How does V1 population activity inform perceptual certainty?

Zoe M. Boundy-Singer Center for Perceptual Systems, University of Texas at Austin, Austin, TX, USA

Corey M. Ziemba Center for Perceptual Systems, University of Texas at Austin, Austin, TX, USA

Olivier J. Hénaff DeepMind, London, UK

Robbe L. T. Goris Center for Perceptual Systems, University of Texas at Austin, Austin, TX, USA

Neural population activity in sensory cortex informs our perceptual interpretation of the environment. Oftentimes, this population activity will support multiple alternative interpretations. The larger the spread of probability over different alternatives, the more uncertain the selected perceptual interpretation. We test the hypothesis that the reliability of perceptual interpretations can be revealed through simple transformations of sensory population activity. We recorded V1 population activity in fixating macaques while presenting oriented stimuli under different levels of nuisance variability and signal strength. We developed a decoding procedure to infer from V1 activity the most likely stimulus orientation as well as the certainty of this estimate. Our analysis shows that response magnitude, response dispersion, and variability in response gain all offer useful proxies for orientation certainty. Of these three metrics, the last one has the strongest association with the decoder's uncertainty estimates. These results clarify that the nature of neural population activity in sensory cortex provides downstream circuits with multiple options to assess the reliability of perceptual interpretations.

Introduction

Perceptual systems infer properties of the environment from sensory measurements that can be corrupted by nuisance variability (Cox & DiCarlo, 2008) and neural noise (Faisal, Selen, & Wolpert, 2008). Consequently, sensory measurements are inherently ambiguous and perceptual inferences can be uncertain. This uncertainty limits the quality of perceptually guided behavior. To mitigate this problem, uncertain observers collect additional sensory evidence (Palmer, Huk, & Shadlen, 2005; Najemnik & Geisler, 2005), combine signals across sensory modalities (Ernst & Banks, 2002; Fetsch, Pouget, DeAngelis, & Angelaki, 2012), and leverage knowledge of statistical regularities in the environment (Weiss, Simoncelli, & Adelson, 2002; Charlton, Młynarski, Bai, Hermundstad, & Goris, 2023). These behavioral phenomena all suggest that the brain can assess the uncertainty of individual perceptual inferences, as do explicit reports of confidence in perceptual decisions (Boundy-Singer, Ziemba, & Goris, 2023; West et al., 2023). How it does so is unknown. The optimal computational strategy ("exact inference") relies on knowing how stimulus and context variables drive neural activity in sensory circuits, expressing this knowledge in a generative model, and then inverting this model for a given sensory measurement (Dayan & Abbott, 2001; Knill & Richards, 1996; Lange, Shivkumar, Chattoraj, & Haefner, 2023; Lange, Chattoraj, Beck, ad Yates, & Haefner, 2021). This operation yields a function that expresses the likelihood of every possible stimulus interpretation. The likelihood function can be used to derive the best perceptual estimate (i.e., the mode) and its uncertainty (i.e., the width). However, exact inference entails complex calculations that are intractable for many real-world perceptual tasks (Beck, Ma, Pitkow, Latham, & Pouget, 2005).

How does the brain estimate perceptual uncertainty? One appealing possibility is that it leverages certain aspects of neural activity as a direct proxy for uncertainty. For example, given that action potentials transmit information, the overall level of responsiveness of a sensory population may provide a reasonable estimate of the certainty of any inference based on that population response (Jazayeri & Movshon, 2006; Ma, Beck, Latham, & Pouget, 2006). Likewise, given

Citation: Boundy-Singer, Z. M., Ziemba, C. M., Hénaff, O. J., & Goris, R. L. T. (2024). How does V1 population activity inform perceptual certainty?. *Journal of Vision*, *24*(6):12, 1–17, https://doi.org/10.1167/jov.24.6.12.

https://doi.org/10.1167/jov.24.6.12 Received February 15, 2024; published June 17, 2024 ISSN 1534-7362

that response variability and inferential uncertainty are intimately related, spatio temporal fluctuations in neural responsiveness may provide a useful indication of perceptual uncertainty (Hoyer & Hyvärinen, 2003; Savin & Denève, 2014; Orbán, Berkes, Fiser, & Lengyel, 2016; Hénaff, Boundy-Singer, Meding, Ziemba, & Goris, 2020; Festa, Aschner, Davila, Kohn, & Coen-Cagli, 2021). And perhaps there are less intuitive aspects of sensory population activity that provide an even better indication of downstream uncertainty (Zemel, Dayan, & Pouget, 1998; Sahani & Dayan, 2003; Salmasi & Sahani, 2022; Walker, Cotton, Ma, & Tolias, 2020). Here, we examined these questions in macaque primary visual cortex (V1). We studied the problem of perceptual orientation estimation in the presence of nuisance variation and signal strength variability. We developed a model-based decoding procedure for deriving exact inference estimates of stimulus orientation and orientation uncertainty from neural population activity. This latter estimate approximates the perceptual uncertainty of a hypothetical observer inferring stimulus orientation from the recorded population responses and varies as predicted with stimulus manipulations that increase behaviorally measured orientation uncertainty (Mareschal & Shapley, 2004; Beaudot & Mullen, 2006). We compared this uncertainty estimate with different aspects of neural activity related to coding fidelity to evaluate their suitability as "candidate representations of uncertainty."

We found that the overall strength of the population response, the cross-neuron dispersion of this response, and cross-neuron variability in response gain each exhibit a modest to strong association with the decoder's orientation uncertainty. Further analysis of the relative importance of these three variables revealed that gain variability is the main driver of this association. This was true both in the presence and absence of external stimulus variability. Neural networks trained to predict orientation uncertainty from the population response reached similar performance levels as gain variability, demonstrating that our handpicked candidate representations are effective proxies for inferential uncertainty. Together, these findings illuminate how the nature of the neural code facilitates the assessment of perceptual uncertainty by circuits downstream of sensory cortex.

Methods

Physiology

All electrophysiological recordings were made from two awake fixating adult male rhesus macaque monkeys (*Macaca mulatta*, both 7 years old at the time of recording). Subjects were implanted with a titanium chamber (Adams, Economides, Jocson, Parker, & Horton, 2011), which enabled access to V1. All procedures were approved by the University of Texas Institutional Animal Care and Use Committee and conformed to National Institutes of Health standards. Extracellular recordings from neurons were made with one or two 32-channel S probes (Plexon), advanced mechanically into the brain with Thomas recording microdrives. Spikes were sorted with the offline spike-sorting algorithm Kilosort2 (Pachitariu, Steinmetz, Kadir, Carandini, & Harris, 2016), followed by manual curation with the "phy" user interface (https://github.com/kwikteam/phy). An example snippet of neural activity is shown in Figure 1B.

Apparatus

Headfixed subjects viewed visual stimuli presented on a gamma-corrected 22-in. CRT monitor (Sony Trinitron, model GDM-FW900) placed at a distance of 60 cm. The monitor had a spatial resolution of 1,280 by 1,024 pixels and a refresh rate of 75 Hz. Stimuli were presented using PLDAPS software (Eastman & Huk, 2012) (https://github.com/HukLab/PLDAPS).

Stimuli

Stimuli consisted of band-pass filtered three-dimensional (3D) luminance noise. The filter was organized around a tilted plane in the frequency domain, which specified a particular direction and speed of image motion (i.e., it was velocity-separable). All stimuli had a central spatial frequency of 2.5 cycles/deg with a bandwidth of 0.5 octaves and a central temporal frequency of 2.5 deg/sec with a bandwidth of 1 octave. Orientation bandwidth was either 3° or 90°, corresponding to the "low" and "high" level of dispersion. Stimulus contrast was computed by normalizing the summed orientation amplitude spectrum of each stimulus frame with the summed amplitude spectrum of a reference grating with matching spatial frequency. We equated contrast across orientation bandwidths by rescaling the filter output appropriately. The stimulus set was composed of four stimulus families (two orientation dispersion levels × two contrast levels), which each contained 16 differently oriented stimuli, evenly spaced between 0° and 337.5°. For each stimulus condition, we generated five unique stimulus versions by using a different noise seed. Stimuli were presented within a vignette with a diameter of 3°, centered on the estimated average receptive field location, determined through a hand-mapping procedure. Stimuli were presented in random order for 1,000 ms each. Trials were excluded from further analysis if fixation was not maintained

within a radius of 0.8 degrees from the fixation point for the duration of the stimulus presentation. The number of repeats per condition varied from session to session. In Experiment 1, the average number of repeats per stimulus condition was 28.3 ± 5.6. In Experiment 2, one stimulus was randomly selected to be overrepresented. In these recordings, the average number of repeats for nonselected stimuli and the single randomly selected stimulus was 16.9 ± 2.1 and 1,089 ± 98.9, respectively.

