

# Stereopsis-related wiring in the primate visual cortex in unnatural visual environments

Sarah Marzen<sup>1,2</sup>, Joel Zylberberg<sup>1,2</sup>, Michael Robert DeWeese<sup>1,2,3,\*</sup>

**1 Department of Physics, University of California, Berkeley, CA, USA**

**2 Redwood Center for Theoretical Neuroscience, University of California, Berkeley, CA, USA**

**3 Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA**

**\* E-mail: deweese@berkeley.edu**

## Abstract

Primates use binocular disparity, the slight differences between the images received by their right and left eyes, to reconstruct their 3-D environment. As a result, binocular disparity statistics of a natural environment will influence how the visual cortex processes the images received by the right and left eyes. In particular, it has been proposed that natural binocular disparity statistics influence the tuning of binocular neurons, the orientation of ocular dominance stripes, and the size of stereoscopic search zones. Here, we calculate binocular disparity statistics in a computer-generated three-dimensional visual environment that has two-point statistics and an observer-object distance distribution similar to that of natural images. Studying binocular disparity statistics using a simulation allows us to ask an additional question—how would a primate primary visual cortex process binocular disparity statistics from an unnatural visual environment (i.e., white noise) using various oculomotor strategies? Our results suggest that the tuning of binocular neurons should be sensitive to the two-point statistics of the visual environment, whereas the orientation of ocular dominance stripes in V1 and stereoscopic search zones are sensitive to the observer-object distance distribution and oculomotor strategy.

## Introduction

Researchers are still trying to figure out how the brain can efficiently combine images from the left and right retina into a single three-dimensional model of the visual environment. The problem itself is deceptively simple: a single point in visual space is imaged onto slightly different locations on our two eyes, and the difference between these two images is referred to as binocular disparity. From the location of these images and knowledge of the distance between the two retina, it is theoretically possible to ray-trace to the position of the original point in visual space and hence construct a three-dimensional model of our visual environment. However, it is not so easy to efficiently find the point in the right retinal image that corresponds to the a given point in the left retinal image, and vice versa. Computational modeling and detailed electrophysiology results are beginning to provide some clues as to how processing in the primary visual cortex allows our brains solve the so-called “correspondence problem” so efficiently and robustly [1].

The primary visual cortex (V1) is the first region to interweave signals from the left and right visual fields, and also appears to be the first stage of concerted neural processing for the correspondence problem [2]. Perhaps unsurprisingly, binocular disparity statistics appear to affect the layout and properties of V1 neurons, in the same way that the receptive fields of V1 neurons seem to shape to optimally encode natural images [3–5]. Previous computational studies have noted that V1 binocular neurons should optimally be tuned to binocular disparities experienced in the natural environment in order to optimally solve the correspondence problem [6, 7]. Another computational study noted that the retinotopic layout of ocular dominance (OD) columns [8–10] was consistent with a minimum wiring length layout [11] of V1 premised on connections between monocular neurons with similar receptive fields in visual space [12]. Similarly, the sensitivity and size of stereoscopic search zones [13, 14] is likely affected by binocular disparity statistics,

though this last area has not yet been explored either computationally or experimentally. Underlying each of these applications is the idea that monocular neurons in V1 that are likely to have similar receptive fields in visual space should somehow be connected, either directly or by sending signals to the same binocular neuron in another layer.

The aforementioned studies that connected binocular disparity statistics to properties of V1 neurons made simplifying assumptions about our visual environment and the way we interact with our visual environment. Reference [6] assumed that all objects were at eye level and used self-reported fixation points. Reference [15] generated a visual environment from spheres with uniformly distributed radii and. Reference [12] simulated binocular disparity statistics using a visual environment composed of point objects and using an eye rotation strategy that only holds for fixation on distant objects. Yet, natural visual scenes have characteristic two-point statistics [16, 17] and observer-object distance distributions [18, 19], and primate eyes rotate according to well-defined mathematical rules [20, 21] when fixating on objects [22–24].

We simulated binocular disparity statistics using a visual environment and eye rotation strategy in ways designed to better match what is known about natural scenes [16–19], eye rotation strategy [20, 21], and eye fixation strategy [22]. From these statistics, we could infer how the tuning and layout of V1 neurons might depend on the structure of visual environments, eye rotation strategy, fixation strategy, and the mapping of retinal ganglion cells to V1. These results have direct bearing on visuomotor optimization theory [25], which quantitatively explains eye rotation strategy as a balance between minimizing the work done by eye muscles and the ease of stereo-matching [13]; the latter variable, in particular, depends strongly on binocular disparity statistics.

## Results

We begin by calculating binocular disparities for a subject in our model world who takes one of three strategies for choosing her fixation points:

- the subject fixates on randomly chosen points on object surfaces;
- the subject fixates on randomly chosen object edges;
- the subject fixates on randomly chosen object centers.

For each fixation strategy, two different visual environments are generated *in silico*: SphereWorld, in which model 3D scenes are generated by placing intersecting, opaque spheres randomly and uniformly over the 3D space surrounding the subject; and PointWorld, in which fixation points and image points are scattered randomly with the constraint that PointWorld must have SphereWorld’s observer-object distance distributions. SphereWorld’s scenes are designed to exhibit naturalistic two-point statistics [16, 17] and naturalistic distance distributions of source points [18, 19]. The model subject then chooses fixation points and rotates her eyes in accordance with the oculomotor strategy known as the binocular extension of Listing’s Law [21], sometimes called L2. Explicitly, this law is as follows: each eye is known to rotate such that its rotation axes are contained in a plane, but during binocular viewing, these planes swing outward by  $\beta + \mu\alpha$  for which  $\alpha$  is the vergence angle,  $\mu$  is the L2 coefficient, and  $\beta$  is the half the exorotation angle of the primary Listing’s planes according to the half-angle rule [21]. See Figure 1. L2 is typically described as having  $\mu = 0.15 - 0.25$  [21, 25], and we will use  $\mu = 0.25$  in our simulations of L2 viewing strategy unless otherwise indicated. We take  $\beta = 2.15^\circ$  based on the results of Reference [26]. This set-up is described schematically in Fig. 1 and described in the Methods section in detail.

