

Transformations of sensory information in the brain reflect a changing definition of optimality

Tyler S. Manning^{1,*}, Emma Alexander², Bruce G. Cumming³, Gregory C. DeAngelis⁴, Xin Huang⁵, and Emily A. Cooper^{1,6}

¹Herbert Wertheim School of Optometry & Vision Science, University of California, Berkeley

²Department of Computer Science, Northwestern University

³Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health

⁴Department of Brain and Cognitive Sciences, University of Rochester

⁵Department of Neuroscience, University of Wisconsin, Madison

⁶Helen Wills Neuroscience Institute, University of California, Berkeley

Neurons throughout the brain modulate their firing rate in response to changes in sensory input. Theories of neural computation posit that these modulations reflect the outcome of a constrained optimization: neurons aim to efficiently and robustly represent sensory information under resource limitations. However, evidence for how this optimization varies across the brain is lacking. Here, we show that sensory representations transform along the dorsal visual stream in a manner consistent with a transition from information preservation to perceptual discrimination. Using binocular disparity as a test bed, we re-analyze measurements from over 1500 neurons characterizing tuning curves in macaque monkey brain regions V1, V2, and MT, and compare these to measurements of the natural visual statistics of binocular disparity (the slight differences in how objects project to two laterally-offset eyes). The changes in tuning curve characteristics at the population level are computationally consistent with a shift in optimization goals from maximizing the information encoded about naturally occurring binocular disparities to maximizing the ability to support fine perceptual discrimination. We find that a change towards larger individual neuron disparity preferences is a key driver of this shift. As a result, we provide new insight into previously-identified differences between disparity-selective regions of cortex and suggest these differences play an important role in supporting visually-guided behavior. Our findings support a fundamental change in how we define “efficient coding” in regions of the brain that contain sensory information, emphasizing the need to consider not just information preservation and neural resources, but also relevance to behavior.

efficient coding | representation | information | binocular disparity

Significance

A major role of the brain is to transform information from the sensory end organs into signals that can be used to guide behavior. From first principles, however, it is unclear how exactly the brain has developed to accomplish this task in a way that simultaneously minimizes how much energy the brain consumes, maximizes information preservation, and maximizes an observer’s ability to utilize the broad range of behaviorally-relevant information in the environment. In this report, we suggest that the brain shifts from a goal of being an optimal conduit of sensory information to a goal of optimally representing sensory information for perceptual discrimination between classically-defined regions of the early and middle regions of the visual system.

An appealing theory of neural computation is that early sensory areas in the brain represent information in a way that maximizes the information carried about stimuli in the world (1, 2) while downstream regions transform this representation so as to best support specific tasks and computations (3–6). A careful test of this theory of sensory transformations requires several elements: we need to characterize the typical probability distribution of a pertinent sensory variable encountered in the environment, we need large-scale measurements of neural responses driven by this sensory variable across multiple brain areas, and we need a single mathematical framework that can be applied to sensory representations shaped by different behavioral or computational objectives. Here, we use binocular disparity in the primate visual system as a model, and combine across diverse datasets to test this theory. Our results provide strong empirical evidence in support of a systematic transformation of sensory representations in the brain: from information preservation at early processing stages to maximizing perceptual discrimination performance at later stages.

Binocular disparity between the retinal images offers an ideal test bed for examining hierarchical sensory representations. In animals with forward-facing eyes, non-fixated points in space tend to fall on disparate retinotopic locations because the eyes are laterally offset from each other (Fig. 1A). Successful integration of information from the two eyes relies on populations of neurons that are tuned for different horizontal binocular disparities—the differences in the retinotopic location of images in the left and right eyes. While neuronal tuning for binocular disparity emerges early in the mammalian visual system (V1), populations of neurons tuned for binocular disparity have also been characterized all along the dorsal and ventral processing streams (7, 8). Beyond just binocular integration, sensing of binocular disparity provides one of the most reliable cues to the relative distances of objects in the environment, and as such this cue supports a variety of high-level perceptual tasks such as figure/ground segmentation, 3D motion perception, and breaking camouflage (9–11). Indeed, the magnitude and direction of binocular disparity varies lawfully as a function of how far objects are from the observer as well as where the observer is fixating, and prior work has shown that these variations result in predictable statistical regularities in the binocular disparities encountered during natural tasks (12–15). We hypothesized that early representations of binocular disparity maximize the in-

formation carried about typical disparities encountered during natural behavior, while later representations instead facilitate discrimination of disparity to support specific perceptual tasks.

Results

Natural distributions of binocular disparity have strong statistical regularities. To test this sensory transformation hypothesis, we first need an understanding of the distribution of binocular disparities that the visual system is tasked with processing. In recent years, there has been a concerted effort to characterize the visual “diet” of binocular disparities that is typical of natural experience (12–15). This work suggests several robust statistical properties of typical binocular disparities, most notably that small disparities (near zero) tend to be much more likely than larger disparities. Optimal sensory representations for binocular integration should therefore differentially allocate processing resources for small disparities versus large ones. Using a previously-collected dataset in which people were eyetracked while performing natural tasks (12), we calculated the probability distribution of binocular disparities in the central 20 degrees of the visual field while people performed two different tasks: food preparation and navigation (Fig. 1B). While the precise distribution shape differed between the tasks, consistent with previous analyses we see that both distributions are approximately zero-mean, symmetric, and highly kurtotic.

Sensory coding theories predict a lawful transition in how these disparity statistics are reflected in the brain. Optimal coding theories provide concrete, testable predictions for how a stimulus probability distribution should be reflected in neural populations. Specifically, if a population maximizes the information carried about a sensory variable, we expect the Fisher Information of the population neural activity (FI ; a measure of the precision with which a sensory variable is encoded) follow a power law in which it is proportionate to the probability squared (Fig. 1C, “infomax” dashed line)(16, 17). This representation can be interpreted as a reference prior (18), which assumes the least possible information about the world, or Jeffrey’s prior, providing invariance across transformation of sensory units (19).

In a standard setting, the $FI \propto p^2$ power law corresponds to a neural code that minimizes the L_0 norm of the stimulus reconstruction error. This means that all error magnitudes are penalized equally and the expected value of these errors is minimized (i.e., error sparsity is maximized; Fig. 1D: top left & right; (5, 6). Prior work has found signatures of this power law across a range of early sensory brain areas (3, 20). However, a representation with this distribution is not optimal for visual tasks that require discrimination between different values of a sensory variable, because error magnitude matters for many tasks. For example, if one is trying to determine whether their hand can fit in between two sharp objects

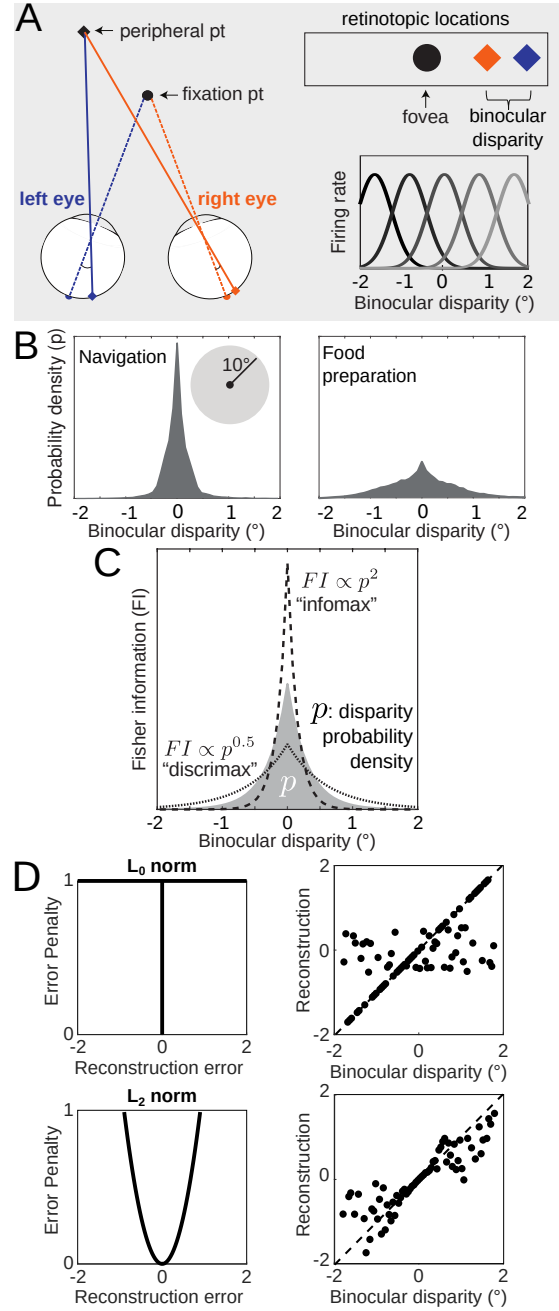


Fig. 1. A. (Top right) Points in the peripheral visual field tend to fall on disparate retinotopic locations because the eyes are laterally offset. The retinotopic difference in these locations is called the horizontal disparity. (Bottom right) Populations of neurons that are tuned for different horizontal binocular disparities are found throughout the visual system. **B.** Binocular disparities encountered in the central visual field (10° of fixation) tend to be small. Each plot shows the probability density distribution of binocular disparity obtained from data collected by (12) while human participants either navigated an indoor environment (Left) or prepared a sandwich (Right). **C.** Information theoretic frameworks indicate that optimal sensory representations can be described by a power law relationship between stimulus probability p (shaded region) and the Fisher Information (FI) of a neuronal population. Information maximizing codes (“infomax”, dashed line) are proportionate to the squared probability, while discrimination maximizing codes (“discrimax”, dotted line) are characterized by a compressive nonlinearity. **D.** (upper and lower left) Error functions that minimize the L_0 and L_2 norms, respectively. (upper and lower right) Example disparity estimates under each of the respective error functions that reconstruct the ground truth from noisy sensory measurements. For clarity, disparities are sampled from a uniform distribution. Error functions yield the same amount of total reconstruction error. Dashed line indicates the ground truth reconstruction.