Data analysis: Single units and unit pairs

We first studied the units' orientation tuning in response to the high-contrast, low-dispersion stimuli (examples shown in Figure 1C). For each unit, we chose a response latency by maximizing the stimulus-associated response variance (Smith, Majaj, & Movshon, 2005) and counted spikes within a 1,000-ms window following response onset. We visually inspected tuning curves and excluded untuned units from further analysis (27% of units). For each unit, we estimated the preferred stimulus orientation by taking the mode of a circular Gaussian function fit to the neural responses. We estimated each unit's orientation selectivity (OSI) using the following equation (Leventhal, Thompson, Liu, Zhou, & Ault, 1995):

$$\mathrm{OSI} = \frac{|\sum_j R_j e^{i2\theta_j}|}{\sum_j |R_j|} \quad (1)$$

where R_j is the mean firing rate and θ_j the orientation for the *jth* stimulus. The OSI values shown in Figure 1D were directly calculated from the observed responses. We next studied the impact of our stimulus manipulations on the units' orientation tuning. For this analysis, we fit four circular Gaussian functions (one per stimulus family) to the responses of each unit. The Gaussian quartet shared the same mode across stimulus families but could vary in their amplitude and bandwidth. The changes in response amplitude and selectivity reported in the Results section were calculated from the fitted functions. Changes in gain variability with stimulus manipulations were computed by using gain variability estimates from fitting the modulated Poisson model (Goris, Movshon, & Simoncelli, 2014) per unit and stimulus family. Finally, we asked how our stimulus manipulations impacted statistical response dependencies among pairs of neurons. For each pair of simultaneously recorded neurons, we estimated their "noise correlation" by computing the Pearson correlation between their responses after removing the effects of stimulus condition on response mean and standard deviation by z-scoring responses.

Data analysis: Population Fisher information

For each recording, we computed population Fisher information (FI) per stimulus family using the method proposed in Kanitscheider, Coen-Cagli, and Kohn Pouget (2015a) (example shown in Figure 1G). This method requires that the number of repeated trials, T, exceeds the number of units, N, by about a factor of 2: $T > (N + 2)/2$. For some of our populations, this requirement is not met. To circumvent this violation, we combined stimulus conditions whose orientation differed by 180° since these only differed in their drift direction. For each population, we summarized the FI per stimulus family by computing the median across stimulus orientations. Finally, to assess the impact of noise correlations, we compared FI with shuffled FI, calculated using the method proposed by Kanitscheider et al. (2015a).

Stimulus encoding model

We fit responses of individual V1 units with the stochastic normalization model, a model inspired by the original work of Hubel and Wiesel (1959) and composed of elements introduced by many later studies (Heeger, 1992; Goris, Simoncelli, & Movshon, 2015; Hénaff et al., 2020; Coen-Cagli & Solomon, 2019). The model consists of a canonical set of linear–nonlinear operations and describes how band-pass filtered noise stimuli are transformed into the firing rate of a V1 cell. Stimuli are first processed by a linear filter whose output is half-wave rectified. This single filter is sufficient to capture the overall response of both simple and complex V1 cells to the diverse oriented content in one stimulus presentation. Following earlier modeling work that involved similar stimuli (Goris et al., 2015), the spatial profile of the linear filter is given by a derivative of a two-dimensional (2D) Gaussian function. At the preferred spatial frequency, the orientation selectivity of this filter depends on the aspect ratio of the Gaussian, α, the order of the derivative, b, and the directional selectivity, d:

$$r_\theta(\theta; \theta_o, \alpha, b, d) \propto \left[1 + \frac{d}{2}(\mathrm{sgn}(\cos(\theta - \theta_0)) - 1)\right] \cdot \left[\cos(\theta - \theta_o) \cdot \exp\left(-\frac{1}{2}(1 - \alpha^2)\cos^2(\theta - \theta_o)\right)\right]^b, \quad (2)$$

where θ is stimulus orientation, θ_o is the filter's preferred orientation, and parameter $d \in [0, 1]$ determines direction selectivity. The function $sgn(\cdot)$ computes the sign of the argument. Because spatial frequency was not systematically varied in our stimulus set, it is not possible to uniquely determine both α and b from the neural responses we observed (Goris et al.,

2015). As such, we set the derivative order to 2 unless the best-fitting aspect ratio reached an upper limit of 5—more extreme values correspond to spatial receptive fields that are atypically elongated for V1 (Goris et al., 2015). The filter's stimulus response, $f(S)$, was computed as the dot-product of the filter and stimulus profile in the orientation domain:

$$f(S) = \sum_{\theta} r_{\theta}(\theta) \cdot S(\theta). \quad (3)$$

The amplitude and width of the stimulus profile directly reflect the stimulus' contrast and dispersion, respectively. In the second stage of the model, the filter's stimulus response is converted into a deterministic firing rate, $\mu(S)$, by subjecting the filter output to divisive normalization and passing the resulting signal through a power-law nonlinearity. This step also involves the inclusion of two sources of spontaneous discharge (one simply adds to the stimulus drive, and the other is suppressed by stimuli that fail to excite the neuron) and a scaling operation:

$$\mu(S) = e_1 + \gamma \left(\frac{e_2 + f(S)}{\beta + \sum_j f_j(S)} \right)^q, \quad (4)$$

where e_1 and e_2 control the spontaneous discharge, γ the response amplitude, q the transduction nonlinearity, while stimulus independent constant, β, and the aggregated stimulus-drive of a pool of neighboring neurons, $\Sigma_j f_j(S)$, provide the normalization signal.

Neural responses vary across repeated stimulus presentations. To capture this aspect of the data, the model describes spikes as arising from a doubly stochastic process—specifically, a Poisson process subject to "gain modulations" originating from noisy normalization signals with standard deviation σ_ϵ (Goris et al., 2014; Hénaff et al., 2020). Under these assumptions, spike count variance, σ_N^2, is given by

$$\sigma_N^2 = \mu(S)\Delta t + \sigma_G^2(\mu(S)\Delta t)^2, \quad (5)$$

where Δt is the size of the counting window and σ_G the standard deviation of the response gain, given by

$$\sigma_G = \frac{\sigma_\epsilon \cdot q}{\beta + \sum_j f_j(S)}. \quad (6)$$

This model has 11 free parameters in total: 4 filter parameters (orientation preference θ_o, spatial aspect ratio α, derivative order b, and directional selectivity d), 4 parameters controlling response range and amplitude (constant β, scalar γ, and maintained discharge e_1 and e_2), 1 parameter for the nonlinearity (exponent q), 1 parameter for the normalization noise (σ_ϵ), and 1 final parameter that controlled the degree to which the normalization signal depended on stimulus dispersion. We computed the model prediction for every trial and used a Bayesian optimization algorithm to find the best-fitting parameters (Acerbi & Ma, 2017). An example model fit is shown in Figure 2A. To assess the goodness of fit, we computed the Pearson correlation between predicted and observed response mean and variance across all stimulus conditions (Figure 2B). Units were excluded from further analysis if the Pearson correlation fell below 0.5 for response mean or below 0.2 for response variance. In total, 352 out of 378 candidate units (93.1%) met this threshold.

Decoding V1 population activity

We leveraged the stimulus encoding model to decode V1 population activity on a trial-by-trial basis, building on the method proposed in Hénaff et al. (2020). Specifically, assuming that the recorded neurons fire independently from one another, we modeled a pattern of spike counts $\{K_i\}$ realized during a window of length Δt using a negative binomial distribution (Goris et al., 2014):

$$\begin{aligned} \log p(\{K_i\}|S) &= \log \prod_{i=1}^{n} p(K_i|S) \\ &= \sum_{i=1}^{n} \log \Gamma(K_i + 1/\sigma_{G,i}^2) - \log \Gamma(K_i + 1) \\ &\quad - \log \Gamma(1/\sigma_{G,i}^2) + K_i \log(\sigma_{G,i}^2 \lambda_i) \\ &\quad - (K_i + 1/\sigma_{G,i}^2) \log(1 + \sigma_{G,i}^2 \lambda_i) \end{aligned} \quad (7)$$

where the rate $\lambda_i = \mu_i(S)\Delta t$ and gain variability $\sigma_{G,i}$ are given by the stochastic normalization model (Equations 4 and 6).

In our experiment, the stimulus was fully defined by three parameters: its peak orientation θ_S, its spatial contrast c_S, and its orientation dispersion σ_S. To obtain the orientation likelihood function for a given trial, we first calculated the likelihood of each possible parameter combination in a finely sampled 3D grid (orientation: [0:.5:360°], contrast: [0.07:.05:1.4], and dispersion: [1:3.4:99°]). We then marginalized across the contrast and dispersion dimension, yielding the orientation likelihood function (examples shown in Figure 2D). This function typically appeared Gaussian in shape (the Pearson correlation coefficient between the likelihood function and best-fitting Gaussian was on average $r = 0.985$). We therefore used the peak of the best-fitting Gaussian as the point estimate of stimulus orientation. The width of the Gaussian defines the uncertainty of this estimate. Alternative ways to

quantify both statistics that did not involve fitting a Gaussian function yielded highly similar values.

Candidate metrics of uncertainty

We evaluated different aspects of V1 population activity as proxies for the decoder's orientation uncertainty. Specifically, we considered response magnitude, R_M (which indicates certainty, the inverse of uncertainty); response dispersion, R_D; and cross-neuron gain variability, R_G, computed as

$$R_M = \frac{1}{M}\sum_{i=1}^{M} r_i, \quad (8)$$

$$R_D = \sqrt{\frac{\sum (r_i - R_M)^2}{M}}, \textit{and} \quad (9)$$

$$R_G = \frac{\overline{\sigma_\epsilon} \cdot \overline{q}}{\overline{\beta} + (s_n \cdot R_M)}, \quad (10)$$

where M is population size; $\overline{\sigma_\epsilon}$, $\overline{q}$, and $\overline{\beta}$ are the mean values of the corresponding parameter estimates across all units (Equations 4 and 6); and s_n is a fixed scalar parameter used to relate the observed response magnitude to the unobserved stimulus-driven component of the normalization signal (its value was estimated through simulation of the stochastic normalization model).