Some of these calculations involve calculating stereopsis-related cortical wiring length in V1 as a function of  $\beta$  and  $\mu$  using the following simplistic approximation for stereopsis-related wiring in V1, inspired by the assumptions made in Reference [12] and by the algorithm proposed in Reference [1]. In our simplistic approximation of stereopsis-related neuronal wiring in V1, the probability that two monocular

neurons in V1 are connected to one another (perhaps through an intermediate binocular neuron in another layer) is proportional to the probability that those two monocular neurons have identical receptive fields in visual space. Mechanistically, this is consistent with Hebbian learning [ ] and requires a high degree of synaptic plasticity in the primary visual cortex [ ]. As shown in the Supplementary Information, if the monocular neurons are directly connected, then the relevant quantity is the average cortical wire length; if the monocular neurons are both connected to binocular neurons in another layer, then the relevant quantity is the average square of cortical wire length.

## Retinal disparity histograms are insensitive to all but observer-object distance distribution and primarily Listing’s plane exorotation

Previous authors have investigated retinal disparity statistics [6, 7, 15] and some debate has centered around an appropriate definition of retinal disparity [27]. When the eyes are fixating on points that are at eye-level elevation, the various definitions tend to coincide<sup>1</sup> to the definition illustrated in Figure ???. However, it is not necessarily clear how to extend the definition of retinal disparity to the case when the eyes fixate on oblique objects that are not at eye-level elevations. The disparity is now a vector with two components, but the axes with which disparity should be defined relative to is unclear. In addition, one could reasonably choose a definition of retinal disparity that projects the spherical coordinates of the retina forwards onto a plane tangent to the eyeball or backwards onto a plane perpendicular to the primary gaze direction, as illustrated in Figure 1. Here, we use a simple definition of retinal disparity that accounts for the rotation of the retinal coordinate system as we fixate on different points in a visual scene. Retinal coordinates are defined relative to the rotated primary position axes such that  $x_r = r_r \cos \theta_r$  and  $y_r = r_r \sin \theta_r$ . Retinal disparity is the difference  $\Delta x_r = r_{r,R} \cos \theta_{r,R} - r_{r,L} \cos \theta_{r,L}$  and  $\Delta y_r = r_{r,R} \sin \theta_{r,R} - r_{r,L} \sin \theta_{r,L}$ . A definition based on a rotated coordinate system, as opposed to a definition that does not rotate with the eyeball’s rotation for a fixation point [7], is useful in two ways. First, **since retinal ganglion cells rotate with the eyeball**, the binocular disparities actually experienced by retinal ganglion cells and the V1 binocular cells that they eventually synapse onto are based on the disparities experienced in a coordinate system that rotates with each eyeball. Second, a definition based on a rotated coordinate system allows one to see more directly how eye rotation strategy affects the binocular disparities experienced by retinal ganglion cells and the V1 binocular cells that they eventually synapse onto. **This definition for retinal disparity sidesteps an ambiguity that arises in the definition of retinal disparity [27] by choosing a coordinate system in which the right and left retinae are apparently mapped onto one another using the identity matrix.** Fortunately, these different definitions of retinal disparity give rise to very similar retinal disparity statistics and predictions for physiology.

As shown in Figure 2(A) and Figure 2(B), retinal disparity histograms are essentially insensitive to the visual environment, fixation strategy, and eye rotation strategy within the range of natural visual environment statistics and physiological eye rotation strategies. The main difference between our results and previously reported retinal disparity histograms [6, 7, 15], then, is in our definition of horizontal retinal disparity and in our method of calculation. The asymmetric horizontal retinal disparity histogram is unlike the symmetric retinal disparity histogram in Reference [15] largely because we choose a definition of retinal horizontal disparity that depends on the rotation of the eye during fixation, which then preserves the natural asymmetry in binocular disparity statistics seen in the histograms of optic disparity in Reference [15] and commented upon in Reference [7]. Negative horizontal retinal disparities are rare when using naturalistic observer-object fixation distances [19] and limiting the maximal rotation of the eye during fixation to  $45^\circ$  as in Reference [12], though observers often do not view objects at eccentricities that large. These results are elaborated upon in Supplementary Information.

Reference [6] also postulated that binocular disparity statistics should be reflected directly in the

<sup>1</sup>Except that in Reference [6], the formula used to calculate retinal horizontal disparities is valid only in the limit that the fixation point lies at eye-level elevation on the midsagittal plane.

frequency of tuned excitatory (TE), tuned inhibitory (TI), near excitatory (NE), and near inhibitory (NI) binocular neurons in V1. It is less clear to us that these disparity histograms should directly relate to frequencies of TE, TI, NE, and NI binocular neurons, since TI and NI neurons are better described as phase disparity sensitive binocular neurons which are thought to act as “lie detectors” of false stereomatches [1]. Rather, the tuning of V1 binocular neurons with zero phase disparity should be tuned to the distribution of natural binocular disparities shown in Figure 2C, and the tuning of V1 binocular neurons with nonzero phase disparities should be sensitive to large vertical disparities as described in Reference [7]. Figure 2C and Figure 2D plot the joint probability distribution of observing a particular vertical disparity and a particular horizontal disparity for SphereWorld and PointWorld, respectively, averaged over all fixation points. **Plan is to also discard all disparities that are not Panum-fusable.** In these two plots, the simulated user had  $\beta = 2.15^\circ$ ,  $\mu = 0.25$ , and an edge-based fixation strategy. The constraint on Panum fusability is implemented here by discarding those data points with disparity greater than 0.13 multiplied by the points eccentricity, as in Reference [12]. Similarly to Figure 4 in Reference [7], Figure 2C here shows that we are far more likely to see large horizontal retinal disparities than large vertical retinal disparities. The joint probability distribution between horizontal and vertical disparities is markedly different for SphereWorld and PointWorld. **The presence of bands in the PointWorld plot is analyzed in the Supplemental Information.**

## Inferred ocular dominance maps are insensitive to all but observer-object distance distribution and primary Listing’s plane orientation

In retinotopic coordinates, the physiologically observed OD maps within and between species show two general trends. Ocular dominance stripes tend to lie in concentric circles parafoveally. For smaller eccentricities, closer to the fovea, ocular dominance stripes tend to lie horizontally [8–10]. The “switching eccentricity” from concentric circles to horizontal stripes appears to occur somewhere near  $15^\circ$  for humans as in Figure 3C, but this eccentricity varies from animal to animal idiosyncratically. Near the fovea, the ocular dominance stripes seem to break up into blobs, which might have more to do with the types of connections than with binocular disparity directionality [28].

