

Effects of Cortical Microstimulation on Confidence in a Perceptual Decision

Christopher R. Fetsch,^{1,4} Roozbeh Kiani,^{2,4} William T. Newsome,³ and Michael N. Shadlen^{1,*}

¹Howard Hughes Medical Institute, Department of Neuroscience and Kavli Institute for Brain Science, Columbia University, New York, NY 10032, USA

²Center for Neural Science, New York University, New York, NY 10003, USA

³Department of Neurobiology and Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

⁴Co-first author

*Correspondence: shadlen@columbia.edu

<http://dx.doi.org/10.1016/j.neuron.2014.07.011>

SUMMARY

Decisions are often associated with a degree of certainty, or confidence—an estimate of the probability that the chosen option will be correct. Recent neurophysiological results suggest that the central processing of evidence leading to a perceptual decision also establishes a level of confidence. Here we provide a causal test of this hypothesis by electrically stimulating areas of the visual cortex involved in motion perception. Monkeys discriminated the direction of motion in a noisy display and were sometimes allowed to opt out of the direction choice if their confidence was low. Microstimulation did not reduce overall confidence in the decision but instead altered confidence in a manner that mimicked a change in visual motion, plus a small increase in sensory noise. The results suggest that the same sensory neural signals support choice, reaction time, and confidence in a decision and that artificial manipulation of these signals preserves the quantitative relationship between accumulated evidence and confidence.

INTRODUCTION

Decision making refers to the process of deliberating toward a commitment to a proposition, hypothesis, or plan of action. Although decisions have a discrete, all or none character—true or false, left or right, option D—they are also associated with a degree of belief that the decision will turn out to be correct. This graded scale of choice certainty, or confidence, affects the way we express our decisions (forcefully or reservedly) and learn from our mistakes and successes. Confidence is critically important when making interrelated decisions without immediate feedback, or when reasoning about a sequence of choices vicariously (Tolman, 1948). For these and other reasons, psychologists have counted confidence among the three main observables of choice behavior (Vickers, 1979), along with the outcome of the decision (correct or incorrect; i.e., accuracy) and the time needed to complete it (reaction time [RT]).

The neural basis of assigning confidence in a decision is not well understood, in part because it is difficult to study in animal models. Recently, methods have been introduced that allow animals to report their confidence, often in the form of a postdecision wager (PDW) (Foote and Crystal, 2007; Hampton, 2001; Kepcs et al., 2008; Kiani and Shadlen, 2009; Middlebrooks and Sommer, 2011; Smith et al., 2008). In one type of PDW, animals indicate their degree of certainty by opting out of the primary behavioral report when a decision is less likely to be successful, instead choosing a guaranteed but smaller (or less preferred) reward (Foote and Crystal, 2007; Hampton, 2001; Kiani and Shadlen, 2009). Monkeys exercise this “sure-bet” option more frequently when the trial is difficult, and they are more accurate when the sure bet is offered and waived versus when it is not offered. This improvement holds within a particular level of stimulus difficulty and even for identical replays of the same stimulus (Kiani and Shadlen, 2009). It suggests that the decision to accept the sure bet is based on an assessment of the reliability of internal sensory evidence, rather than on a simple association with trial difficulty or some property of the stimulus (Smith et al., 2012).

A recent study (Kiani and Shadlen, 2009) reported the activity of decision-related neurons in the lateral intraparietal area (LIP) recorded while monkeys performed a direction discrimination task with PDW. They found that these neurons—previously shown to represent a decision variable (DV) that explains choice and RT (Gold and Shadlen, 2007)—also reflect the degree of confidence in the choice. The results raised the possibility of a common neural mechanism underlying choice, RT, and confidence. This hypothesis makes a clear prediction: if the representation of accumulated evidence used to guide a perceptual decision also supports a degree of confidence in that decision, then a causal manipulation of the evidence will affect PDW in a manner predictable from the effect on choices.

Here we test this hypothesis using electrical microstimulation (μ S). Previous studies showed that μ S of direction-selective neurons in the macaque visual cortex during a direction discrimination task causes monkeys to choose the preferred direction of neurons near the electrode tip more often (Salzman et al., 1990, 1992) and more rapidly (Ditterich et al., 2003). These effects on choice and RT can be quantified as an equivalent change in the motion strength, as though the stimulation effectively added to the visual evidence supporting the preferred direction. Thus, a change in PDW commensurate with the shift

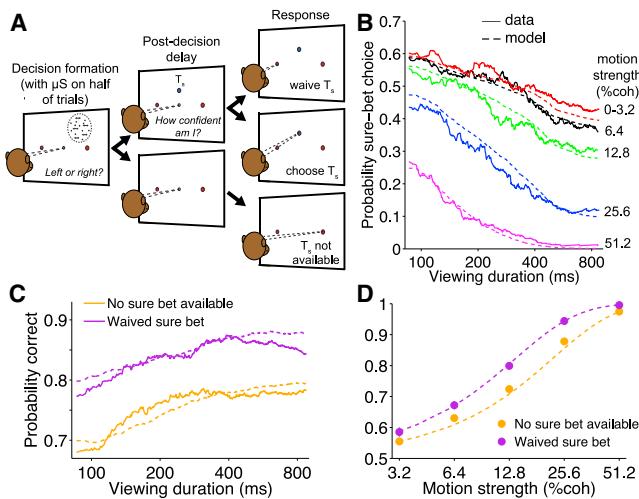


Figure 1. Postdecision Wagering Reflects Confidence in the Motion Decision

(A) Postdecision wagering (PDW) task sequence (see [Experimental Procedures](#)). Red spots indicate direction targets; blue spot is the “sure-bet” target (T_s). (B) Probability of choosing T_s as a function of viewing duration and motion strength (color coded). Combined data from no- μ S trials for two monkeys ($n = 26,924$ trials). Solid traces are running means (proportions) of the data sorted by viewing duration. Dashed traces in all panels are fits to the bounded accumulation model (see text and [Figure 4](#)). (C) Improvement in decision accuracy on no- μ S trials when the sure bet was offered but waived. Solid traces are running means using all nonzero coherences and directions. (D) Same format as (C) but broken down by motion strength (absolute value of coherence) and pooled across viewing durations. Symbols indicate the mean \pm SE.

of the choice function would support the idea that the same sensory signals underlie both the choice and the confidence associated with it. On the other hand, since choice and confidence are known to be dissociated in a variety of settings (Del Cul et al., 2009; Drugowitsch et al., 2014; Kahneman et al., 1982; Komura et al., 2013; Lau and Passingham, 2006; Rahnev et al., 2012; Rounis et al., 2010), we might expect artificial stimulation to induce a discrepancy between the two. Indeed, the effect of μ S on neuronal circuits is unlike anything elicited through natural vision (Histed et al., 2009; Logothetis et al., 2010). Nonetheless, here we show that μ S affects confidence much like a change in the visual stimulus, consistent with a common mechanism characterized by bounded accumulation of evidence.

RESULTS

We trained two rhesus monkeys on a two-alternative direction discrimination task with PDW ([Figure 1A](#); [Experimental Procedures](#)). The monkeys were required to decide between the direction preferred by neurons near the stimulating electrode and the opposite “null” direction and to indicate this choice after a memory delay. Monkeys were rewarded for correct choices and randomly on the neutral (0% coherence) stimulus. During the memory delay, the monkey was sometimes offered a third alternative (the sure-bet target [T_s]) to opt out of the high-stakes direction decision and receive a guaranteed but smaller reward. Monkeys

chose T_s more frequently for shorter viewing durations and weaker motion strengths ([Figure 1B](#)) and showed greater accuracy on waived- T_s trials compared to when T_s was unavailable ([Figures 1C and 1D](#)). [Figure 1C](#) also reveals a saturation in performance with longer viewing durations, suggesting a bounded accumulation process (Kiani et al., 2008; Kiani and Shadlen, 2009).

On half of all trials, electrical μ S (5–10 μ A) was applied to area MT or MST during the presentation of random dot motion. Importantly, the presence or absence of μ S did not alter the designation of correct and incorrect trials. Thus, if microstimulation of rightward preferring neurons were to cause the monkey to answer “right” on a trial in which leftward motion was shown, this would be regarded as an error, hence unrewarded.

Does Microstimulation Affect the Degree of Confidence?

Since microstimulation induces an artificial pattern of activity in the brain, we wondered whether monkeys would simply opt out of the direction decision on μ S trials when given this opportunity. The answer is resoundingly negative. As shown in [Figure 2A](#), monkeys varied their propensity to choose T_s from session to session, but such variation was highly correlated on μ S and no- μ S trials (Pearson’s $r = 0.88$, $p < 10^{-20}$). Averaged within individual experiments, monkeys did not opt out more frequently on μ S trials; indeed, the trend favors a small decrease in T_s choices (see [Supplemental Experimental Procedures](#) available online). The key point is that μ S did not cause indiscriminate uncertainty about perceptual judgments, or we would have observed the opposite trend (more T_s choices on μ S trials).

Although the average frequency of T_s choices was similar on μ S and no- μ S trials, microstimulation nevertheless exerted a substantial effect on confidence judgments, corresponding to a shift of the bell-shaped function along the motion axis ([Figure 2B](#), top). Notice that for both μ S and no- μ S trials, the monkey accepted the sure bet most often for the stimulus conditions that led to the most equivocal choice proportions (i.e., 0.5 preferred-direction choices; [Figure 2B](#), bottom). Across sessions, the shift of the sure-bet function was highly correlated with the shift of the choice function (Pearson’s $r = 0.87$, $p < 10^{-18}$; [Figure S1A](#)). This close association, despite the wide range of magnitudes of both effects, is consistent with the idea that a common neural signal underlies choice and confidence.

Another way to frame this result is to consider each motion direction separately. For motion in the preferred direction (positive coherence), monkeys chose the sure bet less often when μ S was present (two-proportion z test, $p < 10^{-18}$), suggesting that μ S increased confidence by reinforcing the evidence from the visual stimulus. In contrast, for motion in the null direction (negative coherence) monkeys chose the sure bet more often when μ S was present ($p < 10^{-9}$), suggesting that μ S decreased confidence by contradicting the evidence for null-direction motion. The end result is a leftward shift of the curve, as if μ S had injected a signal largely equivalent to a change in motion coherence. Lastly, [Figure 2B](#) (top) clarifies the subtle decrease in the number of T_s choices accompanying μ S (noted above), which is most evident at the peaks of the sure-bet functions. We will explain this apparent increase in confidence using the model described below.