These metrics are intended to reflect properties of the sensory population activity that informs perceptual interpretations. Due to the limited size of our recorded populations, estimated values will typically deviate from the ground truth. We therefore also considered "inferred estimates" obtained by leveraging knowledge of each unit's tuning properties. Specifically, for each trial, we estimated the most likely stimulus orientation, contrast, and dispersion under our encoding model:

$$\hat{\theta}_S, \hat{c}_S, \hat{\sigma}_S = \arg\max_{\theta_S, c_S, \sigma_S} \log p(\{K_i\} | S(\theta_S, c_S, \sigma_S)). \quad (11)$$

We found this maximum likelihood stimulus estimate via gradient descent using a Bayesian optimization algorithm (Acerbi & Ma, 2017). We then used the encoding model to compute the expected response magnitude, response dispersion, and gain variability for this specific stimulus. The first and last estimates were obtained by computing the cross-unit average of Equations 4 and 6, while the expected standard deviation of the population response is given by

$$\sigma_N = \sqrt{\sum_{i=1}^{M} \frac{1}{M}(\sigma_{G,i}^2 \mu_i^2 + \mu_i + \mu_i^2) - \mu^2}, \quad (12)$$

where $\sigma_{G,i}$ and μ_i are the expected gain-variability and response average of neuron i and μ is the expected magnitude of the population response.

We examined our ability to recover each of the direct and inferred metrics from populations of various sizes simulated from the stochastic normalization model. To simulate an idealized population, for most parameters, we used the median value from the model fits to our recorded units. We constrained θ_o to evenly tile the orientation domain, and we randomly drew γ for each simulated unit from an exponential distribution (Baddeley et al., 1997). We simulated responses for our Experiment 1 with two levels of contrast, two levels of orientation dispersion, 16 orientations, and 35 repeated trials per stimulus condition. Each direct and inferred metric was computed from these population responses as in our experimental data and then compared to the ground-truth values from the simulations (results summarized in Figure 3A).

Estimating partial correlations

We developed a method to estimate the association between each candidate metric and orientation uncertainty while controlling for one other candidate metric (i.e., a statistic known as a partial correlation). First, we z-scored each metric and then rank-ordered the trials as a function of the metric we sought to control for. We considered nonoverlapping bins of 50 consecutive trials and computed the variance of the "frozen" metric. If this value did not exceed a threshold level of $\sigma^2 = 0.005$, we proceeded to calculate the Spearman correlation between the "test" metric and the likelihood width for that set of trials. This analysis yielded a distribution of partial correlation values, shown in Figure 4. We computed the average of each distribution on the Fisher z-transformed values and then used the inverse transformation to map this back onto a scale from [–1, 1] (triangles in Figures 4B, C).

Artificial neural network

We trained a family of feed-forward multilayer perceptron (MLP) neural networks on trial-by-trial spike count vectors to predict trial-by-trial ground-truth likelihood width estimates. A unit's spike count was included only if a unit was present for the entire duration of the experiment. On average, 78% of units' spike counts were used. We implemented networks with the TensorFlow framework with the AdamW optimizer with an objective to minimize the mean squared error between ground truth and network-predicted likelihood. Training proceeded in two phases.

Phase 1: For each dataset, we trained 360 unique models that varied in the number of hidden layers

(1, 2, or 3), number of hidden units per layer (10, 20, 30, 40, or 50), dropout rate between layers (0.05, 0.1, 0.15, or 0.2), weight decay (0.0, 0.001, or 0.001), and the learning rate (0.001 or 0.0001). For each configuration of hyper-parameters, we trained five networks on 80% of trials (training/validation set) and obtained a cross-validated prediction on the held-out 20% of trials, rotating trials between training and held-out set such that each trial had a cross-validated prediction. With this grid search, we determined the set of hyper-parameters that minimized each dataset's held-out loss.

Phase 2: Optimal hyper-parameters in Phase 1 differed across datasets. We used this range of optimal hyper-parameters to train 96 networks on each dataset (each one using a different hyper-parameter configuration; hidden layers: 2 or 3; hidden units: 20, 30, 40, or 50; dropout rate: 0.05, 0.1, 0.15, or 0.2; weight decay: 0.0, 0.001, or 0.001; and learning rate: 0.001) and computed their held-out loss as before. These held-out losses represent an ensemble-based estimate (Lakshminarayanan, Pritzel, & Blundell, 2017) of the MLP model class's predictive accuracy, which we report in our results.

Results

Orientation coding in V1 loses precision under high nuisance variation and low signal strength

Our approach to evaluate candidate representations of perceptual uncertainty relies on three elements: first, a rich stimulus set, which elicits variable amounts of perceptual uncertainty about a stimulus feature of interest; second, observation of the joint spiking activity of a population of sensory neurons selective for this feature; and third, a method to obtain ground-truth estimates of perceptual uncertainty on a single-trial basis. We reasoned that the primary visual cortex offers a fruitful test-bed for our approach. V1 neurons are selective for local image orientation (Hubel & Wiesel, 1959), and their activity is thought to inform perceptual orientation estimates (Nienborg & Cumming, 2014; Seidemann & Geisler, 2018; Goris, Ziemba, Stine, Simoncelli, & Movshon, 2017). V1 activity not only reflects stimulus orientation. It is also modulated by factors that impact perceptual uncertainty about stimulus orientation. These include reductions of signal strength by lowering stimulus contrast (Tolhurst Movshon, & Dean, 1983; Mareschal & Shapley, 2004) and increases of nuisance variation by widening orientation dispersion (Goris et al., 2015; Beaudot & Mullen, 2006). Building on these earlier findings, we constructed a stimulus set consisting of filtered 3D luminance noise with filter settings varying in center direction of motion (16 levels), orientation dispersion (2 levels), and amplitude (2 levels; see Methods, Figure 1A).

We used linear multielectrode arrays to record the joint spiking activity of 13 ensembles of V1 units from two fixating macaque monkeys. Ensembles ranged in size from 10 to 61 units. We presented stimuli within a 3° wide circular window, centered on the population receptive field. Stimuli were generally located close to the fovea (mean eccentricity = 2.9°), where receptive fields of V1 cells are comparatively small (Gattass, Gross, & Sandell, 1981). Stimuli thus overlapped with both the classical receptive field and its inhibitory surround (Angelucci et al., 2002). Most units responded selectively to our stimulus set (Figure 1B), and exhibited regular orientation tuning for the high-contrast, narrowband stimuli (Figure 1C). For each unit, we calculated the preferred direction of motion and the sharpness of its orientation tuning (see Methods). Inspection of the distribution of both statistics within each population revealed that orientation preference was typically approximately uniformly distributed (Rayleigh test for nonuniformity was not significant for 10 out of 13 populations with $\alpha = 0.05$; Figure 1D, top) and that tuning sharpness varied considerably across units (Figure 1D, bottom) (Goris et al., 2015).

A stimulus feature can be reliably estimated from neural population activity when that activity reliably varies with the feature. Whether or not this is the case depends on how the mean and variability of the neural response relate to the stimulus. Consider the effects of stimulus contrast and stimulus dispersion on the mean response of an example unit. Lowering contrast reduced the amplitude and, to a lesser degree, the selectivity of the response (Figure 1E, dark vs. light lines). This was true for most of our units (median amplitude reduction: 32.6%, $p < 0.001$, Wilcoxon signed rank test; median OSI reduction: 18.9%, $p < 0.001$). Increasing stimulus dispersion also reduced the example unit's response amplitude. In addition, this manipulation substantially broadened the tuning function (Figure 1E, red vs. blue lines). Again, these effects were exhibited by most units (median amplitude reduction: 31.6%, $p < 0.001$; median OSI reduction: 64.2%, $p < 0.001$).

Our stimulus manipulations also impacted cross-trial response variability. Neurons in visual cortex typically exhibit super-Poisson variability, meaning that spike count variance exceeds spike count mean (Figure 1F, symbols). This behavior is well captured by a statistical model of neural activity in which spikes arise from a Poisson process and response gain fluctuates from trial to trial (Goris et al., 2014; Goris, Ziemba, Movshon, & Simoncelli, 2018) (the modulated Poisson model, Figure 1F, lines). The larger the gain variability, σ_G, the larger the excess spike count variance. Our recent work