The minimal wiring hypothesis demands that ocular dominance stripes orient so that the total neuronal wiring length in V1 is minimized. Applying this principle quantitatively, however, requires assuming something concrete about V1 wiring. Following Reference [12], we assume that OD stripe direction is primarily determined by stereoscopic-related wiring, i.e. connections between two monocular neurons that are likely to map the same point in visual space. These connections could be intermediated by a neuron in another layer of V1, such as a position disparity sensitive binocular neuron [2]. Other types of wiring circuits in V1 are more likely to be symmetric directionally, and hence would have little effect on the direction of the ocular dominance stripe [12].

Our simulations generated a list of points in V1 that corresponded to the same points in visual space for various environments. We calculated the ocular dominance stripe orientation that explicitly minimizes wiring length in each small circular region of a flattened V1 by minimizing an expression for the wiring length  $L(\theta)$  as a function of binocular disparities and ocular dominance stripe angle, as described in Methods. The retinotopic OD maps inferred from binocular disparities are insensitive to the two-point statistics of the world and to fixation strategy. See Figure 3A. The inferred OD maps in Figure 3A for  $\beta = 2.15^\circ$  has circular OD stripes only at eccentricities of  $\sim 90^\circ$ , and almost immediately “switches” to horizontal OD stripes. By comparison, physiological OD maps have circular OD stripes parafoveally and switches to horizontal OD stripes far closer to the fovea, near  $15^\circ$ , as shown in Figure 3C.

One might ask whether such a low “switching eccentricity” is possible for our inferred OD maps with realistic simulation parameters. Rather than computing OD maps for a large part of the parameter space, we can derive how this “switching eccentricity” should vary as a function of primary Listing’s plane exorotation angle  $\beta$ , L2 coefficient  $\mu$ , interocular distance  $d$ , and distance at which the probability

of hitting an object surface is maximized  $D_{max}$ , by extending the arguments given by Chklovskii in Reference [12]. In particular, he noted that these OD maps could be conceptualized as a balance between translational disparity and cyclotorsional disparity. Translational disparity is the component of binocular disparity that resulted only from the lateral separation of the two eyes and biases OD stripes towards forming horizontal stripes. Cyclotorsional disparity is the component of binocular disparity that results only from the torsional difference  $|T_L - T_R|$  between the two eyes and biases OD stripes towards forming concentric circles. The magnitude of translational disparity is essentially constant close to and far from the fovea, but the magnitude of cyclotorsional disparity increases linearly with eccentricity, which explains why OD stripes are horizontal near the fovea and circular at higher eccentricities. As described in the Supplementary Information, the switching eccentricity  $\varepsilon_s$  is roughly given by

$$\varepsilon_s \sim \frac{\langle |\alpha_L - \alpha_R| \rangle_{\vec{r}_F, \vec{r}_P}}{\langle |T_L - T_R| \rangle_{\vec{r}_F, \vec{r}_P}} \quad (1)$$

where  $\langle \dots \rangle_{\vec{r}_F, \vec{r}_P}$  indicates an average over all fixation points and all imaged points. The numerator depends only on the ratio between the two distance scales,  $\frac{d}{D_{max}}$ , while the denominator depends strongly on eye rotation parameters  $\beta$  and weakly on  $\mu$  and  $\frac{d}{D_{max}}$ . As shown in Supplementary Information,

$$\varepsilon_s \propto \frac{(d/D_{max})^2}{\beta + \mu(d/D_{max})^2} \quad (2)$$

which implies the following: that an increase in  $\beta$  decreases the switching eccentricity; that an increase in L2 coefficient  $\mu$  also decreases the switching eccentricity, but with less effect than  $\beta$ ; and that an increase in interocular distance  $d$  will increase the switching eccentricity.

Based on eqn. 2, an environment with a larger  $D_{max}$ , a person with a smaller interocular ocular distance  $d$ , and a person with a larger primary Listing's plane exorotation angle  $\beta$  will produce a map that is more physiologically realistic. Reference [19] has reported  $D_{max}$  as large as 3 m, and  $\beta$  could be as large as  $2.6^\circ$  [26]. However, even these values of parameters do not yield a physiologically realistic OD map; the predicted switching eccentricity is near  $60^\circ$ . See Figure 3(B).

## Visuomotor theory

Binocular disparity statistics are also thought to affect the optimal eye rotation strategy. In order to fuse the images from both eyes, our eyes must fixate on the same point. In theory, they can do this in infinitely many ways, as shown in Figure 1B, but in reality, our eyes choose just one of these rotations in a way that is well-described by a mathematical rule called Listing's Law. This rule states that the position eventually adopted by the eye is the same as the position that would have been adopted if the eye had started from a "primary position" and rotated about an axis in the corresponding Listing's plane, as depicted in Figure 1B. Equivalent definitions in terms of Helmholtz coordinates are given in the Supplementary Information section.

Since Listing's Law was first discovered experimentally by Helmholtz in the eighteenth century, vision scientists have tried to explain why our eyes appear to so consistently choose one of the infinitely many other rotation matrices that would allow us to fixate on the same fixation point. Initially, Fick postulated that Listing's Law was an attempt to decrease the total energy used for eye rotation. This explanation proved insufficient when vision researchers discovered that Listing's planes exorotated slightly when fixating on nearby objects. In 1997, Tweed's pioneering visuomotor optimization theory explained the binocular extension of Listing's Law by postulating that our eyes' rotation strategy is a compromise between the visual system and the motor system [25]. Minimizing the total eye rotation implies the existence of Listing's planes that lie orthogonal to the midsagittal plane and the interocular line, and Tweed further postulated that the ease of stereomatching was directly correlated with the difference in alignment of the horizontal meridians of the two eyes. Later investigators showed that  $\mu = 0.25$  decreased

the area of stereoscopic search zones [13] and increased the surface area of Panum-fusable points in the visual environment [27].

