



# Humans incorporate attention-dependent uncertainty into perceptual decisions and confidence

Rachel N. Denison<sup>a,b,1,2</sup>, William T. Adler<sup>b,1</sup>, Marisa Carrasco<sup>a,b</sup>, and Wei Ji Ma<sup>a,b</sup>

<sup>a</sup>Department of Psychology, New York University, New York, NY 10003; and <sup>b</sup>Center for Neural Science, New York University, New York, NY 10003

Edited by Barbara Anne Doshier, University of California, Irvine, CA, and approved September 6, 2018 (received for review October 11, 2017)

**Perceptual decisions are better when they take uncertainty into account. Uncertainty arises not only from the properties of sensory input but also from cognitive sources, such as different levels of attention. However, it is unknown whether humans appropriately adjust for such cognitive sources of uncertainty during perceptual decision-making. Here we show that, in a task in which uncertainty is relevant for performance, human categorization and confidence decisions take into account uncertainty related to attention. We manipulated uncertainty in an orientation categorization task from trial to trial using only an attentional cue. The categorization task was designed to disambiguate decision rules that did or did not depend on attention. Using formal model comparison to evaluate decision behavior, we found that category and confidence decision boundaries shifted as a function of attention in an approximately Bayesian fashion. This means that the observer's attentional state on each trial contributed probabilistically to the decision computation. This responsiveness of an observer's decisions to attention-dependent uncertainty should improve perceptual decisions in natural vision, in which attention is unevenly distributed across a scene.**

attention | perceptual decision | confidence | Bayesian | criterion

**S**ensory representations are inherently noisy. In vision, stimulus factors such as low contrast, blur, and visual noise can increase an observer's uncertainty about a visual stimulus. Optimal perceptual decision-making requires taking into account both the sensory measurements and their associated uncertainty (1). When driving on a foggy day, for example, you may be more uncertain about the distance between your car and the car in front of you than you would be on a clear day and try to keep further back. Humans often respond to sensory uncertainty in this way (2, 3), adjusting their choice (4) behavior as well as their confidence (5). Confidence is a metacognitive measure that reflects the observer's degree of certainty about a perceptual decision (6, 7).

Uncertainty arises not only from the external world but also from one's internal state. Attention is a key internal state variable that governs the uncertainty of visual representations (8, 9); it modulates basic perceptual properties like contrast sensitivity (10, 11) and spatial resolution (12). Surprisingly, it has been suggested that, unlike for external sources of uncertainty, people fail to take attention into account during perceptual decision-making (13–15), leading to inaccurate decisions and overconfidence—a risk in attentionally demanding situations like driving a car.

However, this proposal has never been tested using a perceptual task designed to distinguish fixed from flexible decision rules, nor has it been subjected to formal model comparison. Critically, as we show in *SI Appendix, section S1*, the standard signal detection tasks used previously cannot, in principle, test the fixed decision rule proposal. In standard tasks, the absolute internal decision rule cannot be uniquely recovered, making it impossible to distinguish between fixed and flexible decision rules (*SI Appendix, Fig. S1A*).

Testing whether observers take attention-dependent uncertainty into account for both choice and confidence also requires a task in which such decision flexibility stands to improve categorization performance. This condition is not met by traditional left versus right categorization tasks, in which the optimal choice boundary is the same (halfway between the means of the left and right category distributions) regardless of the level of uncertainty (*SI Appendix, Fig. S1B*). Optimal performance can be achieved simply by taking the difference between the evidence for left and the evidence for right, with no need to take uncertainty into account. The same principle applies to present versus absent detection tasks.

To overcome these limitations, we used a categorization task—which we call the embedded category task—specifically designed to test whether decision rules depend on uncertainty. Observers categorized stimuli as belonging to one of two distributions, which had the same mean but different variances (Fig. 1A). The task requires distinguishing a more specific from a more general perceptual category (4, 5), which is typical of object recognition (16, 17) (e.g., distinguishing a beagle from other dogs) and perceptual grouping (e.g., distinguishing collinear line segments from other line segment configurations) (18). In the embedded category task, the optimal choice boundaries shift as uncertainty increases, which allowed us to determine whether observers' behavior tracked these shifts, along with analogous shifts in confidence boundaries.

## Significance

**We must routinely make decisions based on uncertain sensory information. Sometimes that uncertainty is related to our own cognitive state, such as when we are not paying attention. Do our decisions about what we perceive take into account our attentional state? Or are we blind to such internal sources of uncertainty, leading to poor decisions and overconfidence? We found that human observers take attention-dependent uncertainty into account when categorizing visual stimuli and reporting their confidence in a task in which uncertainty is relevant for performance. Moreover, they do so in an approximately Bayesian fashion. Human perceptual decision-making can therefore, at least in some cases, adjust in a statistically appropriate way to external and internal sources of uncertainty.**

Author contributions: R.N.D., W.T.A., M.C., and W.J.M. designed research; R.N.D. and W.T.A. performed research; R.N.D. and W.T.A. analyzed data; and R.N.D., W.T.A., M.C., and W.J.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the [PNAS license](#).