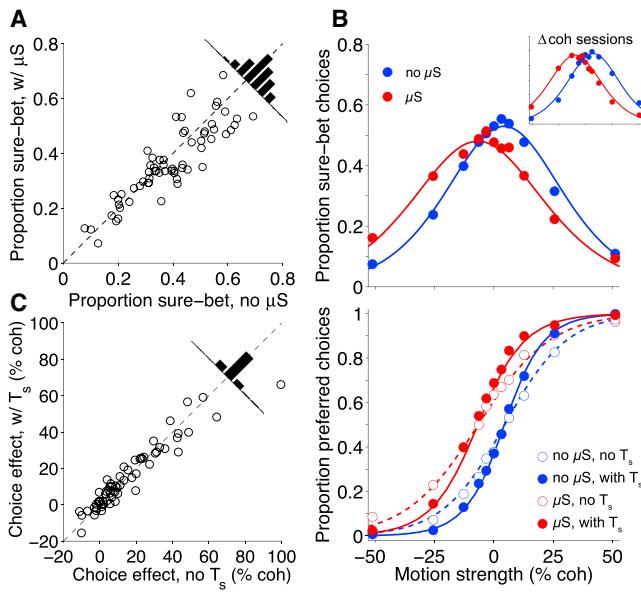


Figure 2. Effects of μ S on PDW and Perceptual Decisions

(A) The proportion of trials in which the monkey opted out of the direction task and chose the sure bet, comparing μ S and no- μ S trials ($n = 63$ sites). (B) Top: proportion of sure-bet (T_s) choices as a function of motion strength (percent coherence; positive = preferred direction of neurons at the stimulation site) for all sessions in both monkeys ($n = 53,134$ trials). Red and blue data points indicate μ S and no- μ S trials, respectively, combining across all viewing durations. Error bars (SE) are smaller than the data points. Top, inset: proportion T_s choices in separate control sessions, for trials with (red) and without (blue) an offset added to the motion coherence in lieu of μ S ("Δcoh," see text and Figure S3). Bottom: proportion of preferred-direction choices as a function of motion strength, plotted separately for the four conditions of the 2×2 design: μ S present (red) or absent (blue), and T_s offered but waived (solid curves and filled symbols) or T_s not offered (dashed curves and open symbols). In both panels, smooth curves represent fits to the bounded evidence-accumulation model (see text), with the exception of the red solid and dashed curves in the bottom panel. These are the predicted μ S choice functions based on a fit to the remaining observations. (C) Comparison of the effect of μ S on choices (represented as an equivalent change in motion strength) on trials with and without the T_s option.

Microstimulation combined with PDW yields four choice functions: $\pm\mu$ S when the sure bet was not offered and $\pm\mu$ S when the monkey could have opted out but instead chose the preferred or null direction of the stimulated neurons. These four conditions are represented by the four curves in the bottom of Figure 2B. There are three salient observations. First, μ S shifted the choice function by the same amount whether or not T_s was available, and this similarity was apparent across the 63 sites (Figure 2C; paired t test, $p = 0.56$). If on some trials μ S had affected confidence in a manner unlike a change in visual evidence, its effect on choices might have been different depending on whether the monkey had the chance to opt out of such trials. The results did not support this possibility. Second, μ S reduced monkeys' sensitivity to motion, consistent with previous studies (Ditterich et al., 2003; Salzman et al., 1992). This is only apparent as a subtle attenuation in the slope of the red curves compared with their blue (no- μ S) counterparts, but the effect is reliable ($17\% \pm 2\%$ change, $p < 10^{-16}$, logistic regression; Equation 2).

It suggests that μ S occasionally weakened the directional signal and/or added a small amount of noise to the decision process. As shown below, an increase in noise can also explain the small decrease in the maximum rate of T_s choices on μ S trials. Third and most importantly, μ S did not abolish the improved sensitivity to motion on trials when T_s was available but waived, as indicated by the steeper slope of the solid compared to the dashed curves (Figure 2B, bottom; logistic regression, $p < 10^{-24}$ for both μ S and no- μ S conditions tested separately). Recall that this improvement is a sign that the monkey evaluated the reliability of the evidence and communicated its direction choice when the reliability seemed high (Kiani and Shadlen, 2009). Its presence on μ S trials implies that such evaluation of evidence is not compromised by artificially manipulating the sensory representation. It also means that μ S did not simply compel the monkey to choose the preferred direction with some probability, irrespective of the state of the perceptual evidence. Rather, it exerted its effects by changing the available evidence for the decision.

Controls: High-Current Stimulation and a Visual Perturbation

The data presented thus far suggest that μ S does not reduce the monkey's overall degree of certainty but instead resembles a change in visual motion. However, it is possible that the absence of an increase in the overall frequency of T_s choices on μ S trials was due to the monkey's inability or unwillingness to choose T_s beyond some rate throughout the experiment. One way to test this possibility is to apply a μ S condition that impairs discrimination performance (e.g., by deliberately weakening the differential directional signals underlying choice). This kind of impairment can be achieved simply by increasing the current amplitude, thus activating indiscriminately a larger population of neurons with a broad distribution of preferred directions (Murasugi et al., 1993). Thus, at eight sites, after completing a block of trials with standard low-amplitude pulses (7.5 μ A), we began a second block with 75 μ A pulses while keeping all other parameters identical. High-current μ S reduced the monkey's sensitivity to motion (logistic regression, $p < 10^{-19}$; Figure 3B) and also led to a greater proportion of T_s choices (no- μ S = 0.45 ± 0.01 ; μ S = 0.49 ± 0.02 ; $p < 0.05$; Figure 3A). The latter effect can be described as primarily a widening, rather than a shift, of the sure-bet function, driven by a pronounced increase in T_s choices for the highest motion strengths. The result implies that PDW does not lack the power to expose a decrease in confidence, and it reinforces the notion that decision accuracy and confidence are linked. Indeed this link was also present in the main experiments: across sessions, flatter choice functions were associated with wider sure-bet functions (Spearman's rank correlation, $\rho = -0.55$, $p < 10^{-5}$), and the modest changes in these two metrics caused by low-current stimulation were correlated ($\rho = -0.37$, $p < 0.004$; see Supplemental Experimental Procedures).

In a second control experiment, instead of stimulating the brain electrically on half the trials, we manipulated the visual stimulus in a manner that mimics the hypothesized effect of μ S. We reasoned that if the brain interprets μ S like a change in motion strength, we should approximate the effects of μ S on choice and PDW by simply adding an offset to the motion

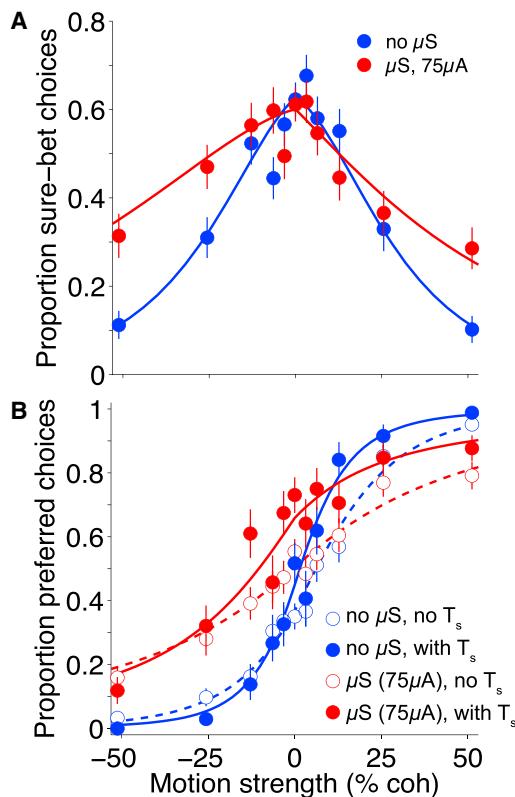


Figure 3. Stimulation with High Current Disrupts Both Accuracy and Confidence

(A and B) Combined data from eight experiments using 75 μ A stimulation ($n = 4,483$ trials). Same conventions as Figure 2B. Smooth curves are best fits of the extended model described in the [Supplemental Experimental Procedures](#) (see also Figure S2).

coherence, termed Δ coh, in one arbitrary “preferred” direction (Salzman et al., 1992). We performed 42 such experiments, using a range of Δ coh values (varied across sessions) to approximate the range of μ S effects in the main experiment. Like μ S, Δ coh trials were rewarded based on the direction of motion that would have occurred in the absence of a coherence offset. As expected, the Δ coh manipulation shifted the pattern of direction and T_s choices by an amount similar to the magnitude of added coherence. The results were largely comparable to μ S sessions, including similar shifts of the T_s curve and the choice function (Figures S1B and S3B), similar effects on direction choices with and without the T_s offer (Figure S3C; $p = 0.26$, paired t test), and improved sensitivity on T_s -waived trials (Figure S3B, bottom; logistic regression, $p < 10^{-17}$). A notable difference from μ S is the lack of an effect on the maximum rate of T_s choices (compare Figure 2B [top] versus inset; see below and [Supplemental Experimental Procedures](#)).