Stereoscopic search zones are the regions over which a stereoscopic match could exist, and sensitivity profiles within these stereoscopic search zones should indicate which areas within this search zone are most likely to match the point of origin [14]. Sensitivity profiles of stereoscopic search zones are likely to reflect the probability of observing a wire in V1 between V1 neurons with those receptive fields [14]. We computed sensitivity profiles for a PointWorld environment, as shown in Figure 4. When  $\beta = 0^\circ$ , increasing  $\mu$  from 0 to 0.25 visibly decreases stereoscopic search zones, but when  $\beta = 2.15^\circ$ , increasing  $\mu$  actually increases stereoscopic search zones. See Figure 4. The shape of the sensitivity profiles of these stereoscopic search zones mimics the OD maps shown in Figure 3; when the zones look square rather than elongated and circular, the corresponding optimal ocular dominance stripe is horizontal rather than circular.

The benefit to minimizing the stereoscopic search zone area could be that it takes less time or that it takes less energy to perform stereoscopic processing in the primary visual cortex. In either case, an “optimal” eye rotation strategy should minimize the total cortical wiring length required to connect the stereoscopic search zones. We calculated the stereopsis-related cortical wiring length in V1 as a function of  $\beta$  and  $\mu$  using the following simplistic approximation for stereopsis-related wiring in V1 described earlier. Though the absolute value of mean cortical wiring length value was highly dependent on the exact assumptions made about stereopsis-related wiring, the dependence of mean cortical wiring length on  $\beta$  and  $\mu$  is essentially independent of these assumptions. For instance, one could imagine excluding wires that connect across hemispheres or that connect beyond Panum’s fusional area. For simplicity, we show the results in which no additional assumptions are made about interhemispherical connections or connections outside of Panum’s fusional area. Results of a PointWorld calculation of average cortical wirelength is shown in Figure 5. These calculations reveal a quadratic surface for cortical wiring length with a very shallow well along one axis and a very steep well along another, which indicates that there are large number of pairs of  $\beta$  and  $\mu$  that appear to give similarly small values for total cortical wiring length. When  $\beta = 0^\circ$ , the optimal  $\mu$  appears to be near 0.25 as shown in Figure 5(C), which is the value of  $\mu$  that minimizes cyclotorsional disparity [25]. However, for  $\beta = 2.15^\circ$ , the optimal  $\mu$  is negative, as shown in Figure 5(D).

## Discussion

We started by postulating an intimate connection between binocular disparity statistics, V1, and oculomotor theory, but many researchers study each of these topics without referencing any of the others. The connection between these three topics arises from three optimization principles: the efficient coding hypothesis, the cortical wiring hypothesis, and visuomotor optimization theory. The efficient coding hypothesis states that our brain has evolved to optimally process information from our environment, as opposed to some other environment [29]. The cortical wiring hypothesis states that the layout of our brain should shift so as to minimize total cortical wiring, typically because large heads are associated with high infant mortality or because the additional cortical wiring leads to energetic inefficiency and time delays [11]. Finally, visuomotor optimization theory is part of a larger class of sensorimotor optimization theories that attempt to derive “optimal” motor behavior [30]. In order to calculate anything using these principles, we must introduce a simplistic model of stereopsis-related wiring in V1. We therefore assume monocular neurons that are likely to have similar receptive fields in visual space are connected to binocular neurons (or at least, those that look for “true matches” [1]). The final word on stereoscopic match-related neuronal circuitry in V1 may come from undertakings similar to EyeWire, a tour de force artificial intelligence attempt to explicitly make the connectome of retinal ganglion cells [1]. Otherwise, psychophysical experiments and computations such as the ones shown here may provide some guesses as to how to interpret the resulting connectome.

The three optimization principles listed above naturally lead to several conclusions when applied to stereopsis-related wiring in V1. First, the efficient coding hypothesis suggests that the V1 binocular neurons primed to find “true” stereomatches [1] should be tuned to sense the range of disparities present in the natural environment seen in Figure 2. Second, the cortical wiring hypothesis suggests that the layout of V1 should be arranged so as to connect monocular neurons with similar receptive fields with minimal wiring. Third, a variant of visuomotor optimization theory suggests that Listing’s Law coefficients should yield small stereoscopic search zones [13]. If these three optimization principles hold and if our simplistic conception of V1 wiring is at all reasonable, then a comprehensive model of binocular disparity statistics should allow one to predict or explain the tuning of V1 neurons, the orientation of ocular dominance stripes, and optimal eye rotation strategy.

To our knowledge, SphereWorld (described in detail in the Model section) represents the most comprehensive computational model for calculating binocular disparities to date. The visual environment takes into account both two-point statistics and observer-object distance distribution, as shown in Figure 6; the simulated eye rotations act according to the binocular version of Listing’s Law; and retinal disparities are mapped to V1 using a Schwarz conformal map with parameters fit to physiological data. Each of these ideas has been present in part in other computational studies, but never have these ideas been combined into one single computational model with which to study binocular vision. These types of computational models are incredibly powerful tools in that we can study the effect of slight variations of physiological and environmental parameters on binocular disparity statistics in ways that are nearly impossible to study empirically. For instance, by comparing binocular disparity statistics between SphereWorld and PointWorld, a visual environment composed of point objects with the same observer-object distance distributions as SphereWorld, we can examine the effect of two-point statistics on binocular disparity statistics.

A number of previous studies have made similar models and calculated similar properties of binocular disparity statistics [6, 7, 15], but combining parts of each of these previous models and adding additional details to the simulations allowed us to refine and sometimes extend the conclusions of our predecessors. Here, we highlight three such refinements.

First, the retinal disparity statistics plotted in Figure 2 indicate that rough disparity histograms (Figures 2A and 2B) are insensitive to the details of the visual environment. However, the joint probability distributions between vertical and horizontal disparities are extremely sensitive to the type of two-point statistics in the environment, as shown by the difference in Figure 2C and 2D. The ellipses of equal probability are far more elongated in the SphereWorld environment than in the PointWorld environment, which suggests that searching for stereomatches with large vertical disparities is far less fruitful in a natural environment than in a world composed of point objects.