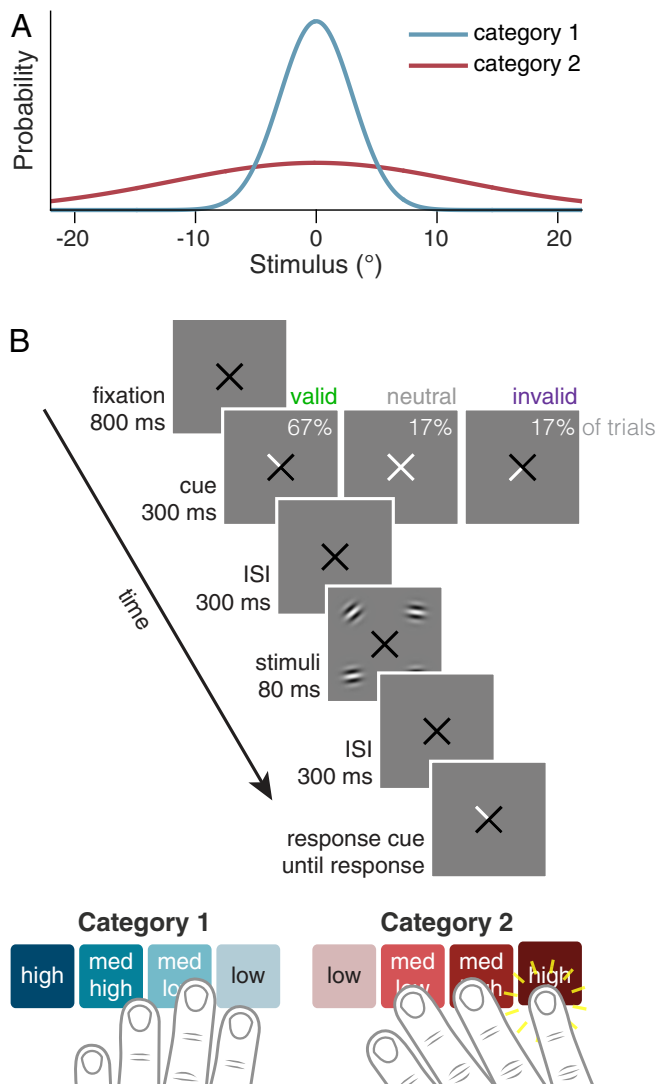
Data deposition: All data and code used for running experiments, model fitting, and plotting are available at <https://doi.org/10.5281/zenodo.1422804>.

<sup>1</sup>R.N.D. and W.T.A. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: [rachel.denison@nyu.edu](mailto:rachel.denison@nyu.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.17117720115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.17117720115/-DCSupplemental).

Published online October 8, 2018.



**Fig. 1.** Stimuli and task. (A) Stimulus orientation distributions for each category. (B) Trial sequence. Cue validity, the likelihood that a nonneutral cue would match the response cue, was 80%. Each button corresponds to a category choice and confidence rating.

## Results

Observers performed the embedded category task in which they categorized drifting grating stimuli as drawn from either a narrow distribution around horizontal ( $SD=3^\circ$ , category 1) or a wide distribution around horizontal ( $SD=12^\circ$ , category 2) (Fig. 1A) (4). Because the category distributions overlap, maximum accuracy on the task is  $\sim 80\%$ . We trained observers on the category distributions in category training trials, in which a single stimulus was presented at the fovea, before the main experiment and in short, top-up blocks interleaved with the test blocks (see *Materials and Methods*). Accuracy on category training trials in test sessions was  $71.9\% \pm 4.0\%$ , indicating that observers knew the category distributions and could perform the task well.

Four stimuli were briefly presented on each trial, and a response cue indicated which stimulus to report. Observers reported both their category choice (category 1 vs. 2) and their degree of confidence on a 4-point scale using one of eight buttons, ranging from high-confidence category 1 to high-confidence category 2 (Fig. 1B). Using a single button press for choice and confidence prevented postchoice influences on the confidence judgment (19) and emphasized that confidence should

reflect the observer's perception rather than a preceding motor response. We manipulated voluntary (i.e., endogenous) attention on a trial-to-trial basis using a spatial cue that pointed to either one stimulus location (valid condition: the response cue matched the cue, 66.7% of trials; and invalid condition: it did not match, 16.7% of trials) or all four locations (neutral condition: 16.7% of trials) (Fig. 1B). Twelve observers participated, with about 2,000 trials per observer.

