990. Cortical microstimulation influences perceptual judgements of motion direction. *Nature* **346**: 174–177.
- Salzman CD, Murasugi CM, Britten KH, Newsome WT. 1992. Microstimulation in visual area MT: Effects on direction discrimination performance. *J Neurosci* **12**: 2331–2355.
- Scott RB, Dienes Z, Barrett AB, Bor D, Seth AK. 2014. Blind insight: Metacognitive discrimination despite chance task performance. *Psychol Sci* **25**: 2199–2208.
- Seward GH. 1928. Recognition time as a measure of confidence, an experimental study of redintegration. *Arch Psychol* **16**: 1–50.
- Shadlen MN, Kiani R. 2013. Decision making as a window on cognition. *Neuron* **80**: 791–806.
- Shadlen MN, Newsome WT. 1996. Motion perception: Seeing and deciding. *Proc Natl Acad Sci* **93**: 628–633.
- Shadlen MN, Roskies AL. 2012. The neurobiology of decision-making and responsibility: Reconciling mechanism and mindedness. *Front Neurosci* **6**: 56.
- Smith JD. 2009. The study of animal metacognition. *Trends Cogn Sci* **13**: 389–396.
- Smith PL, Ratcliff R. 2004. Psychology and neurobiology of simple decisions. *Trends Neurosci* **27**: 161–168.
- Smith JD, Couchman JJ, Beran MJ. 2012. The highs and lows of theoretical interpretation in animal-metacognition research. *Philos Trans R Soc Lond B Biol Sci* **367**: 1297–1309.
- Song JH, Nakayama K. 2009. Hidden cognitive states revealed in choice reaching tasks. *Trends Cogn Sci* **13**: 360–366.
- Stone M. 1960. Models for choice-reaction time. *Psychometrika* **25**: 251–260.
- Stuphorn V, Taylor TL, Schall JD. 2000. Performance monitoring by the supplementary eye field. *Nature* **408**: 857.
- Swets JA, Tanner J, Wilson P, Birdsall TG. 1961. Decision processes in perception. *Psychol Rev* **68**: 301.
- Tanner WP, Swets JA. 1954. A decision-making theory of visual detection. *Psychol Rev* **61**: 401–409.
- Terrace HS, Son LK. 2009. Comparative metacognition. *Curr Opin Neurobiol* **19**: 67–74.
- Treisman M, Faulkner A. 1984. The setting and maintenance of criteria representing levels of confidence. *J Exp Psychol Hum Percept Perform* **10**: 119.
- van Meter D, Middleton D. 1954. Modern statistical approaches to reception in communication theory. *Trans IRE Prof Group Inform Theory* **4**: 119–145.
- Vickers D. 1972. Some general features of perceptual discrimination. In *Psychological aspects of driver behaviour* (ed. Asmussen EG). Institute for Road Safety Research, S.W.O.V., Voorlun, Netherlands.
- Vickers D. 1979. *Decision processes in visual perception*. Academic Press, New York.
- Vickers D. 2001. Where does the balance of evidence lie with respect to confidence? In *Proceedings of the Seventeenth Annual Meeting of the International Society for Psychophysics*, pp. 148–153\*. Pabst Science Publishers, Lengerich, Germany.
- 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**: 179–197.

- Vickers D, Smith P, Burt J, Brown M. 1985. Experimental paradigms emphasising state or process limitations: II. Effects on confidence. *Acta Psychol* **59**: 163–193.
- Volkman J. 1934. The relation of time of judgment to certainty of judgment. *Psychol Bull* **31**: 672–673.
- Zariwala HA, Kepecs A, Uchida N, Hirokawa J, Mainen ZF. 2013. The limits of deliberation in a perceptual decision task. *Neuron* **78**: 339–351.
- Zylberberg A, Barttfeld P, Sigman M. 2012. The construction of confidence in a perceptual decision. *Front Integr Neurosci* **6**: 79.
- Zylberberg A, Fetsch CR, Sigman M, Shadlen MN. 2014. The influence of the strength and variability of sensory evidence on confidence in a perceptual decision. *Program No. 435.16. 2014 Neuroscience Meeting Planner*. Society for Neuroscience, Washington, DC.



## Cold Spring Harbor Symposia on Quantitative Biology

### **Predicting the Accuracy of a Decision: A Neural Mechanism of Confidence**

Christopher R. Fetsch, Roozbeh Kiani and Michael N. Shadlen

*Cold Spring Harb Symp Quant Biol* published online April 28, 2015  
Access the most recent version at doi:[10.1101/sqb.2014.79.024893](https://doi.org/10.1101/sqb.2014.79.024893)

---

<b>P&lt;P</b>	Published online April 28, 2015 in advance of the print volume.
<b>Email alerting service</b>	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or <a href="#">click here</a>

---

---

Advance online articles have not yet appeared in the print volume. Citations to Advance online articles must include the digital object identifier (DOI) and date of initial publication.

---

To subscribe to *Cold Spring Harbor Symposia on Quantitative Biology* go to:  
<http://symposium.cshlp.org/subscriptions>

---