separated in depth, a small error may be okay but a large error in the wrong direction may lead to injury. Thus, for perceptual discrimination, neural codes that minimize some other error metric like mean squared error are often appropriate (L_2 norm; Fig. 1D: bottom left & right). Minimizing these error metrics lowers the value of the exponent n in the $FI \propto p^n$ power law (6, 21). As such, codes that are optimized for downstream discrimination tasks should reduce the concentration of neural processing resources on high probability events and spread FI more equally across the stimulus space (Fig. 1C, “discrimax” dotted line)(5, 17). For neural codes for binocular disparity, we would therefore expect the population FI to be more strongly peaked at zero disparity in early visual regions, and more equally spread out across a broader range of binocular disparities in later visual regions.

Neural populations differ as predicted by a transition from information-preservation to supporting perceptual discriminations. To further test this transformation hypothesis, we must characterize of a large number of neuronal tuning curves for binocular disparity across different brain areas, such that we can calculate the FI associated with these tunings. Since the precise shape of the disparity probability distribution varies between tasks and resource constraints can change the numerical value of the optimal exponent n (5, 21), here we focus on the *relative* transformation of the FI exponent between brain regions rather than on its nominal value. To this end, we compiled a dataset of 1665 neurons’ binocular disparity responses spanning brain areas V1, V2, and MT of the macaque monkey. To ensure that the receptive fields (RFs) of the included neurons were well matched to each other and to the disparity image set, we subsampled this initial set of neurons (see Methods) and obtained a final set of 1056 neurons. The mean responses of each neuron as a function of binocular disparity were fit with a continuous 1D Gabor function and the individual neuron FI associated with each tuning curve was calculated from these fits based on assumption of Poisson spiking (Fig. 2A).

Intuitively, the FI is high when the tuning curve is steep and the spike rate is low, and it is low when the tuning curve is flat and the spike rate is high. The total FI of each population was calculated as a sum across neurons, based on the assumption that each neuron responds to stimuli independently (Fig. 2B). Qualitatively, we see that the population FI is most kurtotic in V1 and least kurtotic in MT, consistent with the hypothesis that the information-maximizing model is a better description of the early visual representation (V1 and V2) and the discrimination-maximizing model is a better description of the downstream representation (MT).

We next directly compared empirical FI distributions to the two probability distributions of binocular disparities in the natural environment. Binocular disparities are distributed non-uniformly across the visual field during natural behavior (12), so we started by resampling the natural disparity distributions based on the specific retinotopic locations of neuronal

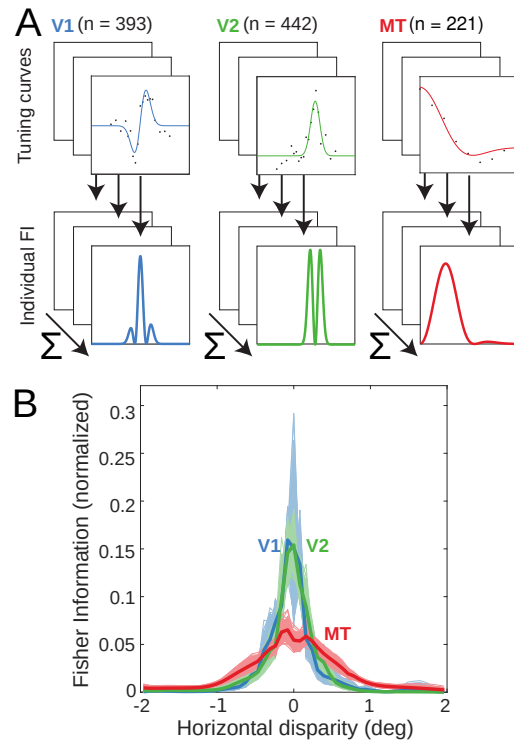


Fig. 2. A. We compiled a dataset of 1056 disparity tuning curves from brain areas V1, V2, and MT of the macaque monkey. The mean responses of each neuron as a function of binocular disparity were fit with a continuous function and the individual neuron FI associated with each tuning curve was calculated from these fits. **B.** The population FI is shown for each brain area (thick lines). Thin lines represent the population FI computed from 500 bootstrapped samples from each brain area.

receptive fields in each population using kernel-smoothed density estimates (see Fig. S1). The resulting disparity distributions were all similar in shape (Supp. Fig. S2A&B), despite the minor differences in sampling density between the different brain regions.

We then calculated the power law that, when applied to the corresponding disparity distributions, resulted in the best fit to the measured FI of each neuronal population. We used bootstrapping of the populations to estimate variability of the best-fit power law. Consistent with our working hypothesis, we observed a systematic decrease in the best fit power law from the lower-level areas V1 and V2 to mid-level area MT for both natural tasks (Fig. 3A&C). To quantify these differences, we fit the bootstrapped power law values for each neuronal population with a Gaussian distribution and measured the effect sizes between populations. The effect sizes, measured as Cohen’s D between pairs of populations, were large between the earlier areas and MT (food preparation: V1 vs. MT = 3.4, V2 vs. MT = 2.9; navigation: V1 vs. MT = 5.7, V2 vs. MT = 7.2). As expected from the similarity in their FI distributions, the effect sizes were small between V1 and V2. (food preparation: 0.15, navigation: V1 vs. V2 = 0.10). Fig. 3B&D show the close match between the population FI and the disparity distributions scaled by the optimal power law values.

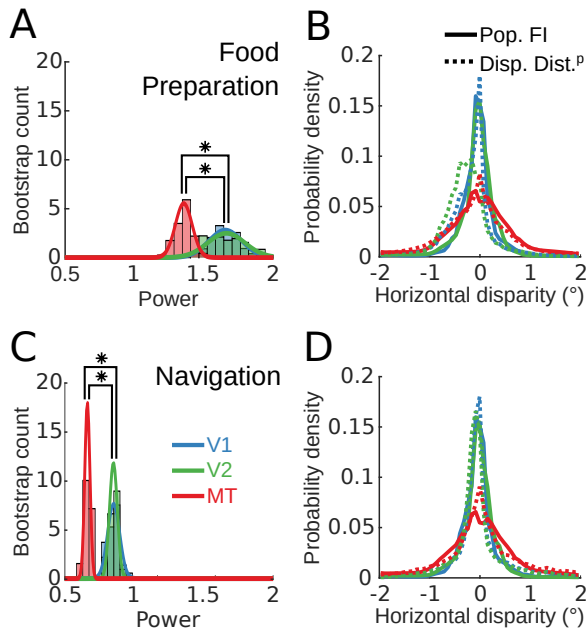


Fig. 3. A. The distribution of best-fit power laws linking population *FI* to the disparity probability densities collected from the food preparation image set (using the kernel-smoothed probability densities to guide sampling) is plotted for the 500 bootstrapped samples for each brain area as bars, along with Gaussian distribution fits to each set of samples. Asterisks indicate significant differences in median power. **B.** Population *FI* plotted from each area overlapped with the disparity distributions in A. scaled by their respective best fit power laws. **C & D.** As in **A & B.**, but for disparity statistics collected during the navigation behavioral task.

These differences correspond to a broad set of changes in individual tuning curve characteristics from V1/V2 to MT. We next asked which aspects of the neural responses to disparity could account for the differences in the *FI* distributions between V1/V2 and MT. To answer this question, we first took a parametric approach and leveraged the Gabor fits to each of the tuning curves (Fig. 4A). For each of the cortical populations, we examined the distributions of each of the six best-fit Gabor parameters (Fig. 4B). In sum, we find that MT neurons generally have higher response offsets, broader envelopes, and lower disparity frequencies than either V1 or V2. The MT population also has a broader distribution of best fit envelope means and phases of the cosine component than the earlier cortical areas. Full results of the statistical comparisons across populations are provided in Supp. Tables S1 & S2. This analysis expands on a previous comparison between V1 and MT (22) by including responses from V2 and a larger number of V1 neurons. One possible explanation for these differences is that they may reflect differences in the retinoptic locations of the receptive fields across the samples from each brain areas. Our subsampling from the initial larger dataset resulted a good match between the brain areas in terms of eccentricity and vertical position within the visual field, although the MT dataset is more concentrated in the left visual field and the V1/V2 datasets are more concentrated in the right visual field (Fig. 5). Since there is no reason to hypothesize that disparity tuning should differ in terms of left or right visual field, these tuning differences most likely reflect differences in the underlying neural representation.