Cue validity increased categorization accuracy, one-way repeated-measures ANOVA,  $F(2, 11) = 95.88$ ,  $P < 10^{-10}$ , with higher accuracy following valid cues, two-tailed paired  $t$  test,  $t(11) = 7.92$ ,  $P < 10^{-5}$ , and lower accuracy following invalid cues,  $t(11) = 4.62$ ,  $P < 10^{-3}$ , relative to neutral cues (Fig. 2A, *Top*). This pattern confirms that attention increased orientation sensitivity (e.g., refs. 11 and 20). Attention also increased confidence ratings,  $F(2, 11) = 13.35$ ,  $P < 10^{-3}$ , and decreased reaction time,  $F(2, 11) = 28.76$ ,  $P < 10^{-6}$ , ruling out speed-accuracy tradeoffs as underlying the effect of attention on accuracy (Fig. 2A, *Center and Bottom*).

Decision rules in this task are defined by how they map stimulus orientation and attention condition onto a response. We therefore plotted behavior as a function of these two variables. Overall performance was a "W"-shaped function of stimulus orientation (Fig. 2B, *Top*), reflecting the greater difficulty in categorizing a stimulus when its orientation was near the optimal category boundaries (at  $\sim 5^\circ$  with no noise). Attention increased the sensitivity of responses to the stimulus orientation (Fig. 2B).

To assess whether observers changed their category and confidence decision boundaries to account for attention-dependent orientation uncertainty, we fit two main models. In one, the Bayesian model, decisions take uncertainty into account, whereas in the other, the Fixed model, decisions are insensitive to uncertainty. Both models assume that, for the stimulus of interest, the observer draws a noisy orientation measurement from a normal distribution centered on the true stimulus value with SD (i.e., uncertainty) dependent on attention. In the Bayesian model, decisions depend on the relative posterior probabilities of the two categories, leading the observer to shift his or her decision boundaries in measurement space, based on the attention condition (4, 5) (Fig. 3A and B; *SI Appendix, Fig. S2*). The Bayesian model maximizes accuracy and produces confidence reports that are a function of the posterior probability of being correct. Note that observers could take uncertainty into account in other ways, but here we began with a normative approach by using a Bayesian model. In the Fixed model, observers use the same decision criteria, regardless of the attention condition (13, 15, 21–27) (i.e., they are fixed in measurement space; Fig. 3A and B). We used Markov chain Monte Carlo (MCMC) sampling to fit the models to raw, trial-to-trial category and confidence responses from each observer separately (*Materials and Methods* and *SI Appendix, Table S1*).

Observers' decisions took attention-dependent uncertainty into account. The Bayesian model captured the data well (Fig. 3C) and substantially outperformed the Fixed model (Fig. 3C and D), which had systematic deviations from the data. Although the fit depended on the full dataset, note deviations of the Fixed fit from the data near zero tilt and at large tilts in Fig. 3C, including failure to reproduce the cross-over pattern of the three attention condition curves that is present in the data and the Bayesian fit. To compare models, we used an approximation of leave-one-out cross-validated log likelihood that uses Pareto-smoothed importance sampling (PSIS-LOO; henceforth LOO) (29). Bayesian outperformed Fixed by an LOO difference (median and 95% CI of bootstrapped mean differences across observers) of 102 [45, 167]. This implies that the attentional state is available to the decision process and is incorporated into probabilistic representations used to make the decision.



again rejected the Fixed model (*SI Appendix, Fig. S5 A and B and Tables S3 and S4*). Therefore, category criteria, independent of confidence criteria, varied as a function of attention-dependent uncertainty.

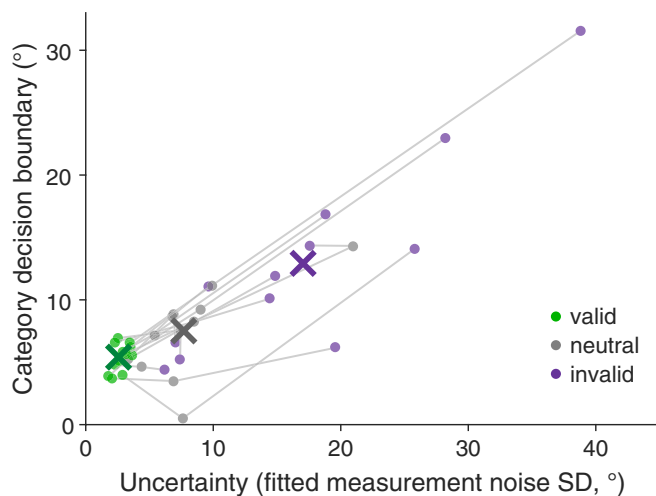
Finally, we directly tested for decision boundary shifts—the key difference between the Bayesian and Fixed models—by estimating each observer’s category decision boundaries non-parametrically. To do so, we fit the category choice data with a Free model in which the category decision boundaries varied freely and independently for each attention condition. The estimated boundaries differed between valid and invalid trials (Fig. 4 and *SI Appendix, Fig. S5C*), with a mean difference of  $7.5^\circ$  ( $SD = 7.8^\circ$ ), two-tailed paired  $t$  test,  $t(11) = 3.33$ ,  $P < 10^{-2}$ . Most observers showed a systematic outward shift of category decision boundaries from valid to neutral to invalid conditions, confirming that their choices accounted for uncertainty.