A Common Mechanism for the Effect of Microstimulation on Choice and Confidence

In the absence of μ S, both direction choices and PDW are explained by the accumulation of noisy evidence bearing on the

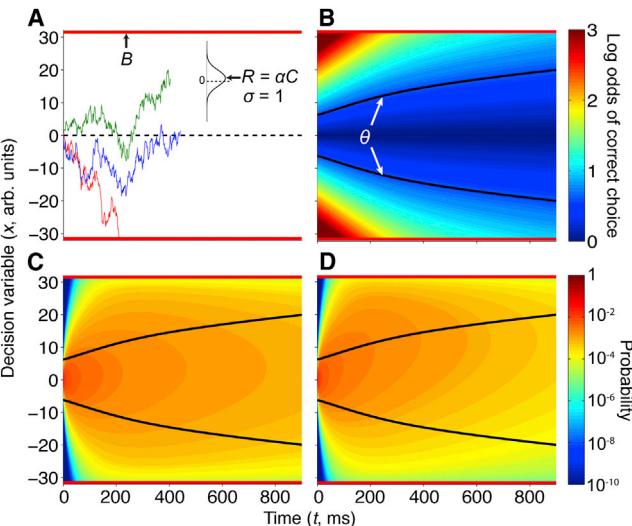


Figure 4. Bounded Evidence Accumulation Model Explains PDW and Effects of μ S

(A) Colored traces represent the accumulation of noisy motion evidence (i.e., from MT/MST) on three individual (simulated) trials. Evidence is drawn from a Gaussian distribution with mean (R) proportional to motion strength and SD (σ) equal to 1. Decision formation terminates when the stimulus is turned off (green and blue trials) or when the accumulated evidence (the decision variable, x) reaches a bound at $\pm B$ (red trial). (B) The model prescribes a sure-bet choice when the logarithm of the odds of being correct is below a fixed threshold, θ , indicated by the black contours which divide the x , t plane into low- and high-confidence regions. For example, the blue trial in (A), but not the green trial, would have ended in a sure-bet choice. (C and D) The probability density of the decision variable across time is shown for a particular motion coherence (3.2%, i.e., weak preferred-direction motion), either without (C) or with (D) μ S. Bias in this example was set to zero for simplicity. Microstimulation shifts the density upward, thereby decreasing the probability of a sure bet and increasing the probability of a preferred choice (see Figure S4). A key assumption of the model is that the brain applies the same mapping between accumulated evidence and the expected log odds of being correct (i.e., confidence), and the same criteria for opting out (black contours in B–D), irrespective of the presence of μ S.

direction of motion (Kiani and Shadlen, 2009) (Figure 4A). The model assumes that (1) a direction choice is based on the sign of the accumulated evidence and (2) a sure-bet choice supersedes a direction choice if the odds that the direction choice will be correct are less than a criterion, θ (Figure 4B). The latter requires the brain to have implicit knowledge of the association between the accumulated evidence, termed a decision variable (DV), and the likelihood that a decision based on this evidence will be correct. We fit this model to the monkey’s direction and T_s choices on no- μ S trials (Figure 2B, top and bottom, blue symbols) and then incorporated the effect of μ S as a perturbation of the evidence, equivalent to a change in motion coherence and/or a change in sensory noise (see [Supplemental Experimental Procedures](#)). The parameters implementing μ S were fit using only the T_s choices on μ S trials (Figure 2B, top, red symbols). We then used the fitted parameters to predict the pattern of direction choices on μ S trials, both with and without the T_s option available (Figure 2B, bottom, red filled and open symbols). Importantly, we did not allow μ S to alter the association between DV and

Table 1. Maximum-Likelihood Estimates of Model Parameters, \pm SE

	μ S Data Set (n = 53,134 Trials)	Δ coh Data Set (n = 43,054 Trials)
α	0.294 ± 0.001	0.291 ± 0.002
B	31.6 ± 1.30	31.2 ± 1.00
γ	-0.0134 ± 0.0002	-0.0224 ± 0.0002
θ	0.609 ± 0.004	0.507 ± 0.004
δ_C	0.112 ± 0.001	0.173 ± 0.001
δ_{σ^2}	0.237 ± 0.015	0.039 ± 0.006

confidence or the criterion for selecting T_s (i.e., θ). This strategy formalizes the qualitative assertion that μ S-induced changes in neural activity are processed like vision-induced changes in neural activity. In other words, the brain does not know that it is being stimulated.

The model fits and predictions are shown by the smooth curves in Figure 2B (and dashed curves in Figures 1B–1D). They capture several key features of the data: (1) the relationship between T_s choices and trial difficulty (i.e., motion strength and viewing duration; Figures 1B and 2B, top, blue curve), (2) the improvement in sensitivity when T_s was offered but waived (Figure 2B, bottom: solid versus dashed curves; Figures 1C and 1D), and (3) the main effects of μ S on choice and PDW (Figure 2B, horizontal shift of red versus blue curves).

Not surprisingly, the model explains the results from the Δ coh experiment as well (Figure S3B), but it also helps explain the key discrepancy between the effects of Δ coh and microstimulation. For the μ S experiments, an adjustment to the variance of the DV (δ_{σ^2} ; Equation S8; *Supplemental Experimental Procedures*) largely accounted for the small decrease in sensitivity to motion on μ S trials (Table 1), whereas this adjustment was negligible for the Δ coh manipulation, which did not affect sensitivity to motion ($p = 0.68$, logistic regression). Interestingly, the added variance also explains the apparent increase in confidence (decrease in T_s choices) associated with stimuli near the point of maximum ambiguity (Figure 2B, top, red curve). It may seem counterintuitive that an increase in variance (i.e., sensory noise) would predict an increase in confidence, but it is readily explained in our framework because dispersion of the DV away from the starting level causes more of its density to lie outside the region for opting out of the direction decision (see Rahnev et al., 2012 for a similar explanation).

We considered and rejected several alternative models, the most important of which allow for the possibility that μ S induces a change in either the mapping of the DV to confidence or the criterion—applied to this mapping—for opting out. Specifically, we relaxed the assumption that the criterion (θ) was unaffected by μ S. This is important because allowing μ S to affect θ is tantamount to accepting that μ S induces a change in neural activity that is processed qualitatively differently than activity caused by visual motion. We found that this extension was not justified for the main data set (see *Supplemental Experimental Procedures*). Note that the effects of μ S on choice (Figure 2B, bottom, solid and dashed red curves) were predictions of the model, based on the fit to the rest of the

data. The impressive agreement to data leads us to conclude that microstimulation did not alter the quantitative relationship between the neural representation of evidence and its mapping to a degree of certainty. In this way, the modeling exercise supports a unified theory of choice, confidence, and—by extension to previous work (Ditterich et al., 2003; Hanks et al., 2006)—reaction time.

DISCUSSION

Cortical microstimulation in behaving monkeys has long been a fruitful approach for exposing causal relationships between neural activity and perception (Bartlett and Doty, 1980; Doty, 1965). The power of the technique lies in its ability to link the functional properties of sensory neurons (e.g., direction selectivity in MT/MST) with psychophysical performance, as shown previously in several brain areas and tasks (Afraz et al., 2006; DeAngelis et al., 1998; Gu et al., 2012; Romo et al., 1998; Salzman et al., 1990). However, subjects in these studies typically report only a primary decision about the sensory stimulus. Here we have stimulated direction-selective neurons while allowing monkeys to report something additional about the decision process: their confidence, or lack thereof, in the choice. We found that μ S affected confidence as if there were an offset to the visual evidence supporting the choice. The results support a quantitative framework in which confidence emerges from the same basic mechanism—bounded evidence accumulation—that successfully accounts for choice and reaction time. Thus, combined with previous studies (Ditterich et al., 2003; Salzman et al., 1990), there is now experimental evidence that links the activity of neurons in extrastriate visual cortex in a causal fashion to all three pillars of choice behavior (Vickers, 1979).

Some might wonder whether this is in any way surprising, given what we know from previous work (Bisley et al., 2001; Ditterich et al., 2003; Romo et al., 1998; Salzman et al., 1992). In fact, our study could have turned out differently because microstimulation induces a pattern of activity that is quite different from that caused by visual stimulation. This pattern could have failed to engage the same networks that normally read out sensory information for the purpose of computing confidence (Bach and Dolan, 2012; Bartfeld et al., 2013). Had μ S induced incongruous changes in choice and confidence—or interfered with the improvement in sensitivity achieved by opting out of select trials—it would not have called into question previous findings of the effects of microstimulation on choice and reaction time. In short, the linking hypothesis tested here was by no means a foregone conclusion. Indeed, one might expect confidence to rest heavily on factors (e.g., metacognitive or affective) beyond operations on evidence and its conversion to a decision, especially considering the proposed role of higher-order structures (Kepcs et al., 2008; Komura et al., 2013; Rounis et al., 2010). Our findings do not directly conflict with these previous studies, but they do support a relatively straightforward mechanism for computing confidence in a perceptual decision (Kiani and Shadlen, 2009)—one that is tightly linked to the decision process itself.

Monkeys can be trained to detect microstimulation in a number of brain areas (Histed et al., 2013; Murphrey and Maunsell, 2007), and we cannot rule out the possibility that they could detect its presence in the current study. What we do know is that they are unable or unwilling to counteract the effects of μ S on choice (and confidence), even though doing so would increase reward rate. The presence of a compensatory choice bias against the stimulated direction (Figure 2B, bottom, rightward shift of blue curves; Salzman et al., 1992) further argues that monkeys do not differentiate between μ S and non- μ S trials. They do not adjust their strategy solely on μ S trials but instead adjust their bias on all trials, reducing errors caused by μ S at the cost of more errors when μ S was absent. Even if μ S were detectable, our results suggest that such detection did not disrupt the critical aspect of the decision process that establishes a level of confidence. From the perspective of downstream brain areas, the additional perturbation caused by μ S of MT/MST is largely equivalent to a change in the neural activity produced by a visual stimulus.

We were able to explain the monkey's PDW behavior using the same bounded drift-diffusion model used to explain direction choices and RT in previous studies (Gold and Shadlen, 2007; Kiani and Shadlen, 2009; Link, 1992; Palmer et al., 2005; Smith and Vickers, 1988). The model exploits the association between the DV and the probability that a choice based on that DV will be correct, predicting a sure-bet choice when this probability is below a fixed threshold. This model can explain the principal effects of μ S by treating it as an offset to the motion strength (Figures 2B and S4). Importantly, the model explains the assignment of confidence in a single decision based on an evolving DV. An alternative is that the monkey identifies the motion coherence and opts out with some frequency based on a learned association between coherence and the probability of being correct. However, this interpretation is contradicted by the improvement in performance—for all motion strengths and durations—on trials where the sure bet was offered and waived. The improvement implies that the brain is opting out selectively, based on a prediction that the decision reached during motion viewing is likely to be correct. The observation is also incompatible with other alternatives, such as selecting T_s following lapses of attention or evading the motion decision entirely on some fraction of trials (i.e., wishing for T_s and simply guessing if it does not become available). Importantly, the model explains the degree of improvement with impressive fidelity (Figure 2B, bottom, blue curves) and is able to predict the similar pattern on μ S trials (Figure 2B, bottom, red curves) based on a fit to the other features of the data (see Supplemental Experimental Procedures). The fact that this improvement is preserved on μ S trials is notable and could not have been predicted from previous work. It suggests that a rather sophisticated capacity to assess the reliability of sensory evidence is maintained despite the unnatural pattern of neural activity induced by μ S.