Second, the ocular dominance maps inferred from binocular disparity statistics are not clearly similar to physiological ocular dominance maps, in opposition to the conclusions of Reference [12]. Perhaps this is because one of our simulation assumptions is incorrect. For instance, if our stereopsis is optimized for long-distance viewing cues [14], then the switching eccentricity will decrease greatly according to eqn. 2. Alternatively, the cortical wiring hypothesis might not apply to ocular dominance columns in the way postulated in Reference [12]; perhaps ocular dominance column width should correspond roughly to the range of natural binocular disparities. The relationship between ocular dominance stripe orientations and binocular disparity maps can be tested empirically by constructing retinotopic ocular dominance maps for animals, who will naturally have different interocular distances and primary Listing’s plane exorotation angle. If the ocular dominance map follows the binocular disparity map, then animals with smaller interocular distances (e.g. squirrel monkeys) and larger primary Listing’s plane exorotation angles will have larger switching eccentricities.

Third, the simulated sensitivity profiles of stereoscopic search zones shown in Figure 4 indicate that only a small part of the stereoscopic search zones shown in Reference [13] are likely to be used. The total wiring length is quantified as a function of  $\beta$  and  $\mu$  in Figure 5, and those plots indicate a long, shallow

well for visual term in visuomotor theory. In the context of visuomotor theory, the degeneracy between equivalent  $\beta$ ,  $\mu$  pairs is broken by the motor term, which prefers  $\beta$  close to the natural resting position of the eyeballs and  $\mu = 0$ . However, for  $\beta \simeq 2^\circ$ , the optimal  $\mu$  is not close to the physiologically observed value of 0.25. This is identically true even for other proxies for visual benefit that one might use in visuomotor optimization theory, as shown in Supplementary Information. This suggests that visuomotor optimization theory needs further refinement.

Overall, our simulations showed that only two simulation parameters had a significant effect on binocular disparity statistics. These null results are, in our minds, good news for the stereovision research community. Other investigators looking to compute binocular disparities *in silico* will only need to make sure that their simulation has a realistic observer-object distance distribution, a realistic value of interocular distance, a realistic value of the primary Listing's plane exorotation angle, and a realistic value of the Listing's Law coefficient. Future simulations of binocular disparity might benefit from using a visual environment with more interesting features, a time-dependent fixation strategy, sensorimotor feedback loops, and a more complex model of stereopsis-related wiring in V1. Eye tracking experiments, such as those currently being undertaken at the Redwood Center for Theoretical Neuroscience, will have the final word on actual binocular disparity histograms. Any significant differences between the binocular disparity histograms generated from eye tracking data and those shown here will signal that sensorimotor feedback loops are key to understanding the binocular disparity statistics received by our brain.

## Models

The *in silico* visual environment was generated using Matlab, and all subsequent computations were done in Matlab as well.

### Generating a *in silico* visual environment

There are two elements of visual environments that have proved key to understanding visual processing in the brain: two-point statistics and the observer-object distance distribution. We generated a visual world *in silico* that had two-point statistics and an observer-object distance distribution representative of two-point statistics and observer-object distance distributions in visual environments.

Images of natural scenes all have similar two-point statistics [17]. Pair-wise statistics describe the fall-off of the correlation function,  $\langle I(\vec{x}_0)I(\vec{x}_0 + x\hat{\theta}) \rangle_{\vec{x}_0, \hat{\theta}} \sim C_1 - \frac{C_2}{x^\eta}$  between two pixels in an image as a function of distance between the pixels,  $x$ , where  $I(\vec{x}_0)$  denotes the pixel intensity at point  $\vec{x}_0$ . The constants  $C_1$ ,  $C_2$  and  $\eta$  vary from image to image, and  $|\eta|$  is typically small. These pair-wise statistics can also be described by the Fourier transform of the correlation function, also called the power spectrum; this falls off as  $\sim \frac{1}{k^{2-\eta}}$ , where  $k$  is the spatial frequency. In general, images of a random grouping of objects will have a correlation function that decreases as a function of distance; pixels that are close together are more likely to represent the same object, and therefore are more likely to have similar pixel values. However, not all images of groupings of objects will have a correlation function that decreases as  $\sim \frac{1}{x^\eta}$ . For instance, when the objects are spheres with radii drawn from a power law probability distribution,

$$P(s) = \begin{cases} (\alpha - 1) \left(\frac{s}{s_0}\right)^{-\alpha} & s \geq s_0 \\ 0 & s < s_0 \end{cases} \quad (3)$$

then images of the resulting visual environment will have a correlation function that falls off as  $\frac{1}{x^{\alpha-3}}$  [16]. Hence, randomly placed spheres with radii drawn from the  $P(s)$  given in eqn. 3 with  $\alpha \simeq 3$  will have a correlation function and thus power spectrum similar to that of natural images.

To our knowledge, no previous computational studies of binocular disparity statistics studied worlds with realistic two-point statistics. In Reference [15] a visual environment was generated as a collage of



spheres with radii drawn from a uniform distribution [15]. This is mathematically equivalent to using  $\alpha = 0$  for  $P(s)$  in eqn. 3 and adding an upper cutoff on the radius. The analysis of Reference [16] suggests that such an environment would yield a correlation function  $\langle I(\vec{x}_0)I(\vec{x}_0 + x\hat{\theta}) \rangle_{\vec{x}_0, \hat{\theta}} \sim C_1 + C_2 x^3$ , where  $C_1$  and  $C_2$  are constants. In Reference [12], the visual environment consisted of point objects. This is mathematically equivalent to using  $\alpha \rightarrow \infty$  for  $P(s)$  in eqn. 3, which would yield a correlation function  $\langle I(\vec{x}_0)I(\vec{x}_0 + x\hat{\theta}) \rangle_{\vec{x}_0, \hat{\theta}} = C_1$ , where  $C_1$  is some constant. Neither of these correlation functions is representative of the correlation functions found in natural scenes.