## Discussion

Using an embedded category task designed to distinguish fixed from flexible decision rules, we found that human perceptual decision-making takes into account uncertainty due to spatial attention, when uncertainty is relevant for performance. These findings indicate flexible decision behavior that is responsive to attention—an internal factor that affects the uncertainty of stimulus representations.

Our findings of flexible decision boundaries run counter to a previous proposal that observers use a fixed decision rule under varying attentional conditions (13–15, 21). This idea originated from a more general “unified criterion” proposal (25, 26), which asserts that in a display with multiple stimuli, observers adopt a single, fixed decision boundary (the unified criterion) for all items (22–27). The unified criterion proposal implies a rigid, sub-optimal mechanism for perceptual decision-making in real-world complex scenes, in which uncertainty can vary due to a variety of factors.

Although the unified criterion proposal has served to explain experimental findings (13–15, 21–27, 34), it is impossible to infer decision boundaries from behavior in the signal detection theory (SDT) tasks used previously (35). In theory, it is always possible



**Fig. 4.** Free model analysis. Group mean MCMC parameter estimates (crosses) show systematic changes in the category decision boundary across attention conditions. The same pattern can be seen for individual observers: Each gray line corresponds to a different observer, with connected points representing the estimates for valid, neutral, and invalid attention conditions. Each point represents a pair of parameter estimates: uncertainty and category decision boundary for a specific attention condition.

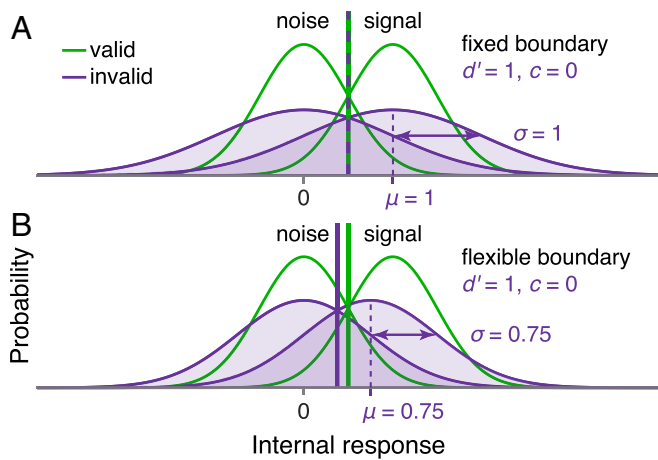
to explain behavioral data from such tasks with a fixed decision rule, as long as the means and variances of the internal measurement distributions are free to vary (*SI Appendix, section S1*).

This issue is particularly thorny for attention studies: SDT works with arbitrary, internal units of “evidence” for one choice or another, and attention could change the means, the variances, or both properties of the internal evidence distributions (10, 11, 36). As a result, the decision boundaries are unconstrained: A fixed decision boundary could be mistaken for a flexible one, and vice versa (Fig. 5). Relatedly, in a perceptual averaging task, confidence data apparently generated by a fixed decision rule can also be explained by a Bayesian decision rule with small underestimations of the internal measurement noise (37). These considerations underscore the importance of doing model comparison even for relatively simple decision models. It may be, then, that decision boundaries did change with attention in previous studies, but these changes were not inferred for methodological reasons.

Alternatively, it may be that decision boundaries truly did not change in previous studies, and task differences underlie our differing results. Studies supporting the unified criterion proposal used either detection or orthogonal discrimination (13, 15, 21–27, 34), which is often used as a proxy for detection (10, 38). In these tasks, the stimuli are low contrast relative to either a blank screen or a noisy background, and performance is limited by low signal-to-noise ratio. In our categorization task, by comparison, although maximum performance is capped due to overlap of the category-conditioned stimulus distributions, variations in performance depend on the precision of orientation representations, just as in a left vs. right fine discrimination task. Therefore, it may be that observers adjust decision boundaries defined with respect to precise features (e.g., What is the exact orientation?) but not boundaries defined with respect to signal strength (e.g., Is anything present at all?).