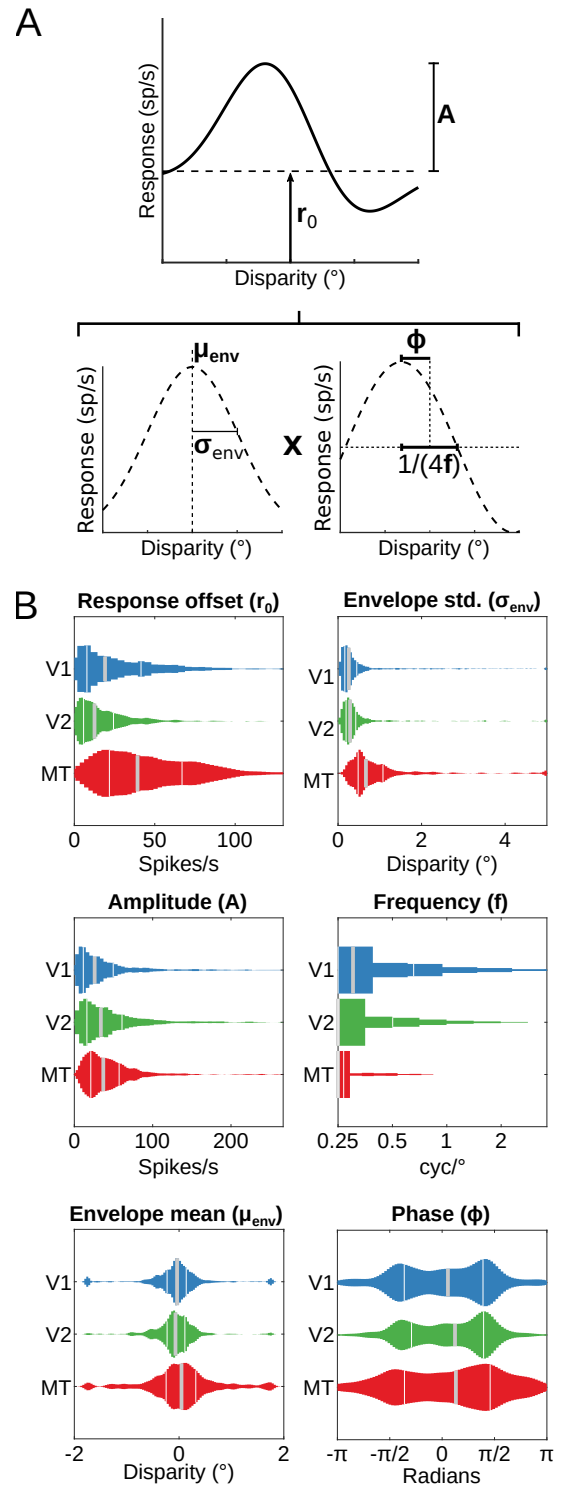


Fig. 4. We fit all the neuronal data with a modified 6 parameter Gabor function and plotted the distribution of best-fitting parameters in each of the cortical areas from which we have data. **A.** Decomposing the Gabor function into Gaussian and cosine components clarifies what each parameter contributes to the shape of the resulting tuning curve. **B.** Distributions of best fitting Gabor parameters for each of the 3 cortical areas. Thin white bars indicate the 25th and 75th quartile and the thick gray bar indicates the median.

An increase in neurons preferring larger disparities is a key factor in the observer coding transformation. There are clear differences in the distribution of best fit Gabor parameter distributions and the preferred disparities between

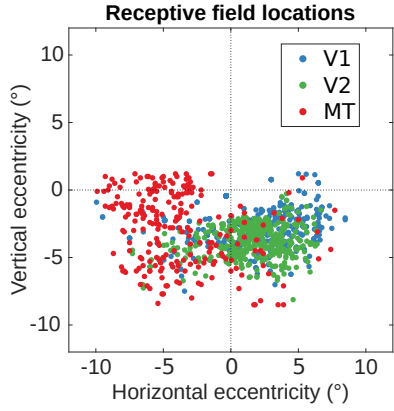


Fig. 5. The distributions of retinotopic eccentricities for the receptive fields in each brain area differed. Each circle represents a single neuron's receptive field. We used the sampling locations to calculate a unique disparity probability distribution associated with these sample locations. The locations here represent the 1056 cells that passed the inclusion criteria (see methods).

the earlier cortical regions and MT. However, the complex and overlapping effects that these parameters have on the tuning curve shape make these differences hard to interpret how these parameter differences contribute to the transformation of the population *FI* distributions between the regions. Therefore, we performed a resampling analysis to 1) see if the changes in any one parameter in particular could account for the difference in the *FI* distributions, and 2) whether swapping the V1 and MT distributions for a parameter was responsible for the changes or just simply resampling the parameters (from within V1) was sufficient. Since the *FI* distributions from V1 and V2 were similar, we did not repeat this analysis for the difference between V2 and MT. The overall approach is outlined in Fig. 6A: for a given tuning curve parameter, we replaced the true V1 fits with ones randomly sampled from the distribution of MT fits. We then summed the individual cell *FI* distributions and normalized by the area under the curve to compare the overall shapes of the resampled distribution and the empirical MT distribution. This process was repeated 500 times for each Gabor parameters to assess the variability in the resulting resampled *FI* distributions. The resulting V1 *FI* distributions for each of the parameters are shown alongside the true MT *FI* distribution in Fig. 6B. Of the six parameters, replacing the V1 envelope mean parameter with those from MT qualitatively results in the closest match with the lowest variability.

To examine the significance of these matches, we calculated the Jensen-Shannon divergence (JSD) between the *FI* of each of the resampled V1 populations and the true MT population *FI* distributions. This information theoretic measure reflects the dissimilarity between two distributions. We first tested whether there were significant differences between the resampled populations for each parameter using a Kruskal-Wallis test and confirmed there were differences in dissimilarity between the parameters ($\chi^2 = 1.65E3$; $df=5$; $p<2.22E-16$). While there were significant differences between each of the sets of resampled populations (see Table S3), the μ_{env} parameter stood out as the least divergent from the empiri-

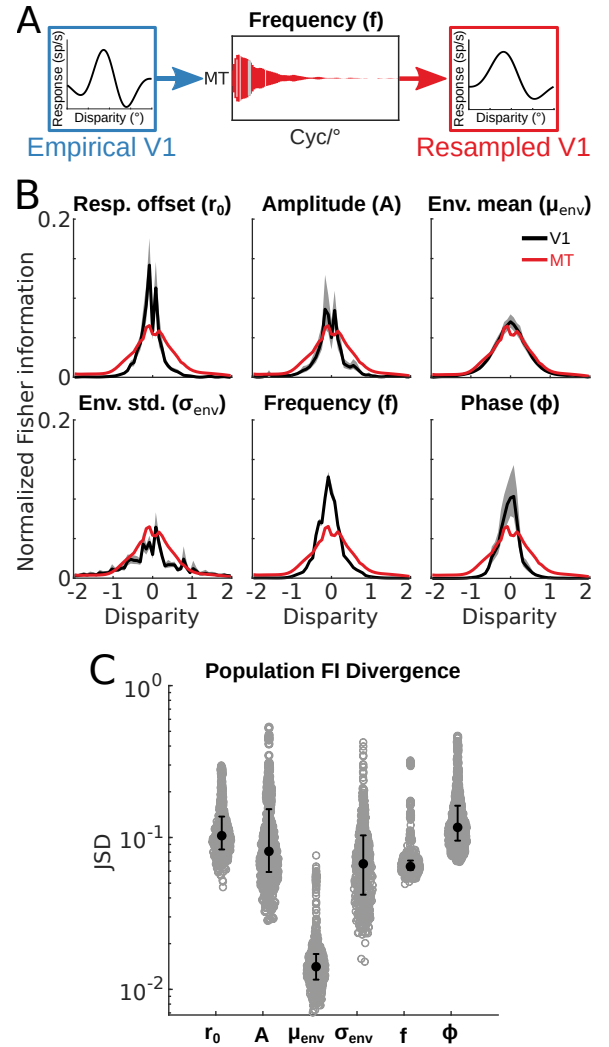


Fig. 6. **A.** We investigated which of the 6 Gabor parameters best explained the difference in the population *FI* between V1 and MT by individually replacing one parameter at a time from each cell in a sample with the best-fitting parameters from the MT cell sample. We repeated this process to create 500 bootstrapped samples. **B.** Mean and interquartile range from the resampled V1 samples (black) compared to the true MT population *FI* (red). **C.** Jensen-Shannon Divergence (JSD) between the resampled V1 population *FI* distributions and the true MT population *FI* distribution. Lower values reflect a closer match between the two distributions. (orange) JSD when parameters were sampled from the MT parameter distributions and (black) JSD when parameters were sampled from the V1 parameter distributions (i.e. shuffled). Asterisks indicate significant differences between the two resampled datasets.

cal MT distribution. We also wanted to ensure that our focus on population *FI* shape instead of also considering the total magnitude of *FI* content for each bootstrap, we also performed the same analysis without normalizing each bootstrap by the AUC and solely by the number of cells in the bootstrapped sample. The results of this analysis are shown in Fig. S4A&B and S5: overall, they again show greater similarity between the resampled μ_{env} parameter populations than any of the other resampled populations.