In addition to shifting the sure-bet curve, μ S also slightly reduced the peak rate of T_s choices (Figure 2B, top). The lack of such an effect in the Δ coh control experiment (Figure 2B, top, inset) suggests that this is a consequence of μ S itself rather than any analysis method or incidental feature of the task, such

as reward contingencies or the compensatory bias. The change in peak T_s frequency can be explained if we assume that μ S affects both the signal and the noise of the sensory representation. In the context of bounded evidence accumulation, adding noise effectively increases the likelihood that the DV will diffuse away from zero (i.e., neutrality) and beyond the threshold for waiving T_s . This explanation is also consistent with the small decrease in sensitivity associated with μ S. Note that an effect on noise is distinct from the proposed mechanism by which high-current stimulation reduces perceptual sensitivity (Figures 3 and S2; see Supplemental Experimental Procedures). The latter is believed to result from the spread of current to multiple columns with different preferred directions (Murasugi et al., 1993)—a dilution of signal rather than an increase in noise. Even low-current stimulation may spread across columns in some cases, but the changes in confidence that we observed suggest an effect on noise per se, the mechanism of which remains unknown. This subtle effect notwithstanding, a key conclusion from the model is that μ S does not influence higher-level aspects of decision strategy, such as the internal mapping between the state of the accumulated evidence and the likelihood of making a correct choice.

It seems uncontroversial that signals in the visual cortex would affect both choice and certainty, but it is remarkable that the coupling should be so well explained by a common mechanism. After all, certainty and confidence invite consideration of temperament, mood, and subjective experience about the decision process itself (e.g., metacognition). Thus, it is noteworthy that the monkeys did not exercise the option to indicate that something was peculiar about the decision process on trials accompanied by μ S. In effect, the monkeys have communicated just the opposite: they "wager" as if they experienced a change in the visual stimulus. Moreover, the high-current μ S experiments (Figure 3A) reassure us that the monkey is in fact able to use PDW to report decreased confidence when it occurs. Of course, we do not know what monkeys experience subjectively when we stimulate the brain, nor can we interrogate the subjective feeling of certainty itself. That said, any neuroscientific investigation is unlikely to furnish this level of explanation. What seems certain is that a quantitative reconciliation of choice, RT, and confidence will provide a basis for extending the neurobiology of decision making to more complex situations in which confidence itself plays a critical role.

EXPERIMENTAL PROCEDURES

Behavioral Task

Two adult male rhesus monkeys (*Macaca mulatta*) were trained to perform a direction discrimination task with postdecision wagering (PDW), as described previously (Kiani and Shadlen, 2009). The task (Figure 1A) was to determine the net direction of motion in a circular patch of dynamic random dots. Motion could be in one of two directions separated by 180°, and difficulty was controlled by varying both the viewing duration (truncated exponential distribution, mean = 270 ms, range = 60–880 ms) and the percentage of coherently moving dots (motion coherence: 0%, 3.2%, 6.4%, 12.8%, 25.6%, or 51.2%). After acquiring central fixation, two direction-choice targets appeared on opposite sides of the fixation point (9°–12° eccentricity), followed by the random dot motion display. After motion offset, the monkey maintained fixation through a variable delay period (range = 500–1,000 ms), during which a third target (the sure-bet target [T_s]) appeared on a random half of trials.

Importantly, the monkeys could not predict whether T_s would appear until at least 500 ms after stimulus offset, strongly encouraging them to complete a direction decision on all trials. T_s differed in color and size from the direction-choice targets and was positioned at an angle perpendicular to the motion axis at 6°–8° eccentricity.

After the delay period, the fixation point disappeared, cueing the monkey to make a saccadic eye movement to one of the targets. When given the opportunity, the monkey could choose T_s and receive a guaranteed reward (drop of water or juice) or waive T_s and make the higher-stakes direction choice. Correct direction choices yielded a larger liquid reward than T_s choices, while errors resulted in a 5–6 s timeout. The ratio of T_s reward size to direction-choice reward size (0.75–0.82 for monkey I, 0.64–0.72 for monkey D) was chosen to encourage the animals to choose T_s approximately 50% of the time at the weakest motion strengths. The ratio was not adjusted during the course of an experiment.

Surgery and Neurophysiological Methods

All procedures were in accordance with National Institutes of Health guidelines and approved by the Institutional Animal Care and Use Committees at the University of Washington and Columbia University. Animals were implanted with a head post and recording chamber using aseptic surgical methods. Electrical microstimulation and multiunit recordings were made with tungsten electrodes (Alpha Omega, impedance = 0.5–2 MΩ measured at 1 kHz). Areas MT ($n = 32$ sites) and MST ($n = 31$) were identified using structural MRI scans and standard physiological criteria, as well as histological analysis in one animal. Stimulation sites were chosen based on strong direction selectivity and consistent tuning (across ~200 μm of cortex) for the direction, speed, and size of the motion stimulus (see *Supplemental Experimental Procedures* for details).

Once we encountered an acceptable site, we positioned the electrode tip near its center and began the discrimination task. On a random half of trials, including both T_s -present and T_s -absent trials, microstimulation was delivered through the recording electrode using a Grass S88 stimulator with two PSIU6 optical isolation units (Grass Technologies). Stimulation trains consisted of square-wave, biphasic pulses with the following parameters: 5, 7.5, or 10 μA; 200, 250, or 333 Hz, 0.4 ms negative and 0.4 ms positive phase (negative phase leading). Within these ranges, no systematic effects of pulse amplitude or frequency were detected. The pulses began 40 ms after motion onset and stopped 40 ms after motion offset to account for visual response latency. For eight sites in one monkey, following the standard block of trials, an additional block was collected in which the amplitude of pulses was increased to 75 μA (“high-current”) while all other parameters remained the same. Note that our pulse duration (0.4 ms) was longer than in previous studies by a factor of 1.33 (Ditterich et al., 2003) or 2 (Murasugi et al., 1993; Salzman et al., 1992).

We performed a set of control experiments with the same task design and stimuli, except that electrical microstimulation was replaced with an offset to the motion coherence assigned by the computer on a given trial (dubbed “added-signal” trials in Salzman et al., 1992). The coherence offset (Δcoh) was fixed for a given session and varied from 5%–40% coh across sessions (see Figures S1B and S1C).

Behavioral Data Analysis

We fit the direction choices to the logistic regression model given by:

$$P_{\text{pref}} = \left\{ 1 + e^{-Q} \right\}^{-1}, \quad Q = \beta_0 + \beta_1 I_E + \beta_2 C \quad (\text{Equation 1})$$

where P_{pref} is the probability of a preferred-direction choice, C is signed motion coherence, I_E is an indicator variable for μS (1 or 0 for trials with/without μS), β_0 is the overall bias, β_1 estimates the effect of μS on the direction choice, and β_1/β_2 expresses this in units of motion coherence. Fitting was performed by the method of maximum likelihood (binomial error), with SEs of the parameters obtained from the inverted Hessian matrix. SEs were used to compute t statistics and thereby evaluate the null hypothesis (e.g., $\beta_1 = 0$). Effects of μS on choice were similar between MT and MST (two-sample K-S test, $p = 0.39$); thus, we pooled the data from the two areas for all analyses.

To quantify the change in sensitivity associated with μS , we fit the logistic model given by:

$$P_{\text{pref}} = \left\{ 1 + e^{-Q} \right\}^{-1}, \quad Q = \beta_0 + \beta_1 I_E + \beta_2 C + \beta_3 I_E C \quad (\text{Equation 2})$$

where β_3 captures the effect on sensitivity. Similarly, the difference in sensitivity with and without T_s present (Figure 2B, bottom) was examined by replacing I_E in *Equation 2* with an indicator term for T_s .

For some analyses, we fit the probability of sure-bet choices as function of signed coherence with a Gaussian function (*Supplemental Experimental Procedures*, Equations S1 and S2). Note that the smooth curves in Figures 2B, 3, S2A, and S3 were generated from the bounded accumulation models (see below), not logistic regression or Gaussian fitting. For additional methods and results related to the Gaussian fits, see *Supplemental Experimental Procedures*.

Model Fits and Predictions

Here we provide an intuitive overview of the model (Figure 4) and fits displayed in Figures 1B–1D, 2B, and S3B. Variables constituting degrees of freedom are identified by bold font and listed in *Table 1*. For mathematical details, see *Supplemental Experimental Procedures*.

We explain choice and PDW using a simplified one-dimensional diffusion process (Gold and Shadlen, 2007; Kiani and Shadlen, 2009; Link, 1992; Ratcliff and Rouder, 1998; Smith and Vickers, 1988) in which noisy evidence favoring either direction (and against the other) is accumulated for its display duration or until the process reaches an upper or lower bound, $\pm B$. The bounds would explain reaction time in other contexts, whereas here they render decision times shorter than the display duration on some trials, and they affect the predicted accuracy on these trials (Kiani et al., 2008). The accumulation has a drift and a diffusion component. The latter is the accumulation of independent random numbers at each time step. The drift is a line with slope (drift rate) proportional to the motion coherence (αC), where the sign of C indicates direction. The sign of the accumulated evidence, termed the decision variable (x), determines the choice. Confidence, in turn, is the log odds that such a choice would be correct. It is a function of both x and time (i.e., the stimulus duration or the time that the accumulation reached a bound). The time dependence arises because the reliability of the evidence (motion strength) varies unpredictably across trials and is not explicitly known by the observer (Drugowitsch et al., 2014; Kiani and Shadlen, 2009). When T_s is offered, we assert that the monkey exercises or waives this option based on a criterion, θ , applied to the log odds of being correct (Figures 4B–4D). The model generates the expected probability of each option by propagating and integrating the probability density of the decision variable within different regions of this space, as partitioned by θ and the bounds (Figures 4C and 4D).

On trials with μS , we assume that the drift rate is offset by δ_C , equivalent to a change in the motion coherence, and allow for the possibility that the diffusion noise is also affected (offset by δ_{σ^2}). On all trials, the drift rate includes an additional offset term γ to account for the compensatory bias that arises in microstimulation experiments (Salzman et al., 1992).

We employed a simple parameterization and tiered fitting strategy designed to minimize the number of degrees of freedom of the model (see *Supplemental Experimental Procedures*). We pursued this strategy to guard against over fitting and to support intuitions about the neural mechanisms. The model furnishes the smooth curves in the analyses of the main data set (low current μS) and the Δcoh control (Figures 2B and S3B, respectively; see also Figure S4), as well as the dashed curves in Figures 1B–1D. A more elaborate model, also described in the *Supplemental Experimental Procedures*, was required to explain the effects of high current stimulation (Figures 3 and S2).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.neuron.2014.07.011>.

AUTHOR CONTRIBUTIONS

M.N.S., C.R.F., and R.K. conceived and designed the experiments and performed the analyses. C.R.F. and R.K. collected the data. All authors wrote the paper.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute, National Eye Institute grant EY11378, NEI Core Grant P30EY01730, and National Center for Research Resources grant RR00166. We thank members of the lab for discussions, G. Horwitz for histology, and A. Boulet, K. Ahl, and K. Morrisroe for technical assistance.