Any computer-generated visual environment that is used to calculate binocular disparity statistics should also have a naturalistic distance distribution of object surfaces. In natural environments, the probability of observing an object surface at a distance  $r$  from a cyclopean eye increases until some distance  $D_{max}$ , after which point the probability appears to decrease exponentially [19]. This probability distribution seems to determine how we infer distances of objects [18]. In a SphereWorld with object density  $\rho$ ,

$$P(r) \sim r^2 e^{-\rho(\int s^2 P(s) ds)r}, \quad (4)$$

as shown in Supplemental Information. In short, the  $r^2$  term comes from the increase in surface area, and the exponential term  $e^{-\rho(\int s^2 P(s) ds)r}$  is a result of object occlusion. From eqn. 4, we see that any computer-generated visual environment with randomly placed occluding objects, e.g. as in Reference [15], will have a naturalistic  $P(r)$  with  $D_{max} \simeq \frac{2}{\rho \int s^2 P(s) ds}$ .

In this study, we generated a visual environment composed of occluding, randomly placed spheres with radii drawn from the distribution  $P(s)$  given in eqn. 3 using parameters  $s_0 = 15$  cm,  $\alpha = 3.2$ , and the number of spheres generated was chosen so that  $D_{max} = 1.3$  m. Smaller  $s_0$  would have been more desirable, but as long as  $s_0$  is small enough, then a simulated viewer will be unable to “see” the environment’s graininess. For ease, we call this environment “SphereWorld”. As was done in Reference [12], we assumed that all object centers were at least  $L_{min} = 30$  cm away, and chose a finite world size of  $L = 60$  m. The finite world size has no significant effect on  $P(s)$  or  $P(r)$ , as shown analytically in Supplemental Information. SphereWorld’s correlation function, power law distribution, and distance distribution of object surfaces in this study are shown in the Supplemental Information.

## Simulating head-fixed visual sampling of this model world

There are two aspects to simulating interaction with the visual environment: first, choosing fixation points that might be chosen by a primate; and second, simulating how a primate would rotate its eyes to fixate on those points.

Fixation strategy is most correctly modeled as a function of time, but we are interested only in the time-averaged difference between the images received by the right and left retina. Hence, we only need to determine which features in the visual world of occluding gray spheres are “informative” or “salient” []. For simplicity, we consider three fixation strategies: one in which the informative features are randomly chosen, unoccluded edges of objects; one in which no points in visual space are more informative than any other; and one in which the unoccluded centers of objects are the most informative features. The first simulated fixation strategy is the closest of the three to a realistic fixation strategy [22]. The other two simulated fixation strategies are control fixation strategies, designed to elucidate the effects of fixation strategy on binocular disparity statistics. The distance distribution of fixation points varies slightly between these three strategies, since centers of objects are closer than the edges of objects. This difference depends on the average radii of the spheres in the environment, as shown in the Supplemental Information section. Furthermore, as was done in Reference [12], we assume that we rotate our eyes to fixate only on objects that have an eccentricity of less than  $\theta_{max} = 45^\circ$  []. Simulation results did not depend on  $\theta_{max}$ , whether it was  $30^\circ$ ,  $45^\circ$ ,  $60^\circ$ , or  $90^\circ$ .

To fixate on these various points, primates move both their heads and eyes. At the risk of oversimplification, we assumed that the head was fixed, and that imaged points could be ray-traced onto the eye

according to the pinhole eye approximation [31]. Head-fixed eye rotation strategy has been well-studied and is well-described by an explicit mathematical rule called Listing’s Law [21,32–35]. There are different ways of stating Listing’s Law and many ways of explaining why Listing’s Law is the optimal eye rotation strategy [20,21,25,32]. For our purposes, it is easiest to state Listing’s Law in terms of the orientation of the Listing’s plane, which is the plane in which all rotation axes of every eye rotation lies. Even when fixating on distant objects, the Listing’s plane is exorotated from a perfectly flat orientation by some angle  $\beta$ . In humans,  $\beta \simeq 2.15^\circ$  [26]. When fixating on nearby objects, Listing’s planes are additionally rotated by  $\mu\alpha$ , where  $\mu$  is the so-called “L2 coefficient” and  $\alpha$  is the vergence angle. For humans, the L2 coefficient  $\mu$  lies somewhere between 0.15 and 0.25 [21]. The meaning of  $\beta$ ,  $\mu$ , and  $\alpha$  are illustrated in Figure 1. Eye rotations can be described by the rotation axis and rotation angle, but they can also be described by Helmholtz coordinates [32]. In the Supplemental Information, we derive the Helmholtz coordinates for the left and right eye as a function of the fixation point location,  $\beta$ , and  $\mu$ . These closed-form expressions for the Helmholtz coordinates appear in Reference [7] and speed up our simulations by orders of magnitude as compared to solving for the rotation axis direction and rotation angle using a numerical solver.

## Mapping retinal images to V1

Retinal coordinates were projected back to a flattened representation of V1 using the Schwartz conformal map [36], matched to the experimental areal cortical magnification factor  $M = a(r_r + b)^{-2}$  of squirrel monkeys [9]. The experimental data do not perfectly fit an areal magnification factor formula of  $M = a(r_r + b)^c$  for any  $a$ ,  $b$ , or  $c$ , and the experimental data used here are subject to small systematic errors, depending on the measurement [37]. However, the results presented here do not depend very strongly on the values of  $a$ ,  $b$ , or  $c$ . The Schwartz conformal mapping takes the form  $w = a \log(z + b)$ , wherein  $w = r_{V1}e^{i\theta_{V1}}$ ,  $z = r_R e^{i\theta_R}$  and  $r$  is the radial/eccentricity coordinate,  $\theta$  is a polar coordinate.

As described in Results, we calculate the ocular dominance stripe orientation that explicitly minimizes wiring length in each small circular region of a flattened V1. Ocular dominance stripes theoretically affect V1 wiring length by stretching a disparity vector/neuronal wire by a factor of  $a \sim 2$  in the direction perpendicular to the ocular dominance stripe [12], as illustrated in Figure 1. Let an ocular dominance stripe in a particular region of V1 made an angle  $\theta$  with the horizontal axis, and that a disparity vector in that region connected the points  $(0, 0)$  and  $(x, y)$ . Calculations in the Supplemental Information show the length of the disparity vector  $l$  ranges from  $\sqrt{x^2 + y^2}$  to  $a\sqrt{x^2 + y^2}$  as a function of  $\theta$  according to the following equation:

$$l(\theta) = \sqrt{(x \cos \theta + y \sin \theta)^2 + a^2(x \sin \theta - y \cos \theta)^2} \quad (5)$$

When there are multiple disparity vectors in a particular region of V1, the total wiring length is a sum of the individual wiring lengths,

$$L(\theta) = \sum_i l_i(\theta) = \sum_i \sqrt{(x_i \cos \theta + y_i \sin \theta)^2 + a^2(x_i \sin \theta - y_i \cos \theta)^2}. \quad (6)$$

Results are not affected by restricting consideration to only the disparity vectors that are within Panum’s fusional area and intra-hemisphere were considered, as in Reference [12], as shown in Supplementary Information. The  $\theta$  that minimizes the total wiring length of that region in Equation 6 is presumed to determine the optimal ocular dominance stripe direction.