To ensure that the results of our resampling approach were due to sampling from the MT parameter distributions themselves and not due to resampling in general, we also repeated our analysis by resampling from the same data sample (i.e.

only other V1 neurons). We established a significant difference between the resampled populations pulling from the V1 and from the MT distributions by performing a Wilcoxon rank-sum test between the two for each Gabor parameter. The results of these tests are shown in Table S4: the divergence between the resampled V1 and empirical MT distributions were only significantly minimized for the μ_{env} and f parameters.

From this pattern of results, we can glean that much of the difference in how disparity information is represented between earlier and later cortical areas can be attributed the presence of a greater number of cells that are selective for larger disparities in MT (or have broader disparity tuning curves).

Discussion

Here we have taken a theoretical prediction about population-level neuronal information along the sensory processing hierarchy and put it to an empirical test. The results are consistent with the prediction that sensory transformations can be understood as the result of a constrained optimization, in which the goal changes sensibly from early to late sensory regions.

While we focused on the power law relationship between stimulus probability, encoding goals, and population FI , the optimal population FI can be influenced by other factors as well. For example, the optimal power law is affected by the resource constraints (21). We assumed that visual brain regions in the same species are subject to the same constraints. Adaptation studies that dynamically shift stimulus statistics without affecting constraints may be able to determine whether or not this assumption holds (23). Similarly, system noise influences the form of FI , both for individual neurons and the population as a whole (see (24)). A noise analysis of our data suggest that, at least in the current dataset, spike count properties were similar across the examined brain areas (see Fig S3). A systematic characterization of neuronal noise properties along the sensory processing hierarchy may ultimately reveal that sensory transformations are subject to differing noise properties as well (25, 26).

Here, we used binocular integration as a model system for sensory transformations more generally. But our results also shed light on specific open questions in the neural underpinning of binocular integration. Previous work has noted that the shape of disparity tuning curves appears to change systematically across brain regions, but the reason for these changes is unknown (22, 27, 28). For example, previous work suggested that MT tuning curves tend to have odd symmetry and broader tunings, whereas in V1, the best fit tuning curves are more even symmetric with a narrower range of preferred disparities (22). Here, we did not see evidence for a difference in even/odd symmetric, but did observe a multifaceted set of differences in the distributions of tuning curve shapes. Some of the difference in tuning shape (e.g. width)

may be a side effect of pooling neurons with different orientation preferences to generate direction selectivity for patterns (29). Our information theoretic analysis suggests these differences in tuning curve shape may also have direct utility: they shift the position of a neuron's peak FI to larger disparities while maintaining a population peak near zero. Over the entire population, this effectively makes the population FI distribution more broad, thus improving disparity discrimination at higher disparity pedestals.

Of note, we do not claim that a shift from “infomax” to “discrimax” representations is the sole difference in how V1, V2, and MT represent visual information. For example, it is well-established that MT encodes binocular disparity in a way more correlated with perception than V1 does: V1 neurons invert their disparity tunings with anticorrelated stereomatches (7) and are affected by vergence and absolute disparity (30, 31), whereas V2 (32, 33) and MT (34) discard false stereo matches and encode relative disparity. It is also important to consider other sensory variables represented in these areas. A recent study used an information theoretic framework to examine the link between the encoding of speed in MT and perceptual biases in speed estimation (35). They found that the population FI for stimulus speed in MT can be related to speed perception via a $p^2(s)$ power law. Our estimates of the power law in MT for binocular disparity domain conflict with this “infomax” representation of speed. However, it is not necessary that every sensory variable encoded within a population is represented at the same level—it is entirely possible for the same brain region to contain an information-maximizing representation of one sensory variable and a discrimination-maximizing representation of another. The information-theoretic framework provides an additional window into how neural representations build and interact along sensory processing streams that can complement other assessments of neural function.

Lastly, our work also highlights the fact that the statistics of sensory input can be task dependent. This task-dependence can pose a problem for assessing the encoding optimality of neural populations on the basis of task-free natural stimulus statistics derived from generic datasets, for example, of natural images and sounds (3, 4). Here, we show that generalizations that are robust across tasks can be made by focusing on relative differences between brain regions. However, every model necessarily relies on a suite of simplifying assumptions, and a notable one here is the assumption that neuronal representations can be well-characterized with tuning curves.

With the information theoretic analyses like those presented here, we gain a more principled understanding of the link between the hierarchy of cortical areas carrying sensory information and the complexity of behaviors that rely on that information. New computational frameworks that can be applied to dynamic populations of neurons, trial-by-trial variations, and spike-count correlations between neurons will contribute to the next step in characterizing hierarchical transformations of sensory signals.

Methods

Definition of Binocular Disparity. We define horizontal binocular disparity (d) as follows:

$$d = \beta_L - \beta_R, \quad (1)$$

where β_L and β_R denote the horizontal eccentricity of an image projected to the left and right retinas, relative to the fovea. We represent eccentricity and binocular disparity in units of visual degrees, with negative disparity values indicating points that are closer than the fixation point and positive values indicating points that are farther than the fixation point.

Natural Scene Statistics of Binocular Disparity. Natural statistics of binocular disparity were re-analyzed from a previous collected dataset (12). In brief, three adult human subjects performed either indoor navigation or food preparation tasks while wearing a custom-designed headset consisting of a pair of stereocameras and eye trackers. Still images were sampled from the stereocamera video footage, transformed into head-centered 3D scene geometry, and finally converted into disparity maps in retina-centered coordinates using the eye-tracking data. Analyses were limited to data within a 10° radius of the point of fixation.

While the original natural disparity statistics measurements were evenly sampled from a 10° radius of fixation, the neural datasets contained neurons with receptive fields that were restricted largely to one hemifield. Given the differences in disparity statistics between the upper and lower hemifields and the expected increase in prevalence of larger disparities with increasing eccentricity (12), we resampled disparities from the maps based on the receptive field centers of each neuron. To do so, we calculated 2D kernel-smoothed probability density distributions (see Fig. S1) for each of the cortical areas using a 2D Gaussian kernel and used these distributions to subsample from the original distribution. To ensure the reliability of this subsampling procedure, we used a bootstrapping procedure to estimate the variability of the resulting distribution (errorbars are smaller than the line width in Fig. 3A&D). In short, we sampled with replacement from each of the image sets (navigation and food preparation) to generate 100 bootstrapped samples of equal size to the empirical distribution, calculated the disparity probability distribution for each bootstrap, and recovered the 95% CIs from the cumulative error distributions.

Neural Recordings. Neural recordings from areas V1, V2, and MT were re-analyzed from a combination of multiple previous studies. All recordings come from awake fixating macaque monkeys and reflect measured action potentials of isolated neurons. The tuning curves measured in V1 and V2 come from multiple studies using highly similar methods measured over the course of a decade in the same laboratory

(36–46). The MT tuning curves were obtained in a single study with slightly different experimental methods (22).

In all cases, the stimuli used to measure response rate (spikes per second) as a function of binocular disparity were random dot stereograms (RDS) with each dot subtending 0.1° degrees. The methods describing the stimuli and data collection methods for the V1 and V2 data are described in (37, 40). Briefly, responses from V1 and V2 were collected using RDS stimuli with no coherent motion, presented on a Wheatstone mirror haploscope for 400–500ms at a display refresh rates ranging from 72–100Hz. Responses from MT were collected using RDS stimuli with 100% coherent motion tailored to each cell's preferred direction, speed, and size (22). For each of the MT recording sessions, stereoscopic stimuli were presented for 1500ms on a single monitor with liquid crystal shutter glasses at a refresh rate of 50Hz for each eye (cross-talk was measured to be $<3\%$). The stimuli in the MT recording sessions were also presented against a non-overlapping background of stationary dots at zero disparity to anchor vergence (this vergence lock was located outside of the neuronal receptive fields in all cases).

Recordings from each area were likewise similar. In brief, after the receptive field of each neuron was localized, the stimulus was adjusted to the optimal size and velocity and the responses to different magnitudes of binocular disparity were recorded (using disparities ranging from -1.6° to 1.6° in steps ranging from 0.04° – 0.4°). Responses from MT and V2 neurons were collected solely with single electrodes, while responses from V1 were collected mostly from single electrodes and a few from multicontact probes (45). The receptive fields of neurons in V1 and V2 were largely restricted to the lower visual field due to cortical topography and the positioning of the implanted recording cylinders. In the MT dataset, all recording cylinders were placed above the right hemisphere, so the receptive fields are limited to the left visual field. Horizontal and vertical eccentricity of the stimuli and receptive field positions were calculated by taking the arctangent of the distance on the screen from the fixation point and the viewing distance.