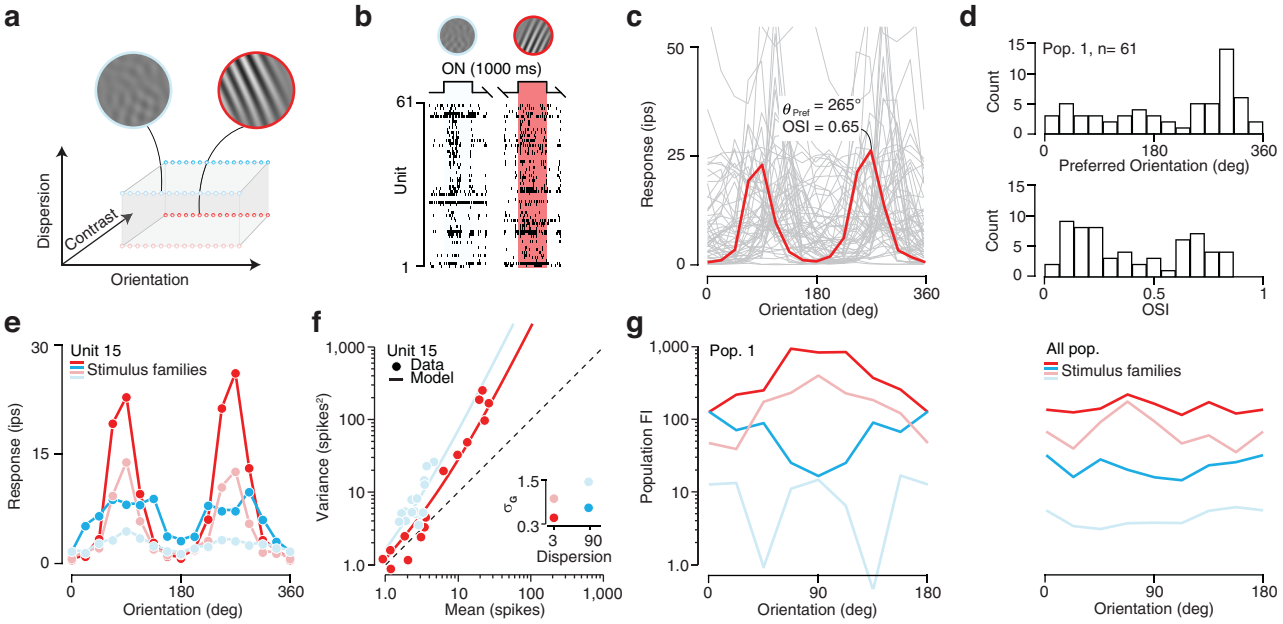

Figure 1. Orientation coding in V1 under different levels of nuisance variation and signal strength. (**a**) We created four stimulus families, each defined by the amount of orientation dispersion and contrast energy. Each family consisted of 16 stimuli with orientations evenly spaced from 0° to 337.5°. The example stimuli share the same orientation but differ along the two other dimensions. (**b**) Raster plot illustrating snippets of neural activity recorded during two different trials from Population 1. An example frame from each trial's dynamic stimulus is shown on top. (**c**) Mean response to low-dispersion, high-contrast stimuli is plotted as a function of stimulus orientation for all units from Population 1 (a single example is highlighted in red). (**d**) Distribution of preferred orientation (top) and orientation selectivity (bottom) for Population 1. (**e**) Mean response as a function stimulus orientation for an example unit. Different colors indicate different stimulus families. (**f**) The example unit's variance-to-mean relationship for two stimulus families (each data point corresponds to a different stimulus orientation). The solid line illustrates the fit of the modulated Poisson model. Inset: gain-variability, σ_G, as estimated by the modulated Poisson model for each stimulus family for this example unit. (**g**) Population Fisher information plotted against stimulus orientation for Population 1, split by stimulus family (left). Population Fisher information plotted against stimulus orientation averaged across all 13 populations (right).

suggests that the strength of gain fluctuations increases with stimulus manipulations that increase orientation uncertainty (Hénaff et al., 2020). Accordingly, we fit the modulated Poisson model to each unit separately per stimulus family (i.e., a specific combination of stimulus contrast and orientation dispersion). Lowering contrast and increasing dispersion both increased gain variability (median increase: 27.1%, $p < 0.001$ for contrast, and 12.8%, $p < 0.001$ for dispersion, Wilcoxon signed rank test), consistent with our previous observations in anesthetized animals (Hénaff et al., 2020).

To quantify the collective impact of these effects on the population representation of stimulus orientation, we computed the population Fisher information (I_θ, see Methods). This statistic expresses how well stimulus orientation can be estimated from population activity by an optimal decoder and is inversely related to the uncertainty of this estimate (Paradiso, 1988). Consider the Fisher information profiles of an example population. Reducing stimulus contrast and increasing orientation dispersion both lowered I_θ, as is evident from the vertical separation of the colored lines (Figure 1G, left). These effects were present in all of our recordings (Figure 1G, right), though the exact impact differed somewhat across populations. Trial-to-trial response fluctuations are often correlated among neurons (Cohen & Kohn, 2011). These so-called noise correlations can affect the coding capacity of neural populations (Moreno-Bote et al., 2014; Kanitscheider, Coen-Cagli, & Pouget, 2015b). However, we found no systematic effect of our stimulus manipulations on the average strength of pairwise response dependencies ($p = 0.61$ for contrast, Wilcoxon rank-sum test and $p = 0.10$ for dispersion; Supplementary Figure S1A). Moreover, a shuffling analysis revealed that noise correlations had minimal impact on Fisher information estimates (see Methods; Supplementary Figures S1B, C). This result may in part be due to the relatively small sizes of our populations (Moreno-Bote et al., 2014). Nevertheless, we conclude that under these experimental conditions, our stimulus reliability manipulations substantially impact orientation coding in V1 because they alter the relation between stimulus orientation, on the one hand, and response mean and response variance, on

the other hand. More generally, these observations indicate that, as expected, the perceptual uncertainty of an observer inferring stimulus orientation from these V1 population responses would increase as contrast decreases (Tolhurst et al., 1983; Mareschal & Shapley, 2004) and orientation dispersion increases (Goris et al., 2015; Beaudot & Mullen, 2006).

The width of the likelihood function captures coding fidelity at the single-trial level

To evaluate candidate representations of uncertainty, we need to know which trials yielded a high-quality representation of stimulus orientation and which did not. Population Fisher information cannot be used for this because its calculation takes into account response variation across trials. Instead, we developed a decoding method to infer stimulus features from population activity on a trial-by-trial basis (Jazayeri & Movshon, 2006; Graf, Kohn, Jazayeri, & Movshon, 2011; Berens et al., 2012; Shooner et al., 2015; Van Bergen, Ma, Pratte, & Jehee, 2015; Arandia-Romero, Tanabe, Drugowitsch, Kohn, & Moreno-Bote, 2016; Walker et al., 2020). Specifically, we built a stimulus–response model that describes how our stimulus set drives spiking activity of V1 neurons and then inverted this model to obtain an estimate of the likelihood of every possible stimulus orientation on the basis of a single-trial population response. We described each unit's responses with a previously proposed model of V1 function (the stochastic normalization model (Heeger, 1992; Hénaff et al., 2020); see Methods; Supplementary Figure S2). In this model, stimuli are processed in a narrowly tuned excitatory channel and a broadly tuned inhibitory channel, and spikes arise from a modulated Poisson process (Goris et al., 2015; Hénaff et al., 2020). As can be seen for an example unit, the model captures how both spike count mean and spike count variance depend on stimulus orientation, stimulus contrast, and stimulus dispersion (Figure 2A). To evaluate the model's goodness of fit, we calculated the association between predicted and observed response mean and variance (Figure 2B). The association was generally high (median Pearson correlation across all units was 0.83 for response mean and 0.80 for response variance; Figure 2C). For each unit, the model predicts how probable a given response is for a given stimulus condition. Conversely, knowledge of this probabilistic relation allows inferring how likely a given stimulus is in light of an observed response (Geisler & Albrecht, 1995). Computing this value for all possible stimuli yields a likelihood function (see Methods). Intuitively, this function summarizes which stimulus interpretations can be ruled out and which cannot. The likelihood function provides a framework to study neural coding that naturally generalizes to the population level. If neural responses are statistically independent, the joint likelihood function of a population of neurons is simply given by the product of the likelihood functions of all units. Given that noise correlations were generally small and did not significantly impact orientation coding capacity in our data (Supplementary Figure S1), we opted to use this formulation (Hénaff et al., 2020) (see Methods).

Consider the likelihood function for two example trials. This function was obtained by marginalizing the full likelihood function over the dimensions of stimulus contrast and stimulus dispersion (see Methods). Population activity differed substantially across these two trials (Figure 1B). Accordingly, the corresponding likelihood functions also differed in central tendency and shape (Figure 2D). For each trial, we used the maximum of the marginalized likelihood function as a point estimate of stimulus orientation. Overall, this estimate correlated well with the veridical stimulus orientation (circular correlation computed across all stimulus conditions, $r = 0.76$ on average and ranged from 0.96 for high-contrast, low-dispersion stimuli to 0.43 for low-contrast, high-dispersion stimuli; Supplementary Figures S3A, B). However, it did not track stimulus orientation perfectly. On some trials, the orientation estimation error could be substantial, especially when stimulus dispersion was high (Figure 2E, left). This pattern reflected an underlying structure in the distribution of estimation error. Specifically, as expected from a well-calibrated model, the spread of the estimation error approximately matched the width of the likelihood function (Figure 2E, right). Consequently, the average width of the likelihood function was strongly associated with the average size of the estimation error (Spearman's rank correlation coefficient: $r = 0.91$, $p < 0.001$; Figure 2F) and with population Fisher information (Spearman's rank correlation coefficient: $r = -0.81$, $p < 0.001$; Figure 2G). Together, these results establish the width of the likelihood function calculated under the stochastic normalization model as a principled metric of coding quality for our dataset. In the following analyses, we will use it as the "ground-truth" estimate of stimulus uncertainty on a trial-by-trial basis.