Accepted: July 7, 2014

Published: August 7, 2014

REFERENCES

- Afraz, S.-R., Kiani, R., and Esteky, H. (2006). Microstimulation of inferotemporal cortex influences face categorization. *Nature* **442**, 692–695.
- Bach, D.R., and Dolan, R.J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nat. Rev. Neurosci.* **13**, 572–586.
- Bartlett, J.R., and Doty, R.W. (1980). An exploration of the ability of macaques to detect microstimulation of striate cortex. *Acta Neurobiol. Exp. (Warsz.)* **40**, 713–727.
- Barttfeld, P., Wicker, B., McAleer, P., Belin, P., Cojan, Y., Graziano, M., Leiguarda, R., and Sigman, M. (2013). Distinct patterns of functional brain connectivity correlate with objective performance and subjective beliefs. *Proc. Natl. Acad. Sci. USA* **110**, 11577–11582.
- Bisley, J.W., Zaksas, D., and Pasternak, T. (2001). Microstimulation of cortical area MT affects performance on a visual working memory task. *J. Neurophysiol.* **85**, 187–196.
- DeAngelis, G.C., Cumming, B.G., and Newsome, W.T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature* **394**, 677–680.
- Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., and Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* **132**, 2531–2540.
- Ditterich, J., Mazurek, M.E., and Shadlen, M.N. (2003). Microstimulation of visual cortex affects the speed of perceptual decisions. *Nat. Neurosci.* **6**, 891–898.
- Doty, R.W. (1965). Conditioned reflexes elicited by electrical stimulation of the brain in macaques. *J. Neurophysiol.* **28**, 623–640.
- Drugowitsch, J., Moreno-Bote, R., and Pouget, A. (2014). Relation between Belief and Performance in Perceptual Decision Making. *PLoS ONE* **9**, e96511.
- Foote, A.L., and Crystal, J.D. (2007). Metacognition in the rat. *Curr. Biol.* **17**, 551–555.
- Gold, J.I., and Shadlen, M.N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574.
- Gu, Y., Deangelis, G.C., and Angelaki, D.E. (2012). Causal links between dorsal medial superior temporal area neurons and multisensory heading perception. *J. Neurosci.* **32**, 2299–2313.
- Hampton, R.R. (2001). Rhesus monkeys know when they remember. *Proc. Natl. Acad. Sci. USA* **98**, 5359–5362.
- Hanks, T.D., Ditterich, J., and Shadlen, M.N. (2006). Microstimulation of macaque area LIP affects decision-making in a motion discrimination task. *Nat. Neurosci.* **9**, 682–689.
- Histed, M.H., Bonin, V., and Reid, R.C. (2009). Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron* **63**, 508–522.
- Histed, M.H., Ni, A.M., and Maunsell, J.H.R. (2013). Insights into cortical mechanisms of behavior from microstimulation experiments. *Prog. Neurobiol.* **103**, 115–130.
- Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. (Cambridge: Cambridge University Press).
- Kepecs, A., Uchida, N., Zariwala, H.A., and Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature* **455**, 227–231.
- Kiani, R., and Shadlen, M.N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science* **324**, 759–764.
- Kiani, R., Hanks, T.D., and Shadlen, M.N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J. Neurosci.* **28**, 3017–3029.
- Komura, Y., Nikkuni, A., Hirashima, N., Uetake, T., and Miyamoto, A. (2013). Responses of pulvinar neurons reflect a subject's confidence in visual categorization. *Nat. Neurosci.* **16**, 749–755.
- Lau, H.C., and Passingham, R.E. (2006). Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc. Natl. Acad. Sci. USA* **103**, 18763–18768.
- Link, S.W. (1992). *The Wave Theory of Difference and Similarity*. (Hillsdale: Lawrence Erlbaum Associates).
- Logothetis, N.K., Augath, M., Murayama, Y., Rauch, A., Sultan, F., Goense, J., Oeltermann, A., and Merkle, H. (2010). The effects of electrical microstimulation on cortical signal propagation. *Nat. Neurosci.* **13**, 1283–1291.
- Middlebrooks, P.G., and Sommer, M.A. (2011). Metacognition in monkeys during an oculomotor task. *J. Exp. Psychol. Learn. Mem. Cogn.* **37**, 325–337.
- Murasugi, C.M., Salzman, C.D., and Newsome, W.T. (1993). Microstimulation in visual area MT: effects of varying pulse amplitude and frequency. *J. Neurosci.* **13**, 1719–1729.
- Murphy, D.K., and Maunsell, J.H.R. (2007). Behavioral detection of electrical microstimulation in different cortical visual areas. *Curr. Biol.* **17**, 862–867.
- Palmer, J., Huk, A.C., and Shadlen, M.N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* **5**, 376–404.
- Rahnev, D.A., Maniscalco, B., Luber, B., Lau, H., and Lisanby, S.H. (2012). Direct injection of noise to the visual cortex decreases accuracy but increases decision confidence. *J. Neurophysiol.* **107**, 1556–1563.
- Ratcliff, R., and Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychol. Sci.* **9**, 347–356.
- Romo, R., Hernández, A., Zainos, A., and Salinas, E. (1998). Somatosensory discrimination based on cortical microstimulation. *Nature* **392**, 387–390.
- Rounis, E., Maniscalco, B., Rothwell, J.C., Passingham, R.E., and Lau, H. (2010). Theta-burst transcranial magnetic stimulation to the prefrontal cortex impairs metacognitive visual awareness. *Cogn Neurosci* **1**, 165–175.
- Salzman, C.D., Britten, K.H., and Newsome, W.T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature* **346**, 174–177.
- Salzman, C.D., Murasugi, C.M., Britten, K.H., and Newsome, W.T. (1992). Microstimulation in visual area MT: effects on direction discrimination performance. *J. Neurosci.* **12**, 2331–2355.
- Smith, P.L., and Vickers, D. (1988). The accumulator model of 2-choice discrimination. *J. Math. Psychol.* **32**, 135–168.
- Smith, J.D., Beran, M.J., Couchman, J.J., and Coutinho, M.V.C. (2008). The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychon. Bull. Rev.* **15**, 679–691.
- Smith, J.D., Couchman, J.J., and Beran, M.J. (2012). The highs and lows of theoretical interpretation in animal-metacognition research. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 1297–1309.
- Tolman, E.C. (1948). Cognitive maps in rats and men. *Psychol. Rev.* **55**, 189–208.
- Vickers, D. (1979). *Decision Processes in Visual Perception*. (New York: Academic Press).

Neuron, Volume 83

Supplemental Information

**Effects of Cortical Microstimulation
on Confidence in a Perceptual Decision**

Christopher R. Fetsch, Rozbeh Kiani, William T. Newsome, and Michael N. Shadlen

Index

page

1 Figures S1-S4

5 Table S1

Supplemental Experimental Procedures:

6 Neurophysiological methods and selection of stimulation sites

7 Addendum to behavioral data analysis

8 Detailed description of bounded accumulation model and tiered fitting procedure

13 Extended model for fitting high-current stimulation data

16 Supplemental References

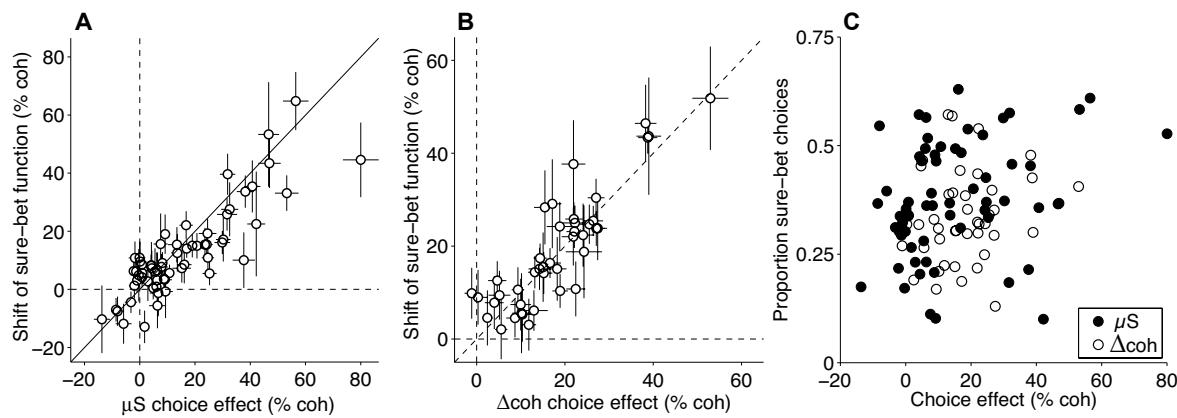


Figure S1 [related to Figure 2B]. Comparison of microstimulation effects on choice and PDW using descriptive fits (logistic regression and Gaussian, respectively; see Experimental Procedures). **(A)** For each stimulation site ($n = 63$), the effect on choice is expressed as a horizontal separation of the sigmoid functions from logistic regression (β_1/β_2 in Equation 1, Experimental Procedures). The effect on PDW is the shift of the fitted Gaussian functions (β_2 in Equation S2). Error bars are SE of the estimates. Pearson's $r > 0.8$ for each brain area and each monkey separately. **(B)** Same format as A, but for Δcoh control sessions ($n = 42$ experiments). **(C)** Variation in the proportion of sure bet choices across experiments is not explained by the size of μS and Δcoh effects on choice (Pearson $r < 0.02$, $P > 0.9$ for both). The graph also demonstrates that the levels of confidence were comparable across μS and Δcoh experiments ($P = 0.67$, 2-dimensional Kolmogorov-Smirnov test).