There were slight differences in experimental protocols between the laboratories (i.e., shutter glasses vs. mirror haploscope, stimulus presentation time, coherent motion vs. incoherent motion), but these differences are unlikely to cause the differences in the FI distributions estimated from the data. Motion and disparity are independently encoded in MT (47), so it is unlikely that the difference in motion energy between the stimuli used by the two laboratories would produce consistent biases in disparity tuning. Most neurons in V1 and MT also do not show a dependence of disparity selectivity on interocular delay, instead showing an inverse relationship between response gain and interocular delay; those that do show disparity-delay inseparability do not exhibit large tuning shifts over the display intervals used in either of the datasets (40, 48).

Tuning Curve Analysis. A subset of the neural recordings described in the previous section were selected for analysis according to set a of inclusion criteria. First, we only included neurons with an average of ≥ 3 repeats per stimulus disparity. We then selected neurons for which a significant amount of the trial-by-trial variance in responses was explained by the disparity of the stimulus (one-way ANOVA at a significance level of $p < 0.01$). For the V1 and V2 recordings, the range of stimulus disparities presented was variable, so to ensure sufficient data to obtain a reliable fit to the tuning curves we only analyzed neurons with a range of at least 1.5 deg between the nearest and the farthest disparity tested. These criteria resulted in 690, 531, and 444 neurons from V1, V2, and MT, respectively.

In each study, stimulus disparity was recorded in screen coordinates rather than retinal coordinates. For example, a point with the same horizontal coordinate on screen for both eyes was coded as having zero disparity. However, in retinal coordinates, the locus of points with zero disparity is a circle that contains the fixation point and the optical center of the two eyes. Therefore, a point with the horizontal coordinates on a planar screen will have an uncrossed (far) retinal disparity. In order to match up the neural recordings to the retinal disparities measured in the scene statistics analysis, we therefore applied a correction factor. In brief, the retinal eccentricity of each on screen stimulus in the left and right eye was determined based on the screen distance, the horizontal screen coordinates in the left and right eye, and an assumed interocular separation of 30mm. These retinal eccentricities were used to calculate the angular disparity on the retinas, which were used in the subsequent analyses.

For each of these neurons, we then fit a continuous tuning curve to the mean responses as a function of the stimulus disparity $h(d)$. Tuning curves were parameterized as with a 6-parameter Gabor function:

$$h(d) = r_0 + A \exp \left[\frac{(d - \mu_{env})^2}{2\sigma_{env}^2} \right] \cos(2\pi f(d - \mu_{env}) + \phi). \quad (2)$$

Best-fitting parameters for each neuron were determined using constrained non-linear optimization in MATLAB (Mathworks, Inc.) minimizing the mean squared error (fmincon). To prevent fits that deviated substantially from the observed range of spike rates, data were up-sampled by a factor of 2 using linear interpolation prior to fitting. Bounds for parameters were as follows: $0 < r_0 < 500$, $0 < A < 500$, $-1.75 < \mu_{env} < 1.75$, $0 < \sigma_{env} < 5$, $0 < f < 4.5$, $-2\pi < \phi < 2\pi$. Fits that resulted in a minimum spike rate of less than 0.05 spikes per second or with a frequency (f) of less than 0.25 were strongly penalized, to avoid instability in the calculation of Fisher Information and the interpretation of the fitted parameters, respectively. The optimization was initialized at 200 randomly selected starting points and optimized according to an interior point algorithm. The parameters with the lowest error across all initializations were taken as the final fit. A subset of neurons (less than 30) were identified with

poor fits on manual inspection, so fitting routines were re-run for these neurons with minor adjustments to the parameter ranges. The R^2 across all fits was greater than 0.3, and the median was above .8 for all three areas.

We further subsampled the neuronal populations to ensure that each included cell was well-fit ($R^2 \geq 0.75$) with a Gabor tuning function, had RF centers within the defined region of the disparity image set (i.e. $\leq 10^\circ$ eccentricity), and had RF centers within the same general subregion of the visual field. To enforce the final criterion, we determined the largest vertical and horizontal components of the RF centers from the V1/V2 datasets to define a rectangular bounding box and then selected only MT cells with RF centers within this region. This was enforced because the MT dataset contained a larger number of cells with RF centers in the upper visual field, which could potentially bias the disparity preferences of the sample (12). The final cell counts from each sample were 393, 442, and 221 from V1, V2, and MT, respectively.

To investigate differences between the resulting parameter distributions, we first performed an Kruskal-Wallis omnibus test to ask if there were any significant differences in median values between the three cortical regions. Since we were primarily interested in differences in distribution breadth, we first took the absolute value of the signed parameters (μ_{env} and ϕ) to test only difference in the median magnitudes of these values. If the omnibus test revealed a significant difference between the distributions, we then followed up with a set of Wilcoxon rank sum tests between each of the pairs of cortical areas. The results of these tests are presented in Tables S1 and S2.

Fisher Information. Based on the simplifying assumption that the spike rates for each neuron are Poisson-distributed, we calculated the Fisher Information (FI_n) of each neuron as described by (49):

$$FI_n(d) = \frac{h'^2(d)}{h(d)}, \quad (3)$$

where $h'(d)$ denotes the first derivative of the tuning curve. Negative FI_n values, resulting from fits with small negative spike rates, were set to zero. Assuming that each neuron's spike rate is independent, the FI in the population level is then:

$$FI = \sum_n FI_n(d) \quad (4)$$

Thus, we summed together the FI_n across the neurons in each neuronal population.

To examine the variance of the population FI , we repeated this analysis on subsamples of 400 neurons (sampled with replacement) 500 times for each region. The sample size of 400 was selected so that we could match sample sizes across all three populations.

To ensure that our neuronal datasets contained disparity responses that were approximately Poisson and did not sub-

stantially vary between areas, we plotted the relationship between mean spike count and spike count variability for each of the cell populations (Fig. S3, top). Note that each point is a unique presented disparity condition and each cell contributes multiple points. Only conditions with ≥ 5 repeated stimulus presentation were included. We then fit a simple power law function to each of the distributions:

$$\sigma^2(r_{mean}) = ar_{mean}^b. \quad (5)$$

Poisson-distributed responses should be best fit with both a , or *slope*, and b , or *power law*, equal to 1. While the best fit slope for each area is slightly greater than 1, the best fit power laws did not substantially differ from 1 (Fig. S3, bottom). Since the populations also did not greatly differ in their best fit values, our assumption of Poisson spiking statistics for the calculation of population *FI* seems well-justified.

Gabor Parameter Resampling Between Cortical Areas.

To identify what aspect of disparity tuning could best explain the differences in the *FI* distributions between populations, we conducted a resampling approach whereby we individually replaced the best-fitting Gabor parameters (r_0 , A , μ_{env} , σ_{env} , f , ϕ) from one population with another (diagrammed in Fig. 6A). Since the *FI* distributions from V1 and V2 did not significantly differ, we restricted the analysis to V1 and MT. For each parameter, we first discarded the best-fit set from the V1 population. We then randomly sampled from a kernel-smoothed probability density derived from the set of best-fits to MT neurons and assigned new values to the parameter of interest for the V1 population. As done previously, we then calculated the single cell *FI* distributions given this new set of tuning curves and summed their values to get the population *FI*. We then repeated this process 500 times to obtain confidence intervals on the population Fisher information under the reparameterization. This process was then repeated for each of the 5 other Gabor parameters, with the median *FI* distributions and interquartile range for the resampled V1 populations shown in Fig. 6B. To quantitatively determine which resampled parameter produced a V1 *FI* distribution that was closest to the empirical MT distribution, we computed the Jensen-Shannon divergence (JSD) between each of the bootstrapped V1 *FI* distributions and the empirical MT distribution. To test whether the shuffling with MT parameters specifically was responsible for a decrease in divergence (rather than shuffling in general), we also performed a control set of bootstraps in which we simply resampled from the V1 parameter distributions instead. Significant differences in JSD between these two sets were determined with a Wilcoxon rank sum test for each parameter.

Comparison Between Fisher Information and Disparity Statistics. We used a grid search to determine the power law that minimized the difference between each population *FI* and the sampled binocular disparity probability distributions. Power laws were applied to the disparity distributions and then differences were calculated as the mean absolute error

between the two distributions sampled at 51 evenly spaced binocular disparities between -2° and 2° . We repeated this minimization for each of the 500 bootstrapped samples from each neuronal population and fit the resulting distributions with a Gaussian distribution using maximum likelihood estimation, which also allowed us to characterize the 95% confidence interval of the average power law for each neuronal population.

Code and Data Availability

The MATLAB code used to pre-process the neuronal data, perform the neuronal analyses, and collect the image statistics is available at <https://github.com/tsmanning/DisparityInfoProject>. The neuronal data and BORIS datasets are available at <https://zenodo.org/doi/10.5281/zenodo.7495820>.

Acknowledgements

We thank the National Eye Institute (grants F32 EY032321 and T32 EY007043) for support.