## Acknowledgments

Very special thanks go to Douglas Adams and Jonathan Horton, who generously supplied us with empirical retinotopic ocular dominance maps with which to compare our predictions. S. Marzen would like

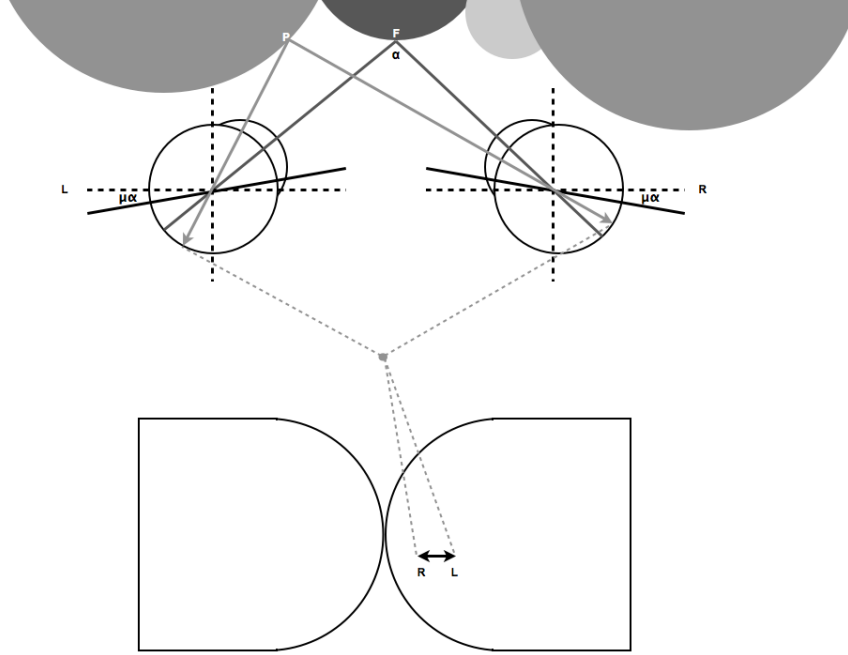
to acknowledge the generous support of the NSF Graduate Research Fellowship and a U.C. Berkeley Chancellor's Fellowship. **Need other names.**

## References

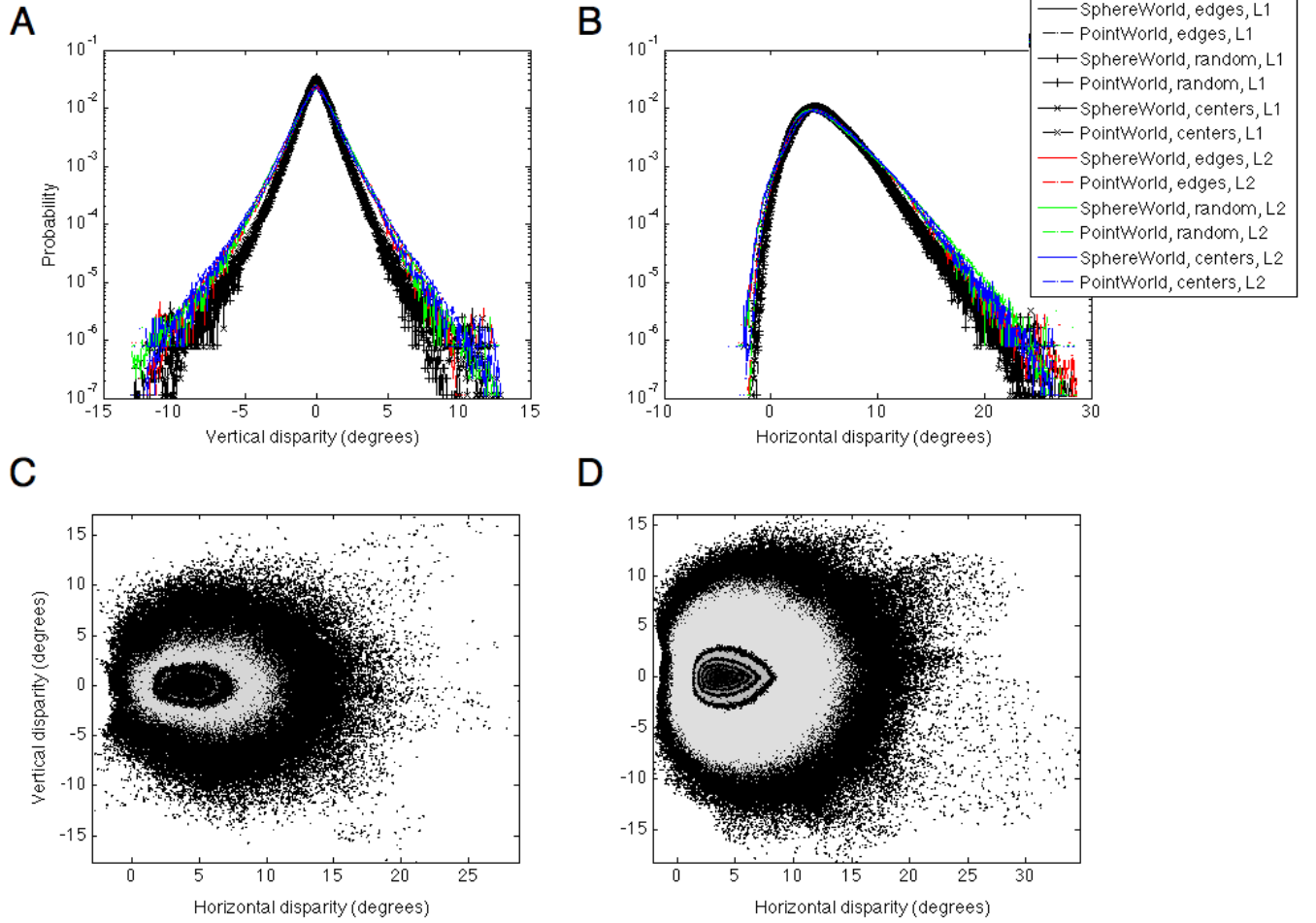
1. Read JCA, Cumming BG (2007) Sensors for impossible stimuli may solve the stereo correspondence problem. *Nature Neuroscience* 10: 1322-1328.
2. Read J (2005) Early computational processing in binocular vision and depth perception. *Progress in Biophysics and Molecular Biology* 87: 77-108.
3. Olshausen B, Field DJ (1996) Emergence of simple cell receptive fields by learning a sparse code for natural images. *Nature* 381: 607 - 609.
4. Karklin Y, Lewicki M (2008) Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* 457: 83 - 86.
5. Karklin Y, Simoncelli EP (2011) Efficient coding of natural images with a population of noisy linear-nonlinear neurons. In: Shawe-Taylor J, Zemel R, Bartlett P, Pereira F, Weinberger K, editors, *Advances in Neural Information Processing Systems* 24. pp. 999-1007.
6. Liu Y, Bovik AC, Cormack LK (2008) Disparity statistics in natural scenes. *Journal of Vision* 8: 1-14.
7. Read J, Cummings B (2004) Understanding the cortical specialization for horizontal disparity. *Neural Comput* 16: 1983-2020.
8. Adams DL, Horton JC (2009) Ocular dominance columns: Enigmas and challenges. *The Neuroscientist* 15: 62-77.
9. Adams DL, Sincich LC, , Horton JC (2007) Complete pattern of ocular dominance columns in human primary visual cortex. *Journal of Neuroscience* 27: 10391-10403.
10. LeVay S, Connolly M, Houde J, Essen DCV (1985) The complete pattern of ocular dominance stripes in the striate cortex and visual field of the macaque monkey. *Journal of Neuroscience* 5: 486-501.
11. Chklovskii DB, Koulakov AA (2004) Maps in the brain: What can we learn from them? *Annual Review of Neuroscience* 27: 369-392.
12. Chklovskii DB (2000) Binocular disparity can explain the orientation of ocular dominance stripes in primate primary visual area (v1). *Vision Research* 40: 1765-1773.
13. Schreiber K, Crawford JD, Fetter M, Tweed D (2001) The motor side of depth vision. *Nature* 410: 819-822.
14. Phillipson GP, Read JCA (2010) Stereo correspondence is optimized for large viewing distances. *European Journal of Neuroscience* : 1-11.
15. Hibbard PB (2006) A statistical model of binocular disparity. *Visual Cognition* 15: 149-165.
16. Ruderman DL (1997) Origins of scaling in natural images. *Vision Research* 37: 3385-3398.
17. Ruderman DL, Bialek W (1994) Statistics of natural images: Scaling in the woods. *Physics Review Letters* 73: 814-817.