Some aspects of population activity predict the decoder's stimulus uncertainty

Might certain transformations of neural activity in sensory cortex provide a direct proxy for perceptual uncertainty? Motivated by previous theoretical proposals (Ma et al., 2006; Jazayeri & Movshon, 2006; Olshausen & Field, 1996; Orbán et al., 2016; Hénaff et al., 2020), we evaluate three different aspects of V1 population activity as candidate representations of coding quality: (1) the overall strength of the population

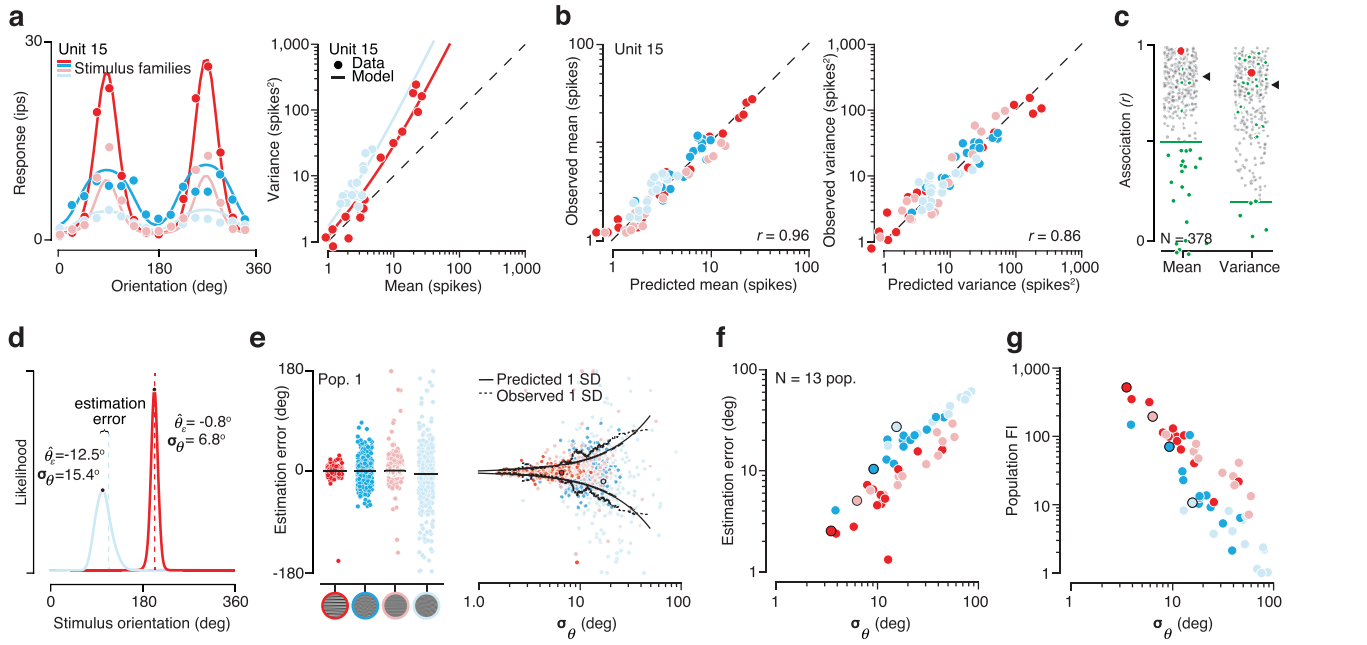

Figure 2. Quantifying orientation uncertainty at the single-trial level. (**a**) Left, mean response of an example unit as a function of orientation for all four stimulus families. Right, response variance plotted as a function of response mean for two stimulus families (one data point per orientation). Solid lines indicate the fit of the stochastic normalization model. (**b**) Observed versus predicted response mean (left) and response variance (right) for the example unit. (**c**) Distribution of the association between predicted and observed response mean (left) and response variance (right) across all units. Black triangles indicate median value across all units. Green line indicates the inclusion criterion. Units were included if the association between predicted and observed mean and variance exceeded both criteria. Excluded units are indicated with green circles. The red circle indicates the example unit illustrated in panel a, b. (**d**) Likelihood functions (solid lines) for two example trials (blue, low-contrast, high-dispersion stimulus; red, high-contrast, low-dispersion stimulus). Dashed line indicates the stimulus orientation. The black circle indicates the maximum likelihood orientation estimate. (**e**) Left, estimation error split by stimulus family for one example population. Each dot indicates a single trial. Black lines indicate the median estimation error. Right, estimation error as a function of likelihood width. Each dot indicates a single trial; color indicates stimulus family. Solid black line indicates the theoretically expected relationship between the variance of estimation error and likelihood width. Dashed line indicates empirically observed variance of estimation error calculated as a moving standard deviation of 100 trials sorted by likelihood width. (**f**) Average absolute estimation error per stimulus family and population as a function of the average likelihood width. Observations from Population 1 are highlighted with a black outline. (**g**) Population Fisher information plotted against average likelihood width; same conventions as panel f.

response ("response magnitude"), (2) cross-neuron variability in responsiveness ("response dispersion" as absolute measure and "relative dispersion" as normalized measure), and (3) cross-neuron variability in excitability ("gain variability"). For each metric, we consider two variants: one that can be estimated directly from the observed population response ("direct estimates") and one that takes into account knowledge of the units' stimulus–response relation ("inferred estimates"; Figure 3A). We selected these metrics because they each capture an aspect of neural activity that is thought to either influence or reflect coding fidelity. Specifically, given that cortical responses appear subject to Poisson-like noise (Tolhurst et al., 1983; Shadlen & Newsome, 1998; Goris et al., 2014), response magnitude is a proxy for the signal-to-noise ratio of the neural response. Cross-neuron response variability, on the other hand, in part reflects the selectivity of the stimulus drive. It will only be high for trials in which a subset of neurons is highly active—an activation pattern that allows for unambiguous stimulus inference (Geisler & Albrecht, 1995) and that satisfies the goal of coding efficiency (Olshausen & Field, 1996). Finally, it has been shown previously that cross-trial gain variability in V1 tracks orientation uncertainty within single units (Hénaff et al., 2020) (also see Figure 1F, inset). This relationship may generalize to cross-neuron gain variability within populations.

We reasoned that proxies for perceptual uncertainty ought to meet three requirements to be useful: (1) distinguish reliable from unreliable stimulus conditions, (2) predict stimulus uncertainty on a trial-by-trial basis, and (3) predict variability in stimulus uncertainty solely due to internal sources. Motivated by this logic, we first

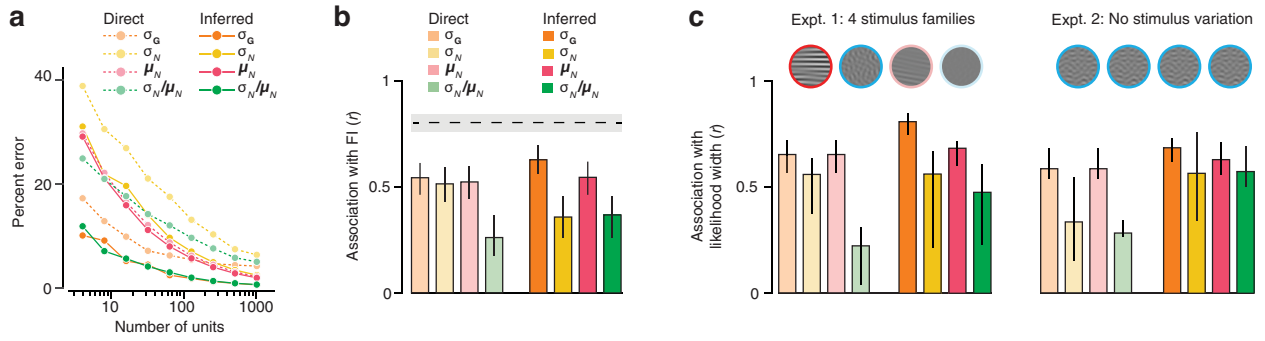

Figure 3. Evaluating different candidate metrics of uncertainty. (**a**) Percent error in recovery of the ground-truth value of each direct and inferred metric for simulated populations of various sizes (gain variability, σ_G; response dispersion, σ_N; response magnitude, μ_N; relative dispersion, σ_N/μ_N). (**b**) Spearman rank correlation between candidate uncertainty metrics and population Fisher information. Light colors correspond to direct metrics; dark colors correspond to inferred metrics. Error bars indicate 75% confidence intervals of bootstrapped correlation values. The dashed line indicates the rank correlation between likelihood width and Fisher information with 75% confidence interval of bootstrapped correlation values indicated with a gray band. (**c**) Left/top, example sequence of stimuli for Experiment 1. Stimuli randomly varied in orientation, family, and noise seed. Left/bottom, bar plot indicating the association between candidate neural metrics of uncertainty and likelihood width. Right/top, example sequence of stimuli for the included Experiment 2 trials with no variation in stimulus orientation, family, or noise seed. A single orientation, family, and noise seed defined a stimulus that was overrepresented relative to all other stimulus conditions. Only trials in which this stimulus was shown were analyzed. Right/bottom, bar plot indicating the association between candidate neural metrics of uncertainty and likelihood width for these trials. Error bars indicate interquartile range.

examined the relationship between the selected metrics and Fisher information. We computed the average value of each metric for each stimulus family, just like we did previously for the width of the likelihood function (Figures 2F, G). Consider the direct estimates. Response magnitude, response dispersion, and gain variability each exhibited a modest association with Fisher information ($r = 0.52$, 0.51, and 0.54, with 95% confidence intervals ranging from 0.31–0.69, 0.30–0.69, and 0.35–0.70; Figure 3B, left) but fell short of the predictive power of the width of the likelihood function ($r = 0.80$, 95% confidence interval 0.68–0.88). By comparison, relative dispersion had a substantially weaker correlated with Fisher information ($r = 0.26$, 95% confidence interval 0.02–0.47; Figure 3B, left). This pattern may in part reflect the limited size of our recorded populations. For each of these metrics, simulated direct estimates better approximate the ground truth as population size grows (Figure 3A). However, the speed of improvement differs across metrics, implying that some will be more hampered by our recording conditions than others. We therefore complemented this analysis with one in which we attempted to get better estimates of each metric by inferring them from the observed population activity while taking into account knowledge of each unit's tuning properties (see Methods). In simulation, this procedure improves estimation accuracy for each metric (Figure 3A, full vs. dotted lines). Inferred response magnitude and gain variability each exhibited a modest association with Fisher information ($r = 0.55$ and 0.63, with 95% confidence intervals 0.35–0.71 and 0.45–0.77) while response dispersion and relative dispersion tended to exhibited a weaker association ($r = 0.36$ and 0.37, with 95% confidence intervals 0.12–0.57 and 0.13–0.56).