Other task differences could play a role as well. Some experiments matched perceptual sensitivity  $d'$  for different attention conditions by changing stimulus contrast, so attention and physical stimulus properties covaried (13, 15). For the metacognitive report, we asked for confidence rather than visibility (13); these subjective measures are known to differ (39). Finally, one study (15) using a signal detection approach suggested that observers rely insufficiently on an instructed prior, especially for unattended stimuli. The question of how attention affects the use of a prior is different from the current question, as incorporating a prior requires a cognitive step beyond accounting for uncertainty in the perceptual representation. In the future, it will be interesting to examine how decision boundaries relate to priors, attention, and uncertainty more generally in this task and other tasks in which absolute decision boundaries can be uniquely inferred.

Despite attention’s large influence on visual perception (8), only a handful of studies have examined its influence on visual confidence, with mixed results. Two studies found that voluntary attention increased confidence (30, 31), one found that voluntary but not involuntary attention increased confidence (33), and another found no effect of voluntary attention on confidence (32). This last result has been attributed to response speed pressures (30, 33). Three other studies suggested an inverse relation between attention and confidence, though these used rather different attention manipulations and measures. One study reported higher confidence for uncued compared with cued error trials (40), one found higher confidence for stimuli with incongruent compared with congruent flankers (41), and a third found that lower fMRI BOLD activation in the dorsal attention network correlated with higher confidence (21). Here, experimentally manipulating spatial attention without response speed pressure revealed a positive, approximately Bayesian, relation between attention and confidence.



**Fig. 5.** Limitations of standard SDT tasks. SDT tasks such as the detection task illustrated here cannot distinguish fixed from flexible decision rules when the means and variances of internal measurement distributions can also vary across conditions. (A) Fixed and (B) flexible decision rules give the same behavioral data (perceptual sensitivity,  $d'$ , and criterion,  $c$ ) in the two depicted scenarios, in which attention affects the measurement distributions differently (compare the invalid distributions in A and B). An experimenter could not infer from the behavioral data which scenario occurred.

The mechanisms for decision adjustment under attention-dependent uncertainty could be mediated by effective contrast (10, 42, 43). Alternatively, attention-dependent decision-making may rely on higher order monitoring of attentional state. For example, the observer could consciously adjust a decision depending on whether he or she was paying attention. Future studies will be required to distinguish between these more bottom-up or top-down mechanisms.

Our finding that human observers incorporate attention-dependent uncertainty into perceptual categorization and confidence reports in a statistically appropriate fashion points to the question of what other kinds of internal states can be incorporated into perceptual decision-making. There is no indication, for example, that direct stimulation of sensory cortical areas leads to adjustments of confidence and visibility reports (21, 44, 45), suggesting that the system is not responsive to every change to internal noise. It may be that the system is more responsive to states that are internally generated or that have consistent behavioral relevance. Attention is typically spread unevenly across multiple objects in a visual scene, so the ability to account for attention likely improves perceptual decisions in natural vision. It remains to be seen whether the perceptual decision-making system is responsive to other cognitive or motivational states.

### Materials and Methods

Extended materials and methods are available in *SI Appendix*.

**Observers.** Twelve observers (seven female) participated in the study. Observers received \$10 per 40- to 60-min session, plus a completion bonus of \$25. The experiments were approved by the University Committee on Activities Involving Human Subjects of New York University. Informed consent was given by each observer before the experiment. All observers were naïve to the purpose of the experiment, and none were experienced psychophysical observers.

**Apparatus, Stimuli, and Task.** Observers were seated in a dark room, at a viewing distance of 57 cm from the screen, with their head stabilized by a chin-and-head rest. Stimuli were presented on a gamma-corrected 100 Hz, 21-inch display (Model Sony GDM-5402) with a gray background (60 cd/m<sup>2</sup>). Stimuli were drifting Gabors with spatial frequency of 0.8 cycles per degree of visual angle (dva), speed of 6 cycles per second, Gaussian envelope with SD 0.8 dva, and randomized starting phase. In category training, the stimuli

were positioned at fixation. In all other blocks, one stimulus was positioned in each of the four quadrants of the screen (45, 135, 225, 315°), 5 dva from fixation. On each trial, each of the four stimuli was drawn independently and with equal probability from one of the two category distributions. The main task is shown in Fig. 1 and described in *Results*. Online eye tracking (Eyelink 1000) was used to ensure fixation.

**Experimental Procedure.** Each observer completed seven sessions: two staircase sessions (training, contrast staircase, prescreening) followed by five test sessions (main experiment). Observers received instructions and training for each task (see *SI Appendix*).

**Staircase Sessions.** Each staircase session consisted of three category training blocks (72 trials each) and three category/attention testing-with-staircase blocks (144 trials each), in alternation.