Bibliography

1. Attneave, F. (1954). Some informational aspects of visual perception. *Psychological review* **61**, 183.
2. Barlow, H. B. (1961) in *Sensory Communication*, ed. Rosenblith, W. A. (The MIT Press), pp. 216–234.
3. Simoncelli, E. P. & Olshausen, B. A. (2001). Natural image statistics and neural representation *Annual review of neuroscience* **24**, 1193–1216.
4. Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes *Annual Review of Psychology* **59**, 167–192.
5. Ganguli, D. & Simoncelli, E. P. (2014). Efficient sensory encoding and bayesian inference with heterogeneous neural populations *Neural Computation* **26**, 2103–2134.
6. Wang, Z., Stocker, A. A., & Lee, D. D. (2016). Efficient neural codes that minimize l p reconstruction error *Neural computation* **28**, 2656–2686.
7. Cumming, B. G. & DeAngelis, G. C. (2001). The physiology of stereopsis *Annual review of neuroscience* **24**, 203–238.
8. Gonzalez, F. & Perez, R. (1998). Neural mechanisms underlying stereoscopic vision *Progress in neurobiology* **55**, 191–224.
9. Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex *Perception & Psychophysics* **55**, 48–121.
10. Regan, D. & Beverley, K. I. (1979). Binocular and monocular stimuli for motion in depth: Changing-disparity and changing-size feed the same motion-in-depth stage *Vision research* **19**, 1331–1342.
11. Adams, W. J., Graf, E. W., & Anderson, M. (2019). Disruptive coloration and binocular disparity: breaking camouflage *Proceedings of the Royal Society B* **286**, 20182045.
12. Sprague, W. W., Cooper, E. A., Tošić, I., & Banks, M. S. (2015). Stereopsis is adaptive for the natural environment *Science advances* **1**, e1400254.
13. Canessa, A., Gibaldi, A., Chessa, M., Fato, M., Solari, F., & Sabatini, S. P. (2017). A dataset of stereoscopic images and ground-truth disparity mimicking human fixations in peripersonal space *Scientific data* **4**, 1–16.
14. Burge, J., McCann, B. C., & Geisler, W. S. (2016). Estimating 3d tilt from local image cues in natural scenes *Journal of Vision* **16**, 2–2.
15. Adams, W. J., Elder, J. H., Graf, E. W., Leyland, J., Lutigheid, A. J., & Murry, A. (2016). The southampton-york natural scenes (syns) dataset: Statistics of surface attitude *Scientific reports* **6**, 1–17.
16. Brunel, N. & Nadal, J.-P. (1998). Mutual information, fisher information, and population coding *Neural computation* **10**, 1731–1757.
17. Wei, X.-X. & Stocker, A. A. (2016). Mutual Information, Fisher Information, and Efficient Coding *Neural Computation* **28**, 305–326.
18. Berger, J. O., Bernardo, J. M., & Sun, D. (2009). The formal definition of reference priors *The Annals of Statistics* **37**, 905–938.
19. Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems *Proc. R. Soc. Lond. A* **186**, 453–461.
20. Ganguli, D. & Simoncelli, E. P. (2016). Neural and perceptual signatures of efficient sensory coding *arXiv:1603.00058 [q-bio]*.

21. Morais, M & Pillow, J. W. (2018) *Power-law efficient neural codes provide general link between perceptual bias and discriminability* eds. Bengio, S, Wallach, H, Larochelle, H, Grauman, K, Cesa-Bianchi, N, & Garnett, R. (Curran Associates, Inc.), Vol. 31, pp. 5071–5080.
22. DeAngelis, G. C & Uka, T. (2003). Coding of horizontal disparity and velocity by mt neurons in the alert macaque *Journal of neurophysiology* **89**, 1094–1111.
23. Rast, L. & Drugowitsch, J. (2020) *Adaptation Properties Allow Identification of Optimized Neural Codes* eds. Larochelle, H, Ranzato, M, Hadsell, R, Balcan, M. F, & Lin, H. (Curran Associates, Inc.), Vol. 33, pp. 1142–1152.
24. Kohn, A, Coen-Cagli, R, Kanitscheider, I, & Pouget, A. (2016). Correlations and Neuronal Population Information *Annu Rev Neurosci* **39**, 237–56.
25. Charles, A. S, Park, M, Weller, J. P, Horwitz, G. D, & Pillow, J. W. (2018). Dethroning the Fano Factor: A Flexible, Model-Based Approach to Partitioning Neural Variability *Neural Computation* **30**, 1012–1045.
26. Goris, R. L. T, Movshon, J. A, & Simoncelli, E. P. (2014). Partitioning neuronal variability *Nat Neurosci* **17**, 858–865. Number: 6 Publisher: Nature Publishing Group.
27. Tanabe, S, Doi, T, Umeda, K, & Fujita, I. (2005). Disparity-Tuning Characteristics of Neuronal Responses to Dynamic Random-Dot Stereograms in Macaque Visual Area V4 *Journal of Neurophysiology* **94**, 2683–2699. Publisher: American Physiological Society.
28. Yoshioka, T. W, Doi, T, Abdolrahmani, M, & Fujita, I. (2021). Specialized contributions of mid-tier stages of dorsal and ventral pathways to stereoscopic processing in macaque *eLife* **10**, e58749. Publisher: eLife Sciences Publications, Ltd.
29. Prince, S. J, Cumming, B. G, & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1 *J Neurophysiol* **87**, 209–21.
30. Cumming, B. G & Parker, A. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity *J Neurosci* **19**, 5602–18.
31. Cumming, B. G & Parker, A. J. (2000). Local Disparity Not Perceived Depth Is Signaled by Binocular Neurons in Cortical Area V1 of the Macaque *J. Neurosci.* **20**, 4758–4767. Publisher: Society for Neuroscience Section: ARTICLE.
32. Thomas, O. M, Cumming, B. G, & Parker, A. J. (2002). A specialization for relative disparity in V2 *Nat Neurosci* **5**, 472–478.
33. Bakin, J. S, Nakayama, K, & Gilbert, C. D. (2000). Visual Responses in Monkey Areas V1 and V2 to Three-Dimensional Surface Configurations *J. Neurosci.* **20**, 8188–8198. Publisher: Society for Neuroscience Section: ARTICLE.
34. Krug, K & Parker, A. J. (2011). Neurons in Dorsal Visual Area V5/MT Signal Relative Disparity *J. Neurosci.* **31**, 17892–17904. Publisher: Society for Neuroscience Section: Articles.
35. Zhang, L.-Q & Stocker, A. A. (2022). Prior Expectations in Visual Speed Perception Predict Encoding Characteristics of Neurons in Area MT *Journal of Neuroscience* **42**, 2951–2962.
36. Cumming, B. G. (2002). An unexpected specialization for horizontal disparity in primate primary visual cortex *Nature* **418**, 633–636.
37. Read, J. C. A & Cumming, B. G. (2003). Measuring V1 Receptive Fields Despite Eye Movements in Awake Monkeys *Journal of Neurophysiology* **90**, 946–960.
38. Read, J. C & Cumming, B. G. (2004). Understanding the Cortical Specialization for Horizontal Disparity *Neural Computation* **16**, 1983–2020.
39. Nienborg, H, Bridge, H, Parker, A. J, & Cumming, B. G. (2004). Receptive Field Size in V1 Neurons Limits Acuity for Perceiving Disparity Modulation *Journal of Neuroscience* **24**, 2065–2076.
40. Read, J. C. A & Cumming, B. G. (2005). Effect of Interocular Delay on Disparity-Selective V1 Neurons: Relationship to Stereoacuity and the Pulfrich Effect *Journal of Neurophysiology* **94**, 1541–1553.
41. Nienborg, H, Bridge, H, Parker, A. J, & Cumming, B. G. (2005). Neuronal Computation of Disparity in V1 Limits Temporal Resolution for Detecting Disparity Modulation *Journal of Neuroscience* **25**, 10207–10219.
42. Nienborg, H & Cumming, B. G. (2006). Macaque V2 Neurons, But Not V1 Neurons, Show Choice-Related Activity *Journal of Neuroscience* **26**, 9567–9578.
43. Tanabe, S & Cumming, B. G. (2008). Mechanisms Underlying the Transformation of Disparity Signals from V1 to V2 in the Macaque *Journal of Neuroscience* **28**, 11304–11314.
44. Tanabe, S, Haefner, R. M, & Cumming, B. G. (2011). Suppressive Mechanisms in Monkey V1 Help to Solve the Stereo Correspondence Problem *Journal of Neuroscience* **31**, 8295–8305.
45. Henriksen, S, Read, J. C. A, & Cumming, B. G. (2016). Neurons in Striate Cortex Signal Disparity in Half-Matched Random-Dot Stereograms *Journal of Neuroscience* **36**, 8967–8976.
46. Clery, S, Cumming, B. G, & Nienborg, H. (2017). Decision-Related Activity in Macaque V2 for Fine Disparity Discrimination Is Not Compatible with Optimal Linear Readout *Journal of Neuroscience* **37**, 715–725.
47. Smolyanskaya, A, Ruff, D. A, & Born, R. T. (2013). Joint tuning for direction of motion and binocular disparity in macaque MT is largely separable *J Neurophysiol* **110**, 2806–16.
48. Pack, C. C, Born, R. T, & Livingstone, M. S. (2003). Two-Dimensional Substructure of Stereo and Motion Interactions in Macaque Visual Cortex *Neuron* **37**, 525–535. Publisher: Elsevier.
49. Seung, H. S & Sompolinsky, H. (1993). Simple models for reading neuronal population codes *Proceedings of the national academy of sciences* **90**, 10749–10753.