We next asked how well these metrics predict stimulus uncertainty on a trial-by-trial basis. For each population, we quantified this association by computing the rank correlation between the metric under consideration and the width of the likelihood function. We summarized results across populations by computing the mean rank correlation. This yielded values that were modest to high for inferred response magnitude ($r = 0.68$, 95% confidence interval of the mean 0.54–0.83), response dispersion ($r = 0.54$, 0.36–0.72), and gain variability ($r = 0.81$, 0.74–0.84), but not for relative dispersion ($r = 0.46$, 0.27–0.64; Figure 3C, left).

Perceptual uncertainty in part arises from internal sources (Faisal et al., 2008; Beck et al., 2005). To isolate these internal variations, we conducted a second experiment across eight recordings. In these sessions, a single, randomly chosen stimulus was overrepresented and shown several hundred times over the course of the experiment. As expected, this set of identical trials elicited considerable variability in the width of the likelihood function (standard deviation = 19.75° on average, 38.9% smaller than the standard deviation across all other trials in these experiments; see Methods). This variability must be largely due to internal sources. We found that the stimulus uncertainty

of these identical trials was predicted similarly well by all four inferred candidate metrics (mean rank correlation $r = 0.63$, 95% confidence interval 0.49–0.77 for response magnitude; 0.54, 0.30–0.77 for response dispersion; 0.69, 0.58–0.79 for gain variability; and 0.55, 0.25–0.82 for relative dispersion; Figure 3C, right).

In summary, our analysis suggests that response magnitude, response dispersion, and gain variability all provide a useful proxy for the fidelity of orientation coding in V1. This was true in both the presence (Figure 3C, left) and absence (Figure 3C, right) of external sources of stimulus variability. By the same standard, the combination of two candidate metrics in the form of relative dispersion did not result in a more predictive metric and thus not does not appear to be a useful proxy for stimulus uncertainty. These results do not appear to depend on the details of the decoding method. Specifically, decoding V1 activity with a conceptually simpler method that treats spikes as if they arise from a pure Poisson process yielded qualitatively similar results, though the association between the candidate metrics and likelihood width was overall weaker (Supplementary Figure S4).

Gain variability directly predicts stimulus uncertainty; the other metrics do not

We have identified three candidate metrics that exhibit modest to strong correlations with the width of the likelihood function. This implies that they may also correlate strongly with each other. Indeed, the average rank correlation between inferred response magnitude and response dispersion was 0.91; for response magnitude and gain variability, it was 0.86 (Figure 4A, left); and for response dispersion and gain variability, it was 0.68. We wondered how controlling for these confounds would impact the metrics' association with stimulus uncertainty. To this end, we rank ordered all trials from a given recording as a function of the metric we sought to control for. We then selected the first 50 trials. By design, these trials will not vary much along the sorting metric (i.e., the "frozen" variable), but they may vary along the two other "nonfrozen" metrics (Figure 4A, left). The correlation between the nonfrozen metrics and the width of the likelihood function provides a measure of the strength of the association in the absence of a confounding variable (Figure 4A, middle and right). We then repeated this analysis for the next sets of 50 trials and thus obtained a distribution of this measure (see Methods). Controlling for response magnitude had a modest impact on the predictive value of gain variability (mean rank correlation: $r = 0.53$ for Experiment 1 and 0.32, for Experiment 2, Figures 4B, C left). Conversely, freezing gain variability all but nullified the association between response magnitude and likelihood width. This was true of Experiment 1 ($r = 0.04$; Figure 4B, left) and Experiment 2 ($r = 0.11$; Figure 4C, left). We found a similar asymmetric pattern for gain variability and response dispersion. Freezing response dispersion did little to the predictive power of gain variability ($r = 0.63$ for Experiment 1 and 0.40 for Experiment 2, Figures 4B, C middle), but freezing gain variability removed most of the association between response dispersion and likelihood width ($r = 0.04$ for Experiment 1, difference with gain variability: $p < 0.001$; $r = 0.10$ for Experiment 2, $p < 0.001$; Figures 4B, C, middle). For all comparisons, the association between gain variability and stimulus uncertainty was greater when holding other metrics frozen than the association between other metrics when holding gain variability frozen ($p < 0.001$, Wilcoxon rank-sum test; Figures 4B, C, left and middle). This approach also showed that response magnitude was more associated with stimulus uncertainty than response dispersion (Figures 4B, C, right; $p < 0.001$, Wilcoxon rank-sum test). Overall, this pattern suggests that out of these three candidate metrics, gain variability has the most direct association with stimulus uncertainty. A complementary analysis that sought to examine how the correlation of each candidate metric with likelihood width depended on the intermetric correlation further corroborated this conclusion (Supplementary Figure S5).

Gain variability rivals performance of artificial neural networks

Our choice of candidate metrics was motivated by previous theoretical proposals (Ma et al., 2006; Jazayeri & Movshon, 2006; Olshausen & Field, 1996; Hoyer & Hyvärinen, 2003; Orbán et al., 2016; Hénaff et al., 2020). It is of course possible that there exists a hitherto unsuspected transformation of population activity that provides an even better indication of the decoder's uncertainty. To complement our analysis with a more agnostic approach, we trained a set of artificial neural networks (ANNs) to predict the width of the likelihood function from the population response. ANNs differed in architecture and training regime. To identify the subset of ANNs most relevant to our purposes, we conducted a two-round analysis. In the first round, we considered a diverse set of ANNs that differed in their architecture (specifically, the number of hidden layers and hidden units) and training regime (specifically, the dropout rate and weight decay; see Methods). For each population, we selected the combination of network architecture and training regime that yielded maximal predictive accuracy on held-out data (see Methods). We found that the best-performing combinations differed across the 13 populations. In the second round, we therefore again considered a set of ANNs, but this

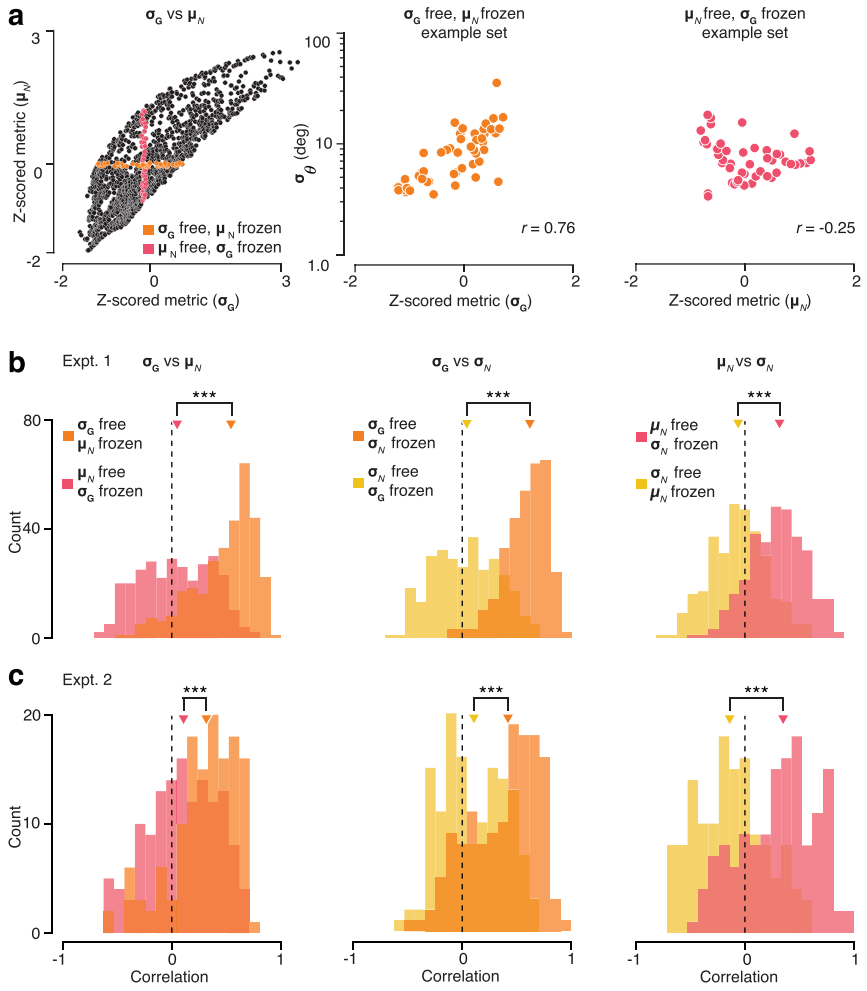

Figure 4. Comparison of partial correlations between three candidate metrics and the decoder's stimulus uncertainty. (**a**) Left, inferred response magnitude is plotted against inferred gain variability for an example population. Each point represents a single trial. Colored points indicate 50 consecutive rank-sorted trials. The two highlighted example sets of data points have identical variance in the nonfrozen metric ($\sigma^2 = 0.35$). Middle, relationship between gain variability and likelihood width for an example set of trials in which the response magnitude rate was kept constant (i.e., the orange trials in the left panel). Right, relationship between response magnitude and likelihood width for an example set of trials in which gain variability was kept constant (i.e., the pink trials in the left panel). (**b**) Distribution of partial correlations across all Experiment 1 recordings for inferred response magnitude and gain variability (left), inferred response dispersion and gain variability (middle), and inferred response dispersion and response magnitude (right). (**c**) Same as panel b for Experiment 2 trials with no variation in stimulus orientation, family, or noise seed. Triangles represent mean correlation; ***$p < 0.001$, Wilcoxon rank-sum test.