In category training blocks, observers learned the stimulus distributions. On each trial, category 1 or 2 was selected with equal probability, the stimulus orientation was drawn from the corresponding stimulus distribution, and the stimulus appeared at fixation for 300 ms at 35% contrast. Observers reported category 1 or 2 and received accuracy feedback after each trial.

In category/attention testing-with-staircase blocks, the trial sequence was identical to the main task (Fig. 1B), except observers reported only category choice. There was no trial-to-trial feedback on this or any other type of attention block.

We used an adaptive staircase procedure (46, 47) to estimate psychometric functions for performance accuracy as a function of contrast, separately for valid, neutral, and invalid trials. Simulations we conducted before starting the study showed that without a sufficiently large accuracy difference between valid and invalid trials, our models would be indistinguishable. Therefore, we used the psychometric function posteriors to determine whether the observer was eligible for the test sessions and, if so, to determine the stimulus contrast for that observer (see *SI Appendix*). Twenty-eight observers were prescreened, 13 were invited to participate in the main study, and 1 dropped out, leaving 12 observers, our target.

**Test Sessions.** Each test session consisted of three category training blocks (identical to staircase sessions but with observer-specific stimulus contrast) and three confidence/attention testing blocks (144 trials each), in alternation. These testing blocks were the main experimental blocks; the trial sequence is shown in Fig. 1B.

**Modeling Procedures.** The modeling procedures were similar to those used by Adler and Ma (5).

We used free parameters to characterize  $\sigma$ , the SD of orientation measurement noise, for all three attention conditions:  $\sigma_{\text{valid}}$ ,  $\sigma_{\text{neutral}}$  and  $\sigma_{\text{invalid}}$ . We added orientation-dependent noise (48).

We coded all responses as  $r \in \{1, 2, \dots, 8\}$ , with each value indicating category and confidence, as in Fig. 1B. The probability of a single trial  $i$  is equal to the probability mass of the internal measurement distribution  $p(x | s_i) = \mathcal{N}(x; s_i, \sigma_i^2)$  in a range corresponding to the observer's response  $r_i$ . We find the boundaries ( $b_{r_i-1}(\sigma_i)$ ,  $b_{r_i}(\sigma_i)$ ) in measurement space, as defined by the fitting model  $m$  and parameters  $\theta$ , and then compute the probability mass of the measurement distribution between the boundaries:

$$p_{m,\theta}(r_i | s_i, \sigma_i) = \int_{-b_{r_i}}^{-b_{r_i-1}} \mathcal{N}(x; s_i, \sigma_i^2) dx + \int_{b_{r_i-1}}^{b_{r_i}} \mathcal{N}(x; s_i, \sigma_i^2) dx, \quad [1]$$

where  $b_0 = 0^\circ$  and  $b_8 = \infty^\circ$ .

To obtain the log likelihood of the dataset, given a model with parameters  $\theta$ , we compute the sum of the log probability for every trial  $i$ , where  $t$  is the total number of trials:

$$\log p(\text{data} | \theta) = \sum_{i=1}^t \log p(r_i | \theta) = \sum_{i=1}^t \log p_\theta(r_i | s_i, \sigma_i). \quad [2]$$

To fit the model, we sampled from the posterior distribution over parameters,  $p(\theta | \text{data})$ . To sample from the posterior, we use an expression for the log posterior

$$\log p(\theta | \text{data}) = \log p(\text{data} | \theta) + \log p(\theta) + \text{constant}, \quad [3]$$

where  $\log p(\text{data} | \theta)$  is given in Eq. 2. We took uniform (or, for parameters that were standard deviations, log-uniform) priors over reasonable, sufficiently large ranges (49), which we chose before fitting any models. We sampled from the probability distribution using a MCMC method, slice sampling (50) (see *SI Appendix*).

To compare model fits while accounting for the complexity of each model, we computed the LOO (29). A bootstrapping procedure was used to compute the group mean with CIs for LOO score differences between models.

**Bayesian Model.** The Bayesian model generates category and confidence responses based on the log posterior ratio,  $d$ , of the two categories:

$$d = \log \frac{p(C=1|x)}{p(C=2|x)} = \log \frac{p(x|C=1)}{p(x|C=2)} + \log \frac{p(C=1)}{p(C=2)}. \quad [4]$$

Given the orientation measurement likelihoods,  $p(x|C)$ , and marginalizing over the stimulus  $s$ , this is equivalent to

$$d = \frac{1}{2} \log \frac{\sigma^2 + \sigma_2^2}{\sigma^2 + \sigma_1^2} - \frac{\sigma_2^2 - \sigma_1^2}{2(\sigma^2 + \sigma_2^2)(\sigma^2 + \sigma_1^2)} x^2 + \log \frac{p(C=1)}{p(C=2)}. \quad [5]$$

The observer compares  $d$  to a set of decision boundaries ( $k_0, k_1, \dots, k_8$ ), which define the eight possible category and confidence responses.  $k_4$  is the category boundary and captures possible category bias, and it is the only boundary parameter in models of category choice only.  $k_0$  is fixed at  $-\infty$  and  $k_8$  is fixed at  $\infty$ , leaving seven free boundary parameters: ( $k_1, k_2, \dots, k_7$ ) =  $k$ .