Supplementary Information

Kernel-smoothed density distributions

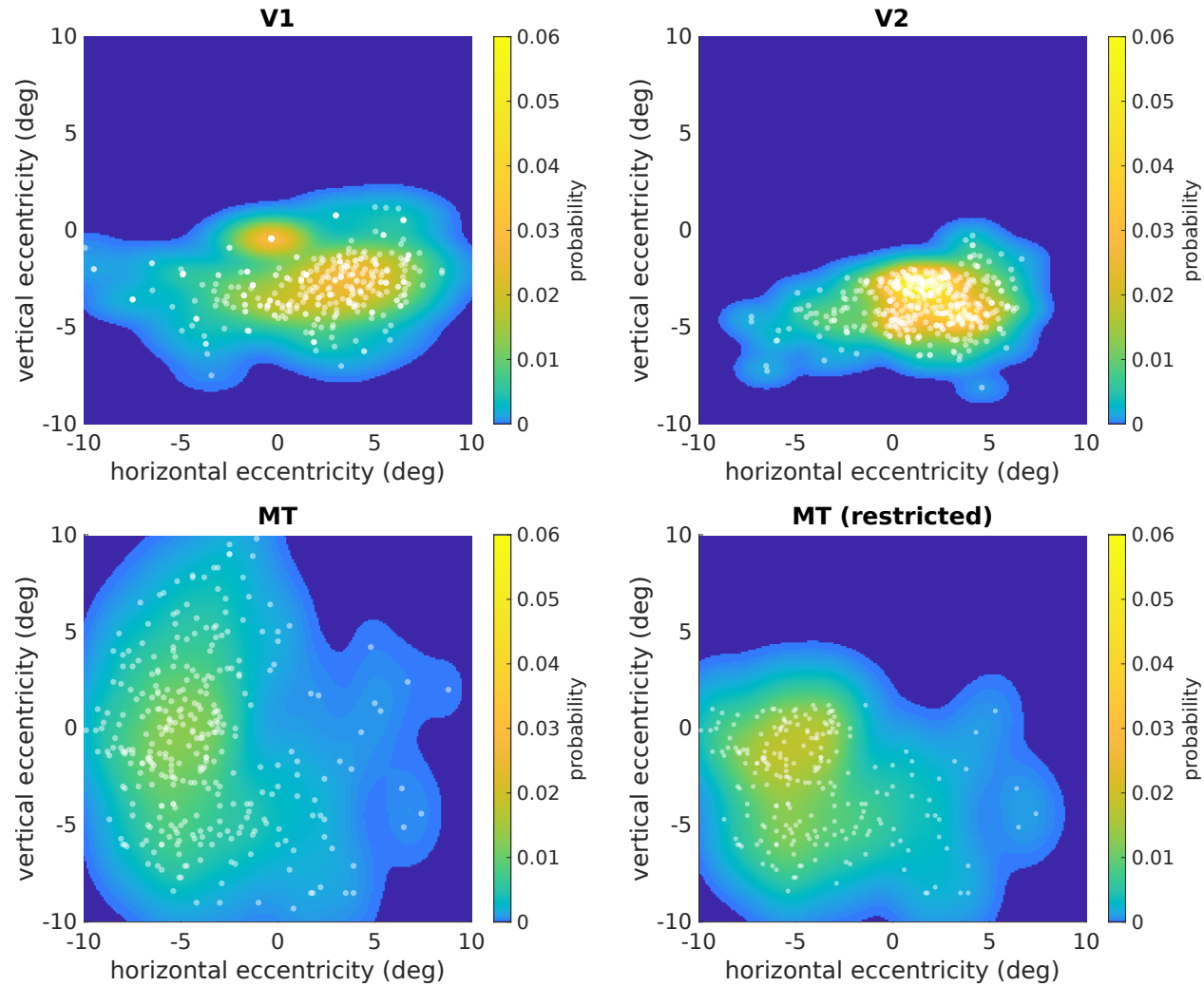


Fig. S1. Kernel-smoothed probability densities (KSDs) from each of the neuronal data sets, as well as the KSD recovered from the MT subsample restricted to cells with receptive field centers within the bounds of the maximal extents of the V1/V2 receptive fields. Each white dot is the center of one cell's receptive field.

Disparity probability densities sampled with KSDs

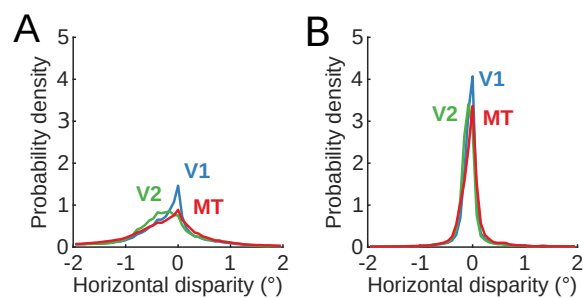


Fig. S2. **A.** The disparity probability densities collected from the food preparation image set using the kernel-smoothed probability densities to guide sampling. The mean and 95% confidence intervals across 100 bootstrapped samples are shown, but the errorbars are too small to be visible. **B.** As in **A.**, but for the navigation image set

Spike count statistics between cortical areas

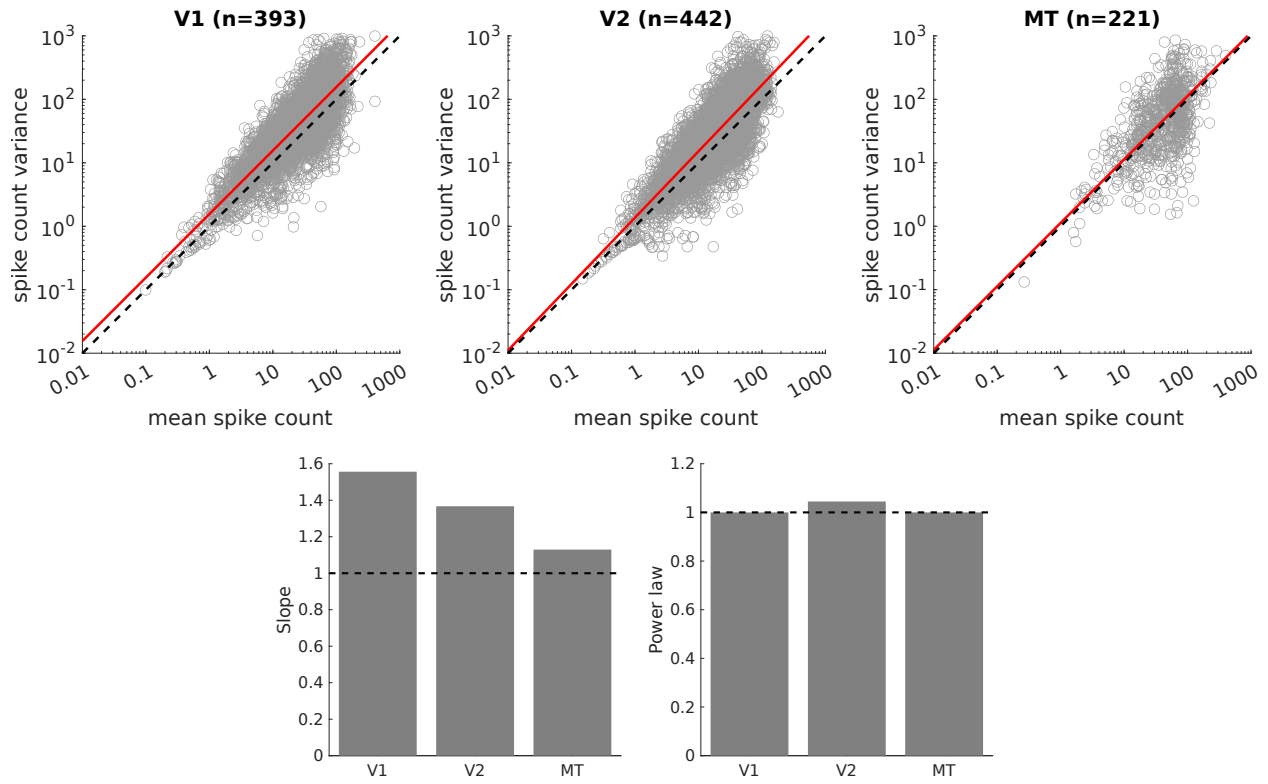


Fig. S3. (Top) Relationship between spike count mean and variance for each unique combination of stimulus disparity and cell in each of the three neuronal datasets. The cell sample numbers in the titles reflect the number of cells that contributed to each plot, but each cell contributes more than one datapoint. Dashed line: unity. Red line: best fit power law function to the distribution. **(Bottom)** The best fit slopes and power laws from each of the red lines in the top set of plots. Dashed line: the expected values of a set of Poisson-distributed spike counts.