time the variation in network architecture and training regime was limited to the cases that had at least once resulted in a best performance in Round 1. We report the average performance across networks as figure of merit. As we did previously for the handpicked metrics, we first computed the correlation between the average predicted uncertainty for each stimulus family and population Fisher information. This association rivaled the best handpicked metric, inferred gain variability ($r = 0.61$ vs. 0.63, with 95% confidence intervals ranging from 0.59–0.62 and 0.45–0.77). We next examined how well the ANNs predicted likelihood width on a trial-by-trial basis. For Experiment 1, we found that likelihood width was on average not significantly better predicted by the family of ANNs than by inferred gain variability (mean $r = 0.79$ with 95% confidence intervals 0.75–0.83 for ANN predicted likelihood width, mean $r = 0.81$, 0.69–0.90 for inferred gain variability; Figure 5B, left). For Experiment 2, we found that likelihood width was somewhat better predicted by ANNs than by inferred gain variability (mean $r = 0.77$ with 95% confidence intervals 0.71–0.83 for ANN predicted likelihood width, mean $r = 0.69$, 0.58–0.81 for inferred gain variability; Figure 5B, right). We conclude that gain variability captures much of the variance of perceptual uncertainty that can be captured

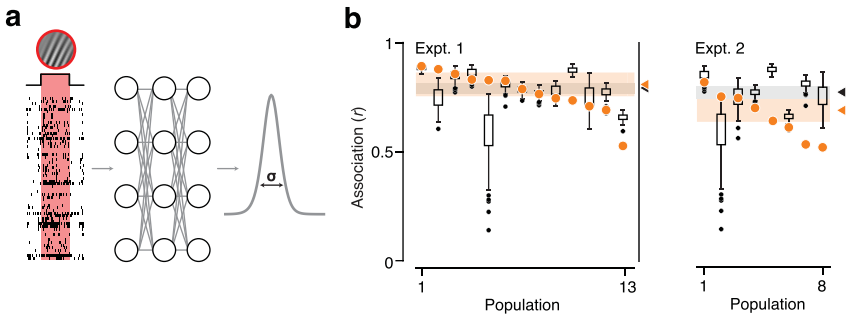

Figure 5. Artificial neural network prediction of likelihood width. (**a**) ANNs were trained on per-trial vectors of spike counts to predict the width of the likelihood function. (**b**) Boxplots indicate association of network predicted and ground-truth likelihood width per population across the set of networks for Experiment 1 (left) and Experiment 2 (right). Box indicates interquartile range. Solid black lines indicate the central 95% of the distribution. Black dots indicate outlier networks. Orange dots indicate inferred gain variability association with likelihood width. Gray and orange shaded regions indicate the 68% confidence interval around the mean association between network or metric and likelihood width. Triangles indicate mean association between network or metric and likelihood width.

by a simple transformation of sensory population activity.

Discussion

Uncertainty is intrinsic to perception. Inevitably, some perceptual interpretations of the environment are more uncertain than others. A prominent hypothesis holds that the same sensory neurons that inform perceptual interpretations represent the uncertainty of these interpretations (Zemel et al., 1998; Hoyer & Hyvärinen, 2003; Ma et al., 2006; Jazayeri & Movshon, 2006; Savin & Denève, 2014; Van Bergen et al., 2015; Orbán et al., 2016; Walker et al., 2020; Hénaff et al., 2020; Festa et al., 2021). The nature of this representation is debated. A major challenge in studying this is that perceptual uncertainty is a property of an observer's belief about the world, not a property of the world itself (Walker et al., 2022). Here, we sought to overcome this by manipulating stimulus reliability in two distinct ways and by using a model-based procedure to decode V1 spiking activity as a stand-in for what downstream areas *ought to* believe about stimulus orientation. We found that response magnitude, response dispersion, and variability in response gain all offer useful proxies for the certainty of stimulus orientation estimates. This was also true when fluctuations in uncertainty were not due to external factors but instead arose from internal sources. Of the metrics we considered, gain variability exhibited the most direct association with stimulus uncertainty.

Our findings offer empirical support for the central tenets of two different theoretical frameworks for the representation of uncertainty in sensory cortex: probabilistic population codes (PPCs) and the sampling hypothesis. The PPC framework is built on the idea that the nature of the neural code should enable implementing statistical inference with simple mathematical operations in a feedforward manner (Zemel et al., 1998; Ma et al., 2006; Beck et al., 2008). The modest to strong association of three simple transformations of population activity with an optimal decoder's uncertainty estimates validates this notion for the primary visual cortex. The core idea of sampling models, on the other hand, is that an aspect of response variability is used to represent uncertainty (Hoyer & Hyvärinen, 2003; Savin & Denève, 2014; Orbán et al., 2016; Festa et al., 2021). The finding that gain variability is the purest predictor of stimulus uncertainty further corroborates this hypothesis for the primary visual cortex (Orbán et al., 2016; Hénaff et al., 2020; Festa et al., 2021). Interestingly, recent theoretical work has shown that some coding regimes are consistent with both the PPC and sampling framework (Lange et al., 2023). It is possible that the primary visual cortex operates in exactly this mode (Lange et al., 2023).

From a statistical standpoint, the three metrics we deemed useful proxies for stimulus uncertainty capture very different aspects of neural population activity: the average of the neural response, the dispersion of these responses, and the variability of a latent modulator in a doubly stochastic process. Empirically, we found these metrics to be highly correlated with each other. Why might this be so? We think that response average and response dispersion are closely related because spike counts in visual cortex tend to be exponentially distributed (Baddeley et al., 1997). In principle, this distribution allows maximal information transmission per spike (Shannon, 1948; Barlow, 1961; Simoncelli & Olshausen, 2001). Under the exponential distribution, response standard

deviation is equal to the response mean, providing an explanation for the strong association between both statistics in our data. We further suggest that response average and gain variability are closely related because of the mechanistic origins of gain variability (Goris, Coen-Cagli, Miller, Priebe Lengyel, 2024). Specifically, we speculate that gain variability in large part arises from stimulus-independent noise in a divisive normalization signal. This normalization signal is thought to reflect aggregated nearby neural activity (Heeger, 1992; Carandini & Heeger, 2012). As shown previously, noisy normalization naturally results in gain variability being inversely proportional to the normalization signal (Coen-Cagli & Solomon, 2019; Hénaff et al., 2020). It follows that a higher mean response, leading to a stronger normalization signal, will generally coincide with a lower level of gain variability. If these interpretations are correct, then these strong associations might not just be specific to this V1 experiment but reflect a general property of visual cortex.

Regardless of its source, the strong association between the different statistics we studied implies that downstream circuits have multiple options to assess the reliability of the sensory messages conveyed by neural populations in visual cortex. The suitability of different proxies for uncertainty may differ across different brain areas or across different perceptual features and tasks. Perhaps a simple transformation is favored in cases where an approximate uncertainty estimate suffices, and more complex transformations are used when achieving a goal critically depends on the quality of the perceptual certainty estimate. More generally, we expect that the quality of perceptual certainty estimates will improve with experience, as is the case for simple perceptual decisions (Goldstone, 1998). Investigating these questions requires measuring sensory population activity from an animal generating a behavior that directly reflects their perceptual certainty on a trial-by-trial basis (Kepecs, Uchida, Zariwala & Mainen, 2008; Walker et al., 2020; Kiani & Shadlen, 2009).

The metrics we considered as candidate representations of uncertainty can be estimated directly from neural population activity without knowledge of the tuning properties of the neurons or of the selected stimulus interpretation. However, for these estimates to be reliable, relatively large populations are required. We therefore resorted to an estimation technique that leveraged knowledge of stimulus–response relations. While this approach is principled, verifying whether these results hold for direct estimation when the recorded population size is substantially larger is an important goal for future work.

Keywords: neural coding, visual cortex, sensory uncertainty, population representation

Acknowledgments

Supported by the US National Science Foundation (Graduate Research Fellowship to Z.M.B.-S.), the US National Institutes of Health (grant nos. T32 EY021462 and K99 EY032102 to C.M.Z. and EY032999 to R.L.T.G.), and the Whitehall Foundation (grant no. UTA19-000535 to R.L.T.G.). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Commercial relationships: none.
Corresponding author: Robbe L. T. Goris.
Email: robbe.goris@utexas.edu.
Address: Center for Perceptual Systems, University of Texas at Austin, Austin, TX 78712, USA.