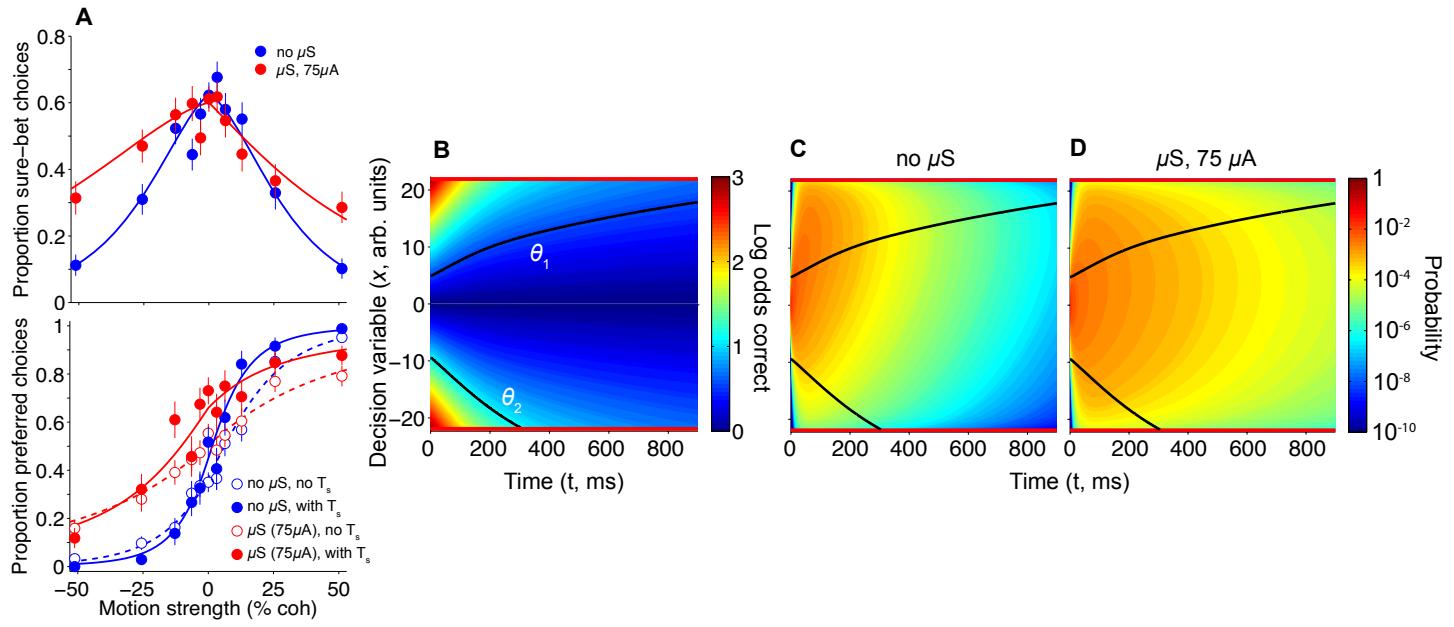


Figure S2 [related to Figure 3 and Suppl. Experimental Procedures]. Extended model used to explain high-current stimulation experiments. **(A)** Data and fits ($n = 4,483$ trials), reproduced from main text Figure 3. **(B-D)** Depiction of the extended model (see p. 13). Heat map in **B** displays the certainty that ought to accompany a decision based on the sign of the accumulated evidence x at decision time t . Units are the natural logarithm of the odds of a correct choice. The mapping is derived identically to the one in main text Figure 4B, independent of valuation, bias, or μ S. It depends only on the possible motion strengths and the parameter α . The criteria for a sure-bet choice, θ_1 and θ_2 , were allowed to differ for preferred and null choices (upper and lower halves of the map). **(C,D)** The probability density of the decision variable across time is shown for 51.2% coherence motion in the preferred-direction, separated for no- μ S **(C)** and high-current μ S **(D)** conditions. Without μ S, the density is shifted strongly upward, generating a high proportion of preferred-direction choices and relatively few sure-bet choices. The effect of high-current μ S is mainly to reduce sensitivity by activating opposing direction columns, implemented in the model by altering the relationship between coherence and drift rate (δ_α in Equation S14; see Table S1). Thus, even for the highest motion strength (51.2% coh), the model produces a decrease in accuracy and an increase in the proportion of sure-bet choices, as observed in the data (panel A, top). In addition, the asymmetric confidence region (difference between θ_1 and θ_2 , panel B) generates a peculiar bias similar to the one observed in the data: more preferred direction choices on T_s -waived relative to no- T_s trials (panel A bottom, solid curves vs. dashed counterparts).

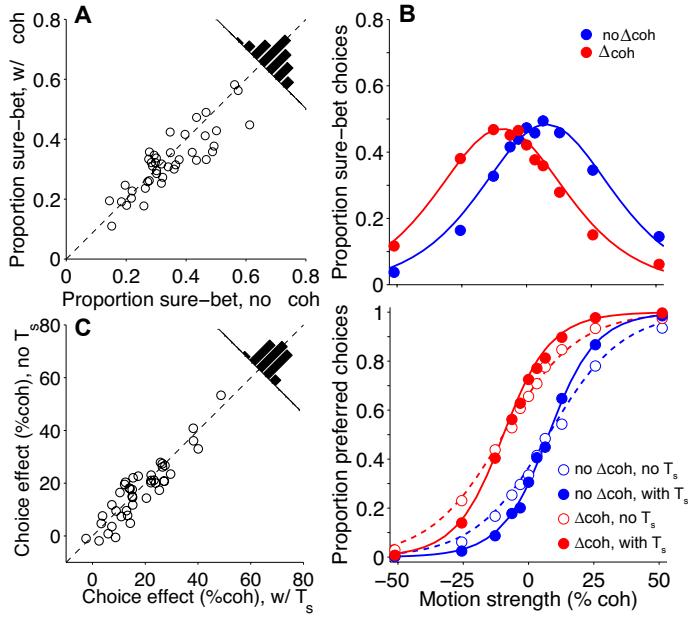


Figure S3 [related to Results: Controls and Figure 2B top, inset]. Similar pattern of PDW and choice behavior with a coherence offset added to the motion instead of μ S. (A–C) Same format as main text Figure 2, but for control ('Δcoh') sessions without electrical μ S. Top panel of B is the same as inset in Figure 2B, top. The coherence offset (fixed within sessions, varied between 5–40% across sessions) was added to a random half of trials in an arbitrary preferred direction. Trial structure and reward contingencies were identical to μ S sessions.

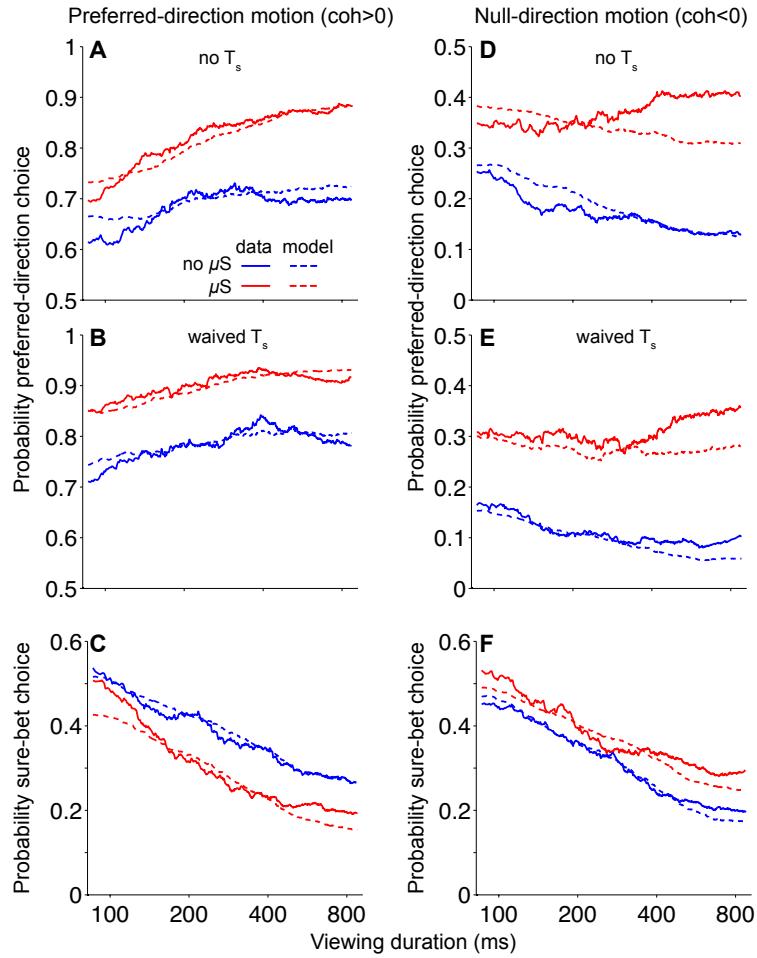


Figure S4 [related to Experimental Procedures and Figure 4]. Model fits to direction choices and sure-bet choices in the main μS experiments, separated by motion direction. (A-C) For preferred-direction motion ($coh>0$), μS increases the probability of a preferred choice (A & B, red vs. blue) and decreases the probability of a sure-bet choice (C). (D-F) μS also increases preferred-direction choices when motion is in the null direction ($coh<0$; D & E), and increases the probability of a sure-bet choice as well (F). Solid traces are running means (proportions) of the data sorted by viewing duration; dashed traces are fits to the simpler bounded accumulation model described in the main text and in detail below. Implementing μS in the model as an equivalent change in motion strength reproduces the basic pattern of effects as a function of viewing duration. One exception is that the model underestimates the probability of a preferred direction choice for $coh<0$ and long durations (red curves in D & E). It appears that longer microstimulation epochs have a greater ability to counteract opposing visual motion than predicted by a static change in motion strength.

Table S1 [related to Figure 3 and Table 1]. Maximum-likelihood estimates of extended model parameters (\pm s.e.). Note that the extended model recovers the main features of the simpler model used to fit the low current μ S experiments (repeated here from Table 1 for ease of comparison).

	High-current μ S, extended model	Low-current μ S, extended model	Low-current μ S, basic model
α	0.345 ± 0.017	0.327 ± 0.010	0.294 ± 0.001
B	21.9 ± 4.95	30.7 ± 9.10	31.6 ± 1.30
γ	-0.021 ± 0.001	-0.016 ± 0.001	-0.013 ± 0.000
θ_1	0.561 ± 0.400	0.551 ± 0.013	0.609 ± 0.004
θ_2	1.118 ± 0.075	0.623 ± 0.017	N/A
δ_c	0.142 ± 0.006	0.129 ± 0.003	0.112 ± 0.001
δ_{σ^2}	0.020 ± 0.049	0.156 ± 0.027	0.237 ± 0.015
δ_α	-0.190 ± 0.010	-0.040 ± 0.001	N/A
β	1.48 ± 0.54	0.91 ± 0.32	N/A

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Neurophysiological methods and selection of stimulation sites

Physiological criteria for identifying MT and MST included (a) audible transitions between white and gray matter, (b) vigorous responses to patches of moving dots, and (c) receptive field (RF) size vs. eccentricity. Stimulation sites classified as MT had RF diameters roughly equal to the eccentricity of the RF center and which rarely included the fovea, whereas MST sites had larger RFs which frequently included the fovea and/or part of the ipsilateral hemifield. When the angle of penetration allowed both MT and MST to be accessed in the same electrode track, we confirmed the depths of both areas as well as the white matter below the superior temporal sulcus, either before or after the behavioral session as time permitted. One monkey was euthanized and histological sections were processed for immunohistochemistry using the SMI32 antibody (neurofilament protein), which produces a characteristic staining pattern in area MT (Hof and Morrison, 1995). Examination of these sections — which contained visible guide tube tracks and an injection site placed near the microstimulation sites (for a separate study) — and comparison to brain atlases confirmed that all sites classified as MT in this animal (N=21) were indeed located in the posterior-medial portion of MT.