FI distribution divergence using cell count normalization

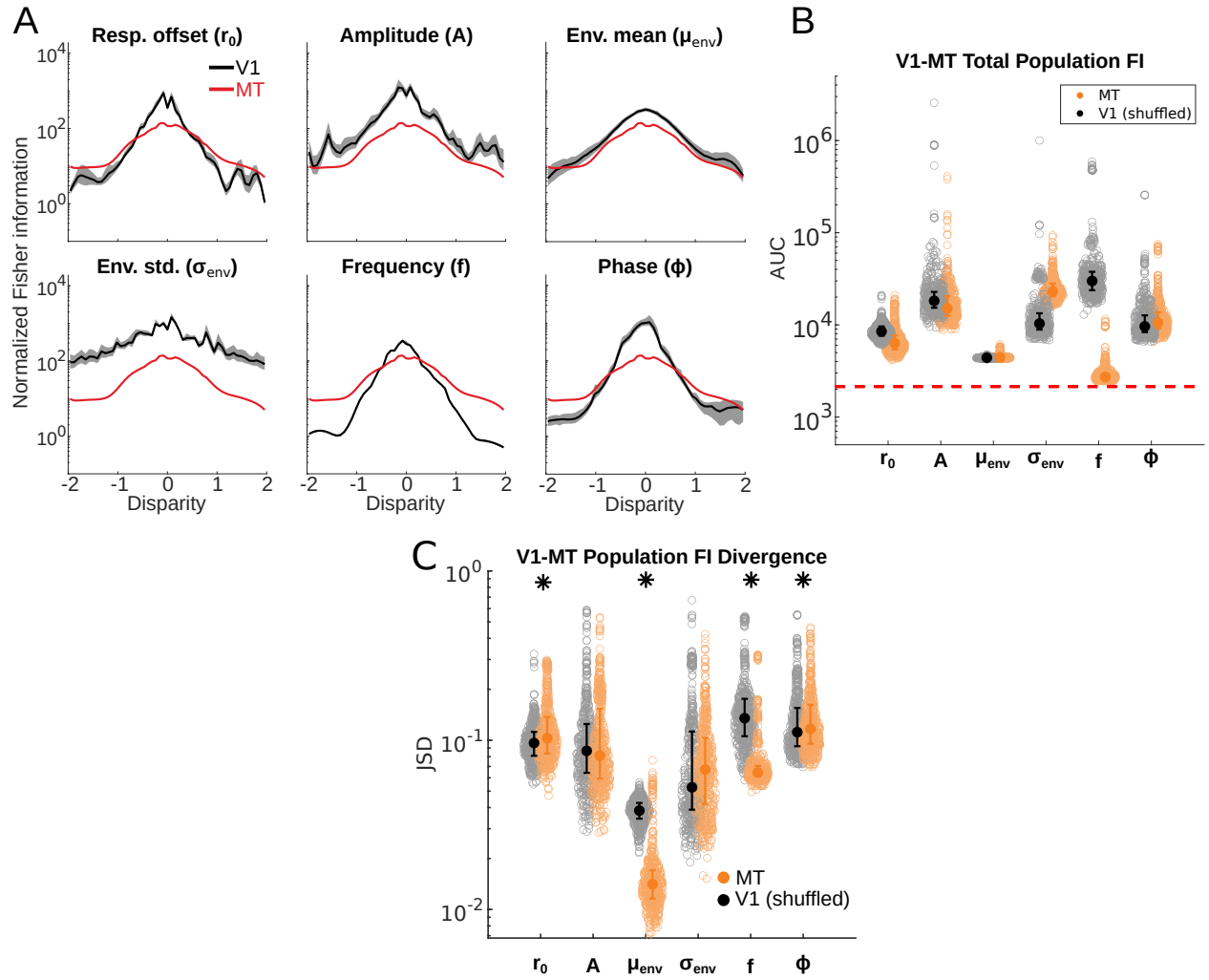


Fig. S4. **A.** Mean and interquartile range from the resampled V1 samples (solid black) compared to the true MT population FI (solid red). **B.** Area under curve (AUC) reflecting total population Fisher information for the resampled V1 populations and the empirical MT distribution (dashed red line). (orange) AUC when parameters were sampled from the MT parameter distributions and (black) AUC when parameters were sampled from the V1 parameter distributions (i.e. shuffled).

Neuronal response statistics

parameter	statistics
r_0	$\chi^2 = 177; df=2; \mathbf{p=4.28E-39}$
A	$\chi^2 = 0.325; df=2; p=0.850$
μ_{env}	$\chi^2 = 11.0; df=2; \mathbf{p=2.95E-6}$
σ_{env}	$\chi^2 = 251; df=2; \mathbf{p=3.02E-55}$
f	$\chi^2 = 47.8; df=2; \mathbf{p=4.07E-11}$
ϕ	$\chi^2 = 21.2; df=2; \mathbf{p=2.45E-05}$
RF center Ecc.	$\chi^2 = 96.9; df=2; \mathbf{p=8.92E-22}$
Pref. Disp.	$\chi^2 = 5.16; df=2; p=0.0760$

Table S1. Test statistics and significance for Kruskal-Wallis one-way test for differences between distribution medians.

parameter	V1-V2	V1-MT	V2-MT
r_0	$z=7.93; \mathbf{p=1.28E-15}$	$z=-5.98; \mathbf{p=2.22E-9}$	$z=-13.0; \mathbf{p=1.18E-38}$
A	N/A	N/A	N/A
μ_{env}	$z=-1.27; p=0.202$	$z=-4.80; \mathbf{p=1.57E-6}$	$z=-4.14; \mathbf{p=3.42E-5}$
σ_{env}	$z=-4.83; \mathbf{p=1.34E-6}$	$z=-14.8; \mathbf{p=1.57E-49}$	$z=-13.0; \mathbf{p=1.27E-38}$
f	$z=2.94; \mathbf{p=0.00325}$	$z=6.63; \mathbf{p=3.43E-11}$	$z=4.98; \mathbf{p=6.48E-7}$
ϕ	$z=2.48; \mathbf{p=0.0132}$	$z=-2.44; \mathbf{p=0.0148}$	$z=-4.53; \mathbf{p=5.82E-6}$
RF center Ecc.	$z=-1.87; p=0.0617$	$z=-9.03; \mathbf{p=1.66E-19}$	$z=-8.64; \mathbf{p=5.40E-18}$
Pref. Disp.	N/A	N/A	N/A

Table S2. Test statistics and significance for two-tailed Wilcoxon rank sum test for differences between distribution medians. Pairs of distributions tested are indicated in the top row.

Resampled V1 neuronal statistics

	r_0	A	μ_{env}	σ_{env}	f	ϕ
r_0	-	$z=5.87$; p=4.26E-9	$z=27.2$; p=2.85E-163	$z=14.1$; p=3.60E-45	$z=19.0$; p=4.12E-80	$z=-7.80$; p=6.40E-15
A	-	-	$z=26.5$; p=5.01E-155	$z=8.94$; p=4.06E-19	$z=7.37$; p=1.67E-13	$z=-10.4$; p=1.67E-25
μ_{env}	-	-	-	$z=-25.3$; p=1.66E-141	$z=-26.7$; p=1.71E-157	$z=-27.3$; p=1.74E-164
σ_{env}	-	-	-	-	$z=-3.63$; p=2.89E-4	$z=-17.7$; p=3.70E-70
f	-	-	-	-	-	$z=-21.8$; p=2.88E-105
ϕ	-	-	-	-	-	-

Table S3. Test statistics and significance for two-tailed Wilcoxon rank sum test for differences between bootstrapped V1 populations resampled from MT parameter distributions for each resampled parameter. Pairs of distributions tested are indicated in the top row.

parameter	MT vs. V1 resampling
r_0	$z = 3.70$; p=2.12E-4
A	$z = -1.35$; p=0.177
μ_{env}	$z = -22.1$; p=5.01E-108
σ_{env}	$z = 0.374$; p=0.708
f	$z = -21.1$; p=1.84E-98
ϕ	$z = 2.85$; p=4.37E-3

Table S4. Test statistics and significance for two-tailed Wilcoxon rank sum test for median Jensen-Shannon Divergence (JSD) differences between bootstrapped V1 populations resampled from V1 and MT parameter distributions for each resampled parameter

parameter	MT vs. V1 resampling
r_0	$z = -14.95$; p=1.56E-50
A	$z = -9.72$; p=2.60E-22
μ_{env}	$z = 5.03$; p=4.98E-7
σ_{env}	$z = 19.8$; p=2.59E-87
f	$z = -27.4$; p=5.86E-165
ϕ	$z = 5.36$; p=8.42E-8

Table S5. Test statistics and significance for two-tailed Wilcoxon rank sum test for median total Fisher Information (TFI) differences between bootstrapped V1 populations resampled from V1 and MT parameter distributions for each resampled parameter