In the Bayesian models with  $d$  noise, we assume that, for each trial, there is an added Gaussian noise term on  $d$ ,  $\eta_d \sim p(\eta_d)$ , where  $p(\eta_d) = \mathcal{N}(0, \sigma_d^2)$  and  $\sigma_d$  is a free parameter.

**Fixed Model.** The observer compares the measurement to a set of boundaries that are not dependent on  $\sigma$ . We fit free parameters  $k$  and use measurement boundaries  $b_r = k_r$ .

**Linear and Quadratic Models.** The observer compares the measurement to a set of boundaries that are linear or quadratic functions of  $\sigma$ . We fit free parameters  $k$  and  $m$  and use measurement boundaries  $b_r(\sigma) = k_r + m_r\sigma$  (Linear) or  $b_r(\sigma) = k_r + m_r\sigma^2$  (Quadratic).

**Free Model.** To estimate the category boundaries with minimal assumptions, we fit free parameters  $k_{4, \text{valid}}$ ,  $k_{4, \text{neutral}}$ , and  $k_{4, \text{invalid}}$  and used measurement boundaries  $b_{4, \text{attention condition}} = k_{4, \text{attention condition}}$ .

**Data and Code Availability.** All data and code used for running experiments, model fitting, and plotting are available at <https://doi.org/10.5281/zenodo.1422804>.

**ACKNOWLEDGMENTS.** The authors thank Roshni Lulla and Gordon C. Bill for assistance with data collection and helpful discussions; Luigi Acerbi for helpful ideas and tools related to model fitting and model comparison; and Stephanie Badde, Michael Landy, and Christopher Summerfield for comments on the manuscript. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Award DGE-1342536 (to W.T.A.) and by the National Eye Institute of the National Institutes of Health under Awards T32EY007136 (to New York University) and F32EY025533 (to R.N.D.).