We isolated multiunit (MU) activity using a dual voltage-time window discriminator (BAK Electronics, Sanford, FL), with voltage and slope thresholds configured to include events slightly above the background ‘hash’ (typical pre-stimulus baseline event rate = 30-100 events/s). Once the electrode was positioned near a site with strong visual responses to flashing random-dot patches, we measured MU tuning for motion direction, speed, patch size, and position while the monkey fixated a central target. Parametric functions were fit to the MU event rates (circular Gaussian and log-normal for direction and speed tuning, respectively), and the preferred direction and speed were taken as the peaks of their respective fitted functions. Preferred size (approximate RF diameter) was taken as the patch size that elicited the strongest response, or in

the case of saturating/sigmoidal size tuning, as the smallest value in the saturated portion of the tuning function.

Tuning measurements were repeated 2-4 times at intervals of 80-120 μm along the electrode track. A site was considered acceptable for microstimulation if (a) tuning/RF parameters remained relatively stable (e.g., Δ pref. dir. $< 50^\circ$) for at least 200 μm , and (b) if direction selectivity was sufficiently strong, with at least 2 standard deviations separating preferred- and null-direction MU responses.

Addendum to behavioral data analysis

We fit the probability of sure-bet choices as function of signed coherence with a Gaussian function of the form:

$$P_{sb}(C) = Ae^{-(C-\mu)^2/(2\sigma^2)} \quad (\text{Equation S1})$$

The three free parameters (A , μ , σ) and their standard errors were estimated using the same methods as Equation 1 (main text). To evaluate the significance of changes in each parameter with μS , we fit an expanded function with three additional free parameters β_i :

$$P_{sb}(C) = (A + \beta_1 I_E) \cdot e^{-(C-\mu+\beta_2 I_E)^2/(2(\sigma+\beta_3 I_E)^2)} \quad (\text{Equation S2})$$

As in Equation 1, I_E is the μS indicator variable (0 or 1). Having found that the primary effect of μS was a change in the mean (i.e., a horizontal shift; $\beta_2 > 0$), we quantified this shift (e.g., for comparison with β_1/β_2 in Equation 1; see Figures S1A and S1B) using a simplified version of Equation S2 with β_1 and β_3 set to zero. Rank correlations reported in the main text (Results: Controls) refer to the following comparisons: σ in Equation S1 versus β_2 in Equation 1, and β_3 in Equation S2 versus β_3 in Equation 2.

To compare the frequencies of T_s choices with vs. without μS (or Δcoh), we used a paired t-test on logit-transformed proportions. The raw proportions are shown in Figures 2A and S3A, but statistical tests were performed after correcting for a potential artifact arising from (i) the nonuniform sampling of coherences (e.g., points concentrated around 0% coh) and (ii) the fact

that the shifts were not fully counteracted by the compensatory bias. For each session, we effectively slid the coherence axis so that 0% was equidistant from the means of the fitted Gaussians (Equation S1) for μ S and no- μ S (or Δ coh and no- Δ coh) conditions. The Gaussians provided interpolated proportions at the log-spaced coherences (e.g., 0%, $\pm 3.2\%$, ..., $\pm 51.2\%$) on the shifted abscissa. Averaging these proportions yielded a corrected estimate of sure-bet frequency, free from sampling confounds. Results of paired t-tests on the corrected values suggested no change in the overall sure-bet frequency for Δ coh vs. no- Δ coh ($P=0.88$; Figure S3A) and a decreasing, nonsignificant trend for μ S vs. no- μ S ($P = 0.18$). To better quantify this trend, we examined the change in the height parameter, A , of the fitted Gaussians across sessions. This comparison showed a significant decrease in the height of the sure-bet function for μ S (paired t-test, $P < 0.004$), but, as expected, no change for Δ coh ($P = 0.47$). Incidentally, the decrease in A did not entail a reduction in (corrected) sure-bet frequency because it was compensated by a small increase in width (σ). All of these subtle details are well explained by the bounded accumulation model, in particular the addition of noise on μ S trials (δ_{σ^2} ; see Table 1 and above text).

Detailed description of bounded accumulation model and tiered fitting procedure

As described briefly in Experimental Procedures, we modeled the results using a simple one-dimensional bounded-accumulation (or drift-diffusion) model. Although some task variants (e.g., reaction time studies) require more complex models, such as nonstationary bounds and races between competing processes, here we opted for the simplest model framework capable of explaining our results.

The accumulated evidence (or decision variable, x) undergoes deterministic drift plus diffusion, conforming to the discrete stochastic difference equation

$$\Delta x = R\Delta t + N\left\{0, \sqrt{\sigma^2 \Delta t}\right\}, \quad x(0) = 0 \quad (\text{Equation S3})$$

where R is the drift rate and $N\{\dots\}$ is a Normal (Gaussian) distribution with mean 0 and standard deviation $\sigma\sqrt{\Delta t}$. In the absence of bias and microstimulation, R is proportional to signed motion coherence ($R = \alpha C$; $\alpha > 0$), where positive values of C indicate motion in the preferred direction, and the variance $\sigma^2 = 1$. The process terminates when the stimulus is turned off ($t_d = \tau$) or when x reaches a bound at $\pm B$, whichever occurs first (Kiani et al., 2008).

When T_s is offered, we assert that the monkey exercises or waives this option based on the probability that a decision based on the sign of $x(t_d)$ will be correct. The model assumes implicit knowledge of the mapping between $[x(t_d), t_d]$ and this probability. First, we calculate the log likelihood ratio that any $x(t_d)$ will be achieved with motion in the preferred direction (Kiani and Shadlen, 2009):

$$\Lambda_{pref}(x(t_d), t_d) = \log \frac{\sum_i P[x(t_d)|C_i]P[C_i]}{\sum_i P[x(t_d)|-C_i]P[-C_i]} \quad (\text{Equation S4})$$

Preferred choices ($x(t_d) > 0$) are associated with positive $\Lambda_{pref}(x(t_d), t_d)$. The log likelihood ratio for the null motion direction can be easily calculated by inverting the ratio on the right hand side of Equation S4 (i.e., $\Lambda_{null}(x(t_d), t_d) = -\Lambda_{pref}(x(t_d), t_d)$), making Λ_{null} a flipped version of Λ_{pref} around $x=0$. Because the two directions of motion are equally probable, *a priori*, we can calculate the log posterior odds of a correct choice as $\Lambda_{cor} = |\Lambda_{pref}|$. Thus, Λ_{cor} is a symmetrical map around zero, and $\Lambda_{cor}(0, t_d) = 0$. Equation S4 requires estimates of the likelihoods, $P(x(t_d), t_d | C_i)$. In the present experiment, we do not measure response times, so we lack experimental access to the decision time, t_d , but we can calculate its distribution for each trial type using the model.

A central assumption of the model is that Λ_{cor} is not affected by μ S (see discussion of extended model, below). It is a mapping of certainty about evidence, independent of valuation, bias and stopping criterion. To calculate Λ_{cor} we use $R = \alpha C$ and $\sigma^2 = 1$. Therefore, Λ_{cor} depends on only one model parameter, α (Beck et al., 2008; Moreno-Bote, 2010), and the set of motion strengths in the experiment (Kiani and Shadlen, 2009).

As in previous studies, our monkeys exhibited a bias against the preferred direction of the stimulated neurons. This compensatory ‘null choice’ bias is thought to arise from the subject’s attempt to equalize the two choices in the face of an increasing tendency (caused by μ S) to choose the preferred direction (Salzman et al., 1992). Although there are several ways to implement such a bias in diffusion models (Ratcliff and McKoon, 2008; van Ravenzwaaij et al., 2012), we find that we can account for it most parsimoniously with an offset (γ) to the drift rate, which approximates a dynamic bias signal (Hanks et al., 2011). Thus for no- μ S trials,

$$R_{-\mu S} = \alpha C + \gamma \quad (\text{Equation S5})$$

$$\sigma_{-\mu S}^2 = 1 \quad (\text{Equation S6})$$

This bias is also present on μ S trials. In addition, μ S can potentially affect both the drift rate and the sensory noise:

$$R_{+\mu S} = \alpha(C + \delta_C) + \gamma \quad (\text{Equation S7})$$

$$\sigma_{+\mu S}^2 = 1 + \delta_{\sigma^2} \quad (\text{Equation S8})$$

where adjustments to constants are indicated by δ_x using the subscript to indicate the modified term. The term δ_C would cast the effect of μ S in units of motion strength.

For each trial, we obtain the probability distribution of the decision variable $P(x,t)$ by numerical solution to the Fokker-Planck equation (Chang and Cooper, 1970) using the definitions of R and σ^2 in Equations S5–S8. The solution establishes the distributions of upper

and lower bound absorption — $P(x = B, t)$ and $P(x = -B, t)$, respectively — and the distribution of x within the bounds (unabsorbed probability, $P(-B < x < B, t)$). These distributions are then used to calculate the expected probabilities of each choice under the model parameters.

On trials when T_s is not shown (T_s^-), the probability of a preferred direction choice is the probability that x has terminated in the upper bound or that it has a final positive value:

$$P_{pref}^{T_s^-}(C, \tau) = \int_0^\tau P(x = B, t_d) dt_d + \int_{x>0}^{x<B} P(x, t_d = \tau) dx + 0.5 P(x = 0, t_d = \tau) \quad (\text{Equation S9})$$

where τ is the stimulus duration. $P_{null}^{T_s^-}(C, \tau)$ can be simply calculated as $1 - P_{pref}^{T_s^-}(C, \tau)$.