References

Acerbi, L., & Ma, W. J. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Advances in Neural Information Processing Systems, 30*, 1834–1844.

Adams, D., Economides, J., Jocson, C., Parker, J., & Horton, J. (2011). A watertight acrylic-free titanium recording chamber for electrophysiology in behaving monkeys. *Journal of Neurophysiology, 106*(3), 1581–1590.

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupe, J.-M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience, 22*(19), 8633–8646.

Arandia-Romero, I., Tanabe, S., Drugowitsch, J., Kohn, A., & Moreno-Bote, R. (2016). Multiplicative and additive modulation of neuronal tuning with population activity affects encoded information. *Neuron, 89*, 1305–1316.

Baddeley, R., Abbott, L. F., Booth, M. C. A., Sengpiel, F., Freeman, T., Wakeman, E. A., . . . Rolls, E. T. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 264*(1389), 1775–1783.

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory Communication, 1*(1), 217–233.

Beaudot, W. H., & Mullen, K. T. (2006). Orientation discrimination in human vision: Psychophysics and modeling. *Vision Research, 46*(1–2), 26–46.

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., . . . Pouget, A.

(2008). Probabilistic population codes for Bayesian decision making. *Neuron, 60*(6), 1142–1152.

Beck, J.M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2005). Not noisy, just wrong: The role of suboptimal inference in behavioral variability. *Neuron, 74*(1), 30–39.

Berens, P., Ecker, A., Cotton, R., Ma, W., Bethge, M., & Tolias, A. (2012). A fast and simple population code for orientation in primate V1. *Journal of Neuroscience, 32*(31), 10618–10626.

Boundy-Singer, Z. M., Ziemba, C. M., & Goris, R. L. T. (2023). Confidence reflects a noisy decision reliability estimate. *Nature Human Behaviour, 7*(1), 142–154.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience, 13*(1), 51–62.

Charlton, J. A., Młynarski, W. F., Bai, Y. H., Hermundstad, A. M., & Goris, R. L. T. (2023). Environmental dynamics shape perceptual decision bias. *PLoS Computational Biology, 19*(6), e1011104.

Coen-Cagli, R., & Solomon, S. (2019). Relating divisive normalization to neuronal response variability. *Journal of Neuroscience, 39*(37), 7344–7356.

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience, 14*(7), 811–819.

Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Computational Biology, 4*(1), e27.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: The MIT Press.

Eastman, K. M., & Huk, A. C. (2012). PLDAPS: A hardware architecture and software toolbox for neurophysiology requiring complex visual stimuli and online behavioral control. *Frontiers in Neuroinformatics, 1*, 6.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9*(4), 292–303.

Festa, D., Aschner, A., Davila, A., Kohn, A., & Coen-Cagli, R. (2021). Neuronal variability reflects probabilistic inference tuned to natural image statistics. *Nature Communications, 12*(1), 3635.

Fetsch, C., Pouget, A., DeAngelis, G., & Angelaki, D. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience, 1*, 146–154.

Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. *Journal of Comparative Neurology, 201*(4), 519–539.

Geisler, W. S., & Albrecht, D. G. (1995). Bayesian analysis of identification performance in monkey visual cortex: Nonlinear mechanisms and stimulus certainty. *Vision Research, 35*(19), 2723–2730.

Goldstone, R. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612.

Goris, R., Simoncelli, E., & Movshon, J. (2015). Origin and function of tuning diversity in macaque visual cortex. *Neuron, 88*(4), 819–831.

Goris, R. L. T., Coen-Cagli, R., Miller, K. D., Priebe, N. J., & Lengyel, M. (2024). Response sub-additivity and variability quenching in visual cortex. *Nature Reviews Neuroscience, 25*(4), 237–252.

Goris, R. L. T., Movshon, J. A., & Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience, 17*(6), 858–865.

Goris, R. L. T., Ziemba, C. M., Movshon, J. A., & Simoncelli, E. P. (2018). Slow gain fluctuations limit benefits of temporal integration in visual cortex. *Journal of Vision, 18*(8), 1–13, https://doi.org/10.1167/18.8.8.

Goris, R. L. T., Ziemba, C. M., Stine, G. M., Simoncelli, E. P., & Movshon, J. A. (2017). Dissociation of choice formation and choice-correlated activity in macaque visual cortex. *Journal of Neuroscience, 37*(20), 5195–5203.

Graf, A. B. A., Kohn, A., Jazayeri, M., & Movshon, J. A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neuroscience, 14*(2), 239–245.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*(2), 181–197.

Hoyer, P., & Hyvarinen, A. (2003). Interpreting neural response variability as Monte Carlo sampling of the posterior. *Advances in Neural Information Processing Systems, 17*(1), 293–300.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of Physiology, 148*(3), 574–591.

Henaff, O. J., Boundy-Singer, Z. M., Meding, K., Ziemba, C. M., & Goris, R. L. T. (2020). Representation of visual uncertainty through neural gain variability. *Nature Communications, 11*(1), 2513.

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience, 9*(5), 690–696.

Kanitscheider, I., Coen-Cagli, R., Kohn, A., & Pouget, A. (2015a). Measuring Fisher information accurately in correlated neural populations. *PLoS Computational Biology, 11*(6), e1004218.

Kanitscheider, I., Coen-Cagli, R., & Pouget, A. (2015b). Origin of information-limiting noise correlations. *Proceedings of the National Academy of Sciences, 112*(50), E6973–E6982.

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature, 455*(7210), 227–231.

Kiani, R., & Shadlen, M. N. (2009). Representation of confidence associated with a decision by neurons in the parietal cortex. *Science, 324*(5928), 759–764.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.

Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in Neural Information Processing Systems,* 30.

Lange, R., Chattoraj, A., Beck, J. M., Yates, J., & Haefner, R. (2021). A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *PLoS Computational Biology, 17*(11), e1009517.

Lange, R. D., Shivkumar, S., Chattoraj, A., & Haefner, R. M. (2023). Bayesian encoding and decoding as distinct perspectives on neural coding. *Nature Neuroscience, 26*, 2063–2072.

Leventhal, A., Thompson, K., Liu, D., Zhou, Y., & Ault, S. (1995). Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *Journal of Neuroscience, 15*(3), 1808–1818.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience, 9*(11), 1432–1438.

Mareschal, I., & Shapley, R. M. (2004). Effects of contrast and size on orientation discrimination. *Vision Research, 44*(1), 57–67.

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience, 17*(10), 1410–1417.

Najemnik, J., & Geisler, W. (2005). Optimal eye movement strategies in visual search. *Nature, 434*, 387–391.

Nienborg, H., & Cumming, B. (2014). Decision-related activity in sensory neurons may depend on the columnar architecture of cerebral cortex. *Journal of Neuroscience, 34*(10), 3579–3585.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*(6583), 607–609.

Orban, G., Berkes, P., Fiser, J., & Lengyel, M. (2016). Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron, 92*(2), 530–543.

Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M., & Harris, K. D. (2016). Fast and accurate spike sorting of highchannel count probes with KiloSort. In *Advances in neural information processing systems* (Vol. *29*).

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*(5), 1.

Paradiso, M. A. (1988). A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. *Biological Cybernetics, 58*(1), 35–49.

Sahani, M., & Dayan, P. (2003). Doubly distributional population codes: Simultaneous representation of uncertainty and multiplicity. *Neural Computation, 15*(10), 2255–2279.

Salmasi, M., & Sahani, M. (2022). Learning neural codes for perceptual uncertainty. *IEEE International Symposium on Information Theory (ISIT),* 2463–2468.

Savin, C., & Deneve, S. (2014). Spatio-temporal representations of uncertainty in spiking neural networks. *Advances in Neural Information Processing Systems,* 27.

Seidemann, E., & Geisler, W. S. (2018). Linking v1 activity to behavior. *Annual Review of Vision Science, 4*, 287–310.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience, 18*(10), 3870–3896.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379–423.

Shooner, C., Hallum, L. E., Kumbhani, R., Ziemba, C., Garcia-Marin, V., Kelly, J., . . . Kiorpes, L. (2015). Population representation of visual information in areas V1 and V2 of amblyopic macaques. *Vision Research, 114*, 56–67.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*(1), 1192–1216.

Smith, M. A., Majaj, N. J., & Movshon, J. A. (2005). Dynamics of motion signaling by neurons in macaque area MT. *Nature Neuroscience, 8*(2), 220–228.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research, 23*(8), 775–785.

Van Bergen, R. S., Ma, W. J., Pratte, M. S., & Jehee, J. F. M. (2015). Sensory uncertainty decoded from visual cortex predicts behavior. *Nature Neuroscience, 18*(12), 1728–1730.

Walker, E. Y., Cotton, R. J., Ma, W. J., & Tolias, A. S. (2020). A neural basis of probabilistic computation in visual cortex. *Nature Neuroscience, 23*(1), 122–129.

Walker, E. Y., Pohl, S., Denison, R. N., Barack, D. L., Lee, J., Block, N., . . . Meyniel, F. (2022). Studying the neural representations of uncertainty. *arXiv preprint*, arXiv:2202.04324.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598–604.

West, R., Harrison, W., Matthews, N., Mattingley, J., Sewell, D., & Mathys, C. (2023). Modality independent or modality specific? Common computations underlie confidence judgements in visual and auditory decisions. *PLoS Computational Biology, 19*(7), e1011245.

Zemel, R., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation, 10*(2), 403–430.