- Knill DC, Richards W, eds (1996) *Perception as Bayesian Inference* (Cambridge Univ Press, Cambridge, UK).
- Trommershäuser J, Kording K, Landy MS, eds (2011) *Sensory Cue Integration* (Oxford Univ Press, Oxford).
- Ma WJ, Jazayeri M (2014) Neural coding of uncertainty and probability. *Annu Rev Neurosci* 37:205–220.
- Qamar AT, et al. (2013) Trial-to-trial, uncertainty-based adjustment of decision boundaries in visual categorization. *Proc Natl Acad Sci USA* 110:20332–20337.
- Adler WT, Ma WJ (2018) Comparing Bayesian and non-Bayesian accounts of human confidence reports. bioRxiv:093203. Preprint, posted December 11, 2016.
- Massamian P (2016) Visual confidence. *Annu Rev Vis Sci* 2:459–481.
- Fleming SM, Daw ND (2017) Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psych Rev* 124:91–114.
- Carrasco M (2011) Visual attention: The past 25 years. *Vis Res* 51:1484–1525.
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647.
- Carrasco M, Penpeci-Talgar C, Eckstein M (2000) Spatial covert attention increases contrast sensitivity across the CSF: Support for signal enhancement. *Vis Res* 40:1203–1215.
- Lu ZL, Doshier BA (1998) External noise distinguishes attention mechanisms. *Vis Res* 38:1183–1198.
- Anton-Erxleben K, Carrasco M (2013) Attentional enhancement of spatial resolution: Linking behavioural and neurophysiological evidence. *Nat Rev Neurosci* 14:188–200.
- Rahnev DA, Maniscalco B, Graves T, Huang E, de Lange FP, Lau H (2011) Attention induces conservative subjective biases in visual perception. *Nat Neurosci* 14:1513–1515.
- Rahnev DA, Bahdo L, de Lange FP, Lau H (2012) Prestimulus hemodynamic activity in dorsal attention network is negatively associated with decision confidence in visual perception. *J Neurophysiol* 108:1529–1536.
- Morales J, et al. (2015) Low attention impairs optimal incorporation of prior knowledge in perceptual decisions. *Atten Percept Psychophys* 77:2021–2036.
- Liu Z, Knill DC, Kersten D (1995) Object classification for human and ideal observers. *Vis Res* 35:549–568.
- Sanborn AN, Griffiths TL, Shiffrin RM (2010) Uncovering mental representations with Markov Chain Monte Carlo. *Cogn Psychol* 60:63–106.
- Geisler WS, Perry JS (2009) Contour statistics in natural images: Grouping across occlusions. *Vis Neurosci* 26:109–121.
- Navajas J, Bahrami B, Latham PE (2016) Post-decisional accounts of biases in confidence. *Curr Op Behav Sci* 11:55–60.
- Cameron EL, Tai JC, Carrasco M (2002) Covert attention affects the psychometric function of contrast sensitivity. *Vis Res* 42:949–967.
- 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.
- Caetta F, Gorea A (2010) Upshifted decision criteria in attentional blink and repetition blindness. *Vis Cognit* 18:413–433.
- Gorea A, Caetta F, Sagi D (2005) Criteria interactions across visual attributes. *Vis Res* 45:2523–2532.
- Zak I, Katkov M, Gorea A, Sagi D (2012) Decision criteria in dual discrimination tasks estimated using external-noise methods. *Atten Percept Psychophys* 74:1042–1055.
- Gorea A, Sagi D (2000) Failure to handle more than one internal representation in visual detection tasks. *Proc Natl Acad Sci USA* 97:12380–12384.
- Gorea A, Sagi D (2001) Disentangling signal from noise in visual contrast discrimination. *Nat Neurosci* 4:1146–1150.
- Gorea A, Sagi D (2002) Natural extinction: A criterion shift phenomenon. *Vis Cognit* 9:913–936.
- Giordano AM, McElree B, Carrasco M (2009) On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *J Vis*, 10.1167/9.3.30.
- Vehdari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27:1413.
- Zizlsperger L, Sauvigny T, Haarmeier T (2012) Selective attention increases choice certainty in human decision making. *PLoS One* 7:e41136.
- Zizlsperger L, Sauvigny T, Händel B, Haarmeier T (2014) Cortical representations of confidence in a visual perceptual decision. *Nat Commun* 5:3940.
- Wilimzig C, Tsuchiya N, Fahle M, Einhäuser W, Koch C (2008) Spatial attention increases performance but not subjective confidence in a discrimination task. *J Vis*, 10.1167/8.5.7.
- Kurtz P, Shapcott KA, Kaiser J, Schmiedt JT, Schmid MC (2017) The influence of endogenous and exogenous spatial attention on decision confidence. *Sci Rep* 7:6431.
- Solovey G, Graney GG, Lau H (2015) A decisional account of subjective inflation of visual perception at the periphery. *Atten Percept Psychophys* 77:258–271.
- Kontsevich LL, Chen C-C, Verghese P, Tyler CW (2002) The unique criterion constraint: A false alarm? *Nat Neurosci* 5:707.
- Doshier BA, Lu ZL (2000) Noise exclusion in spatial attention. *Psych Sci* 11:139–146.
- Zylberberg A, Roelfsema PR, Sigman M (2014) Variance misperception explains illusions of confidence in simple perceptual decisions. *Consc Cognit* 27:246–253.
- Thomas JP, Gille J (1979) Bandwidths of orientation channels in human vision. *J Opt Soc Am* 69:652–660.
- Rausch M, Zehetleitner M (2016) Visibility is not equivalent to confidence in a low contrast orientation discrimination task. *Front Psychol* 7:591.
- Baldassi S, Megna N, Burr DC (2006) Visual clutter causes high-magnitude errors. *PLoS Biol* 4:e56.
- Schoenherr JR, Leth-Steensen C, Petrusic WM (2010) Selective attention and subjective confidence calibration. *Atten Percept Psychophys* 72:353–368.
- Ling S, Carrasco M (2006) Sustained and transient covert attention enhance the signal via different contrast response functions. *Vis Res* 46:1210–1220.
- Carrasco M, Ling S, Read S (2004) Attention alters appearance. *Nat Neurosci* 7:308–313.
- Fetsch CR, Kiani R, Newsome WT, Shadlen MN (2014) Effects of cortical microstimulation on confidence in a perceptual decision. *Neuron* 83:797–804.
- Peters MAK, et al. (2017) Transcranial magnetic stimulation to visual cortex induces suboptimal introspection. *Cortex* 93:119–132.
- Kontsevich LL, Tyler CW (1999) Bayesian adaptive estimation of psychometric slope and threshold. *Vis Res* 39:2729–2737.
- Prins N (2012) The psychometric function: The lapse rate revisited. *J Vis*, 10.1167/12.6.25.
- Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* 14:926–932.
- Acerbi L, Vijayakumar S, Wolpert DM (2014) On the origins of suboptimality in human probabilistic inference. *PLoS Comput Biol* 10:e1003661.
- Neal RM (2003) Slice sampling. *Ann Stat* 31:705–741.