On trials when T_s is shown (T_s^+), the probability that the monkey chooses this option depends on a criterion, θ , which is in units of Λ_{cor} (log odds correct). By inverting Λ_{cor} we obtain a criterion in units of the decision variable such that $x_\theta(t) = |\Lambda_{cor}^{-1}(\theta, t)|$. $x_\theta(t)$ defines the decision variable that makes $\Lambda_{cor} = \theta$ at each moment in time (black curves in Figures 4B–4D). Decision variables in the range $[-x_\theta(t), +x_\theta(t)]$ lead to sure-bet choices. Decision variables larger than $+x_\theta(t)$ lead to a preferred direction choice and those smaller than $-x_\theta(t)$ lead to a null choice. Thus, the probability of a sure-bet choice is

$$\begin{aligned} P_{sb}^{T_s^+}(C, \tau) = & \int_{-x_\theta(\tau)}^{x_\theta(\tau)} P(x, t_d = \tau) dx + \\ & \int_0^\tau P(x = B, t_d) F(\Lambda_{cor}(B, t_d)) dt_d + \\ & \int_0^\tau P(x = -B, t_d) F(\Lambda_{cor}(-B, t_d)) dt_d \end{aligned} \quad (\text{Equation S10})$$

where $F(\Lambda)$ equals 1 if $\Lambda < \theta$ and 0 otherwise. The last two integrals in Equation S10 capture the possibility that a decision that terminates at a bound would lead to a T_s choice — that is, when $\Lambda_{cor}(x = |B|, t_d) < \theta$ (e.g., for small B ; not the case in Figures 4B–4D). The probability of a

preferred direction choice is given by

$$P_{pref}^{T_s^+}(C, \tau) = \int_{x>x_\theta(\tau)}^B P(x, t_d = \tau) dx + \int_0^\tau P(x = B, t_d) (1 - F(\Lambda_{cor}(B, t_d))) dt_d \quad (\text{Equation S11})$$

and

$$P_{null}^{T_s^+}(C, \tau) = 1 - P_{pref}^{T_s^+}(C, \tau) - P_{sb}^{T_s^+}(C, \tau). \quad (\text{Equation S12})$$

We used a tiered maximum likelihood fitting procedure to obtain the model parameters. First, we fit the three basic model parameters (α, B , and θ) (Kiani and Shadlen, 2009) and the compensatory bias (γ) by maximizing the joint probability of the observed outcomes (i.e., preferred choice, null choice, or T_s choice when available) for the no- μ S trials only (blue filled and open data points in Figure 2B). With these four parameters held fixed, we then estimated δ_C and δ_{σ^2} using only the probability of T_s choices on μ S trials (Figure 2B, top, red data points). Finally, having estimated all the model parameters, we generated *predictions* for the stimulation effect on direction choices (Figure 2B, bottom, red curves) and the improvement in sensitivity on μ S trials when T_s was waived (Figure 2B, bottom, solid vs. dashed red curve).

In total, the model has six free parameters, but note that only two of them (δ_C and δ_{σ^2}) are used to explain the effects of μ S on T_s choices and no additional parameters are used for the effects of μ S on the choice functions. Moreover, our tiered fitting procedure prevents fine tuning of the basic model parameters by the stimulated trials which might have artificially boosted the explanatory power of the model. We repeated the fitting procedure multiple times from different starting points to ensure our results were not driven by local minima. Standard errors (s.e.) for the parameters were calculated using a bootstrap procedure (see Table 1).

In addition to the ‘extended’ model described below, we considered two alternative nested models to test whether μ S affects the sure-bet criterion (θ). In the first, the parameter δ_{σ^2} was set to zero and a new free parameter was added to implement an offset to θ on μ S trials (6

free parameters); that is, $\theta_{-\mu S} = \theta$ and $\theta_{+\mu S} = \theta + \delta_\theta$. In the second nested model, δ_{σ^2} was allowed to vary in addition to δ_θ (7 free parameters). These two alternative models were compared with the basic model described above using the Bayesian Information Criterion (BIC), where $\Delta\text{BIC} = \text{BIC}_{\text{alt}} - \text{BIC}_{\text{basic}}$. The value of ΔBIC was positive for both comparisons (basic vs. 6-parameter alternative, $\Delta\text{BIC} = 130.5$; basic vs. 7-parameter alternative, $\Delta\text{BIC} = 10.9$), indicating a better fit of the basic model while taking into account the number of free parameters.

Extended model for fitting high-current stimulation data

In the main text, we reported that stimulation of MT/MST with 75 μA current increased the proportion of T_s choices, consistent with lower confidence on these trials. Here we describe a model that explains this and other features of the high-current data, and we compare this model to the simpler one used for the main experiments (see above and Experimental Procedures). This modeling exercise is instructive, and while the key findings are statistically reliable, they should be regarded as tentative because they are based on only eight stimulation sites and relatively few trials (4,483; compare to 53,134 for low-current μS). To obtain a larger data set on this control would have necessitated many fewer trials with low current.

The extended model contains 9 degrees of freedom. To avoid local minima, we fit the model from several random starting points. We adopted the best of these to estimate the parameters and repeated this process in a bootstrap analysis to estimate standard errors for the parameters. That is, the s.e. are the standard deviations of the best fit to a resampled (with replacement) data set, applying the same procedure of multiple random starting points for each iteration. The parameter values and their s.e. are displayed in Table S1.

A summary of the model is as follows. The deterministic drift without μS is identical to the simpler model:

$$R_{-\mu S} = \alpha C + \gamma \quad (\text{Equation S13})$$

where α converts motion coherence to the signal component of signal-to-noise and γ is the bias term. As in the simpler model, microstimulation can affect the drift rate as an offset to the motion strength, replacing αC with $\alpha(C + \delta_C)$. However, we also consider the possibility that stimulation spreads to direction columns that supply negative evidence, thereby changing the relationship between C and drift rate. Thus we allow for a change to α :

$$R_{+\mu S} = (\alpha + \delta_\alpha)(C + \delta_C) + \gamma \quad (\text{Equation S14})$$

In other words, the model attributes the μ S-induced change in sensitivity to a reduction in signal-to-noise as if the population response were less direction selective. Thus, the slope of the choice function is flatter and T_s choice frequency does not fully attenuate with high motion strength (Figures 3, S2A). This is explained by a degradation of signal (δ_α) rather than an increase in noise (δ_{σ^2}), as if high current μ S reduces the difference between firing rates of neurons with different direction preferences, as previously suggested (Murasugi et al., 1993). This effect on signal could be achieved by changes to the visual stimulus, albeit using a different type of random dot display than the one we used. Importantly, the offset to α affects only the distribution of the decision variable during μ S trials (Equations S9–S11), not the mapping of DV to confidence (Equation S4). The value of δ_α was small but still significantly different from zero for the low-current dataset, suggesting that the decrease in sensitivity (Figure 2B, bottom) was not fully explained by increased noise (δ_{σ^2}). This is expected, however, since even low-current stimulation would occasionally recruit a broader distribution of direction signals than intended.

The extended model also includes a more elaborate description of the noise itself, irrespective of microstimulation. For all trials, we relaxed the constant variance assumption and included a coherence-dependent term:

$$\sigma_{-\mu S}^2 = 1 + \beta |C|$$

$$\sigma_{+\mu S}^2 = 1 + \delta_{\sigma^2} + \beta |C| \quad (\text{Equation S15})$$

The idea that noise scales with motion strength is consistent with the observation that the MT response to preferred and null direction motion is asymmetric. Specifically, an increase in rightward motion coherence causes an increase in the firing rate of cells that prefer rightward which is about 3-fold larger than the decrease in firing rate of cells that prefer leftward (Britten et al., 1993). This implies that the difference in firing rates of pools of MT neurons with opposing direction preferences should have a variance that is proportional to $|C|$ via multiplication with β . We typically ignore this signal-dependent noise for the sake of simplicity and because the impact happens to be less pronounced for the low-current dataset (Table S1), but including it for the smaller high-current dataset qualitatively improved the fit.

Lastly, we found that the high-current fits could be further improved by defining the boundaries of the choose/waive T_s option with greater flexibility. Recall that this is captured by a single scalar value, θ , in the simpler model used in the main text. Here we allow an asymmetry in this value for preferred and null direction choices (Figure S2B) — i.e., θ_1 and θ_2 — allowing the model to capture a tendency to opt out more often for preferred or null direction decisions. The resulting asymmetry decouples the choice functions on T_s -waived vs. T_s -absent trials (Figure S2A, bottom, solid vs. dashed curves), suggesting an alteration of the mapping between DV and certainty in those sessions in which high current stimulation was present. Note that the values of θ_1 and θ_2 are not nearly as discrepant for low current μS (Table S1), which maintains the approximate symmetry of the mapping as assumed in the simple model in the main text. We also tested a more elaborate model which allowed an offset to θ_1 and θ_2 on μS trials only (not shown), but these offset parameters were not significantly different from zero (high-current data, $P > 0.81$; low-current data, $P > 0.30$, consistent with model comparisons using BIC, mentioned above). Again, however, we are not prepared to draw strong conclusions about this feature of the

data from only a few stimulation sites in one monkey. The primary role of the high-current experiment is to show that it is possible in our paradigm to induce a decrease in confidence (i.e., an increase in T_s choices).

SUPPLEMENTAL REFERENCES

- Beck, J.M., Ma, W.J., Kiani, R., Hanks, T., Churchland, A.K., Roitman, J., Shadlen, M.N., Latham, P.E., and Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron* *60*, 1142-1152.
- Britten, K.H., Shadlen, M.N., Newsome, W.T., and Movshon, J.A. (1993). Responses of neurons in macaque MT to stochastic motion signals. *Vis Neurosci* *10*, 1157-1169.
- Chang, J.S., and Cooper, G. (1970). A practical difference scheme for Fokker-Planck equations. *J Comput Phys* *6*, 1-16.
- Hanks, T.D., Mazurek, M.E., Kiani, R., Hopp, E., and Shadlen, M.N. (2011). Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J Neurosci* *31*, 6339-6352.
- Hof, P.R., and Morrison, J.H. (1995). Neurofilament protein defines regional patterns of cortical organization in the macaque monkey visual system: a quantitative immunohistochemical analysis. *J Comp Neurol* *352*, 161-186.
- Moreno-Bote, R. (2010). Decision confidence and uncertainty in diffusion models with partially correlated neuronal integrators. *Neural Computation* *22*, 1786-1811.
- Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput* *20*, 873-922.
- van Ravenzwaaij, D., Mulder, M.J., Tuerlinckx, F., and Wagenmakers, E.J. (2012). Do the dynamics of prior information depend on task context? An analysis of optimal performance and an empirical test. *Front Psychol* *3*, 132.