18. Yang Z, Purves D (2003) A statistical explanation of visual space. *Nature Neuroscience* 6: 632-640.
19. Yang Z, Purves D (2003) Image/source statistics of surfaces in natural scenes. *Network: Computation in Neural Systems* 14: 371-390.
20. Hepp K (1994) Oculomotor control: Listing's law and all that. *Current Opinion in Neurobiology* 4: 862-868.
21. Wong AMF (2004) Listing's law: Clinical significance and implications for neural control. *Survey of Ophthalmology* 49: 563-575.
22. Reinagel P, Zador AM (1999) Natural scene statistics at the centre of gaze. *Network: Computational Neural Systems* 10: 1-10.
23. Schumann F, Einhäuser W, Vockeroth J, Bartl M, Schneider E, et al. (2008) Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *J Vis* 8: 1 - 17.
24. 'tHart BM, Vockeroth J, Schumann F, Bartl K, Schneider E, et al. (2009) Gaze allocation in natural stimuli: comparing free exploration to head-fixed viewing conditions. *Vis Cog* 17: 1132- 1158.
25. Tweed D (1997) Visual-motor optimization in binocular control. *Vision Research* 37: 1939-1951.
26. P Bruno AVvdB (1997) Relative orientation of primary positions of the two eyes. *Vision Research* 37: 9350947.
27. Schreiber KM, Tweed DB, Schor CM (2006) The extended horopter: Quantifying retinal correspondence across changes of 3d eye position. *Journal of Vision* 6: 64-74.
28. Chklovskii DB, Koulakov AA (2000) A wire length minimization approach to ocular dominance patterns in mammalian visual cortex. *Physica A* 284: 318-334.
29. Simoncelli EP, Olshausen BA (2001) Natural image statistics and neural representation. *Annual Review of Neuroscience* 24: 1193-1216.
30. Tweed D (2007) Sensorimotor optimization in higher dimensions. *Progress in Brain Research* 165: 181-191.
31. Grand YL, Hage SGE (1980) *Physiological optics*. Springer-Verlag.
32. Haslwanter T (1995) Mathematics of three-dimensional eye rotations. *Vision Res* 35: 1727-1739.
33. Hooge ITC, van den Berg AV (2000) Visually evoked cyclovergence and extended listing's law. *Journal of Neurophysiology* 83: 2757-2775.
34. Rijn LJ, van den Berg AV (1993) Binocular eye orientation during fixation: Listing's law extended to include eye vergence. *Vision Research* 33: 691-708.
35. Mok D, Ro A, Cadera W, Crawford JD, Vilis T (1992) Rotation of listing's plane during vergence. *Vision Research* 32: 2055-2064.
36. Schwartz EL (1980) Computational anatomy and functional architecture of striate cortex: A spatial mapping approach to perceptual coding. *Vision Research* 20: 645-669.
37. Giacomantonio CE, Goodhill GJ (2007) The effect of angioscotomas on map structure in primary visual cortex. *Journal of Neuroscience* 27: 4935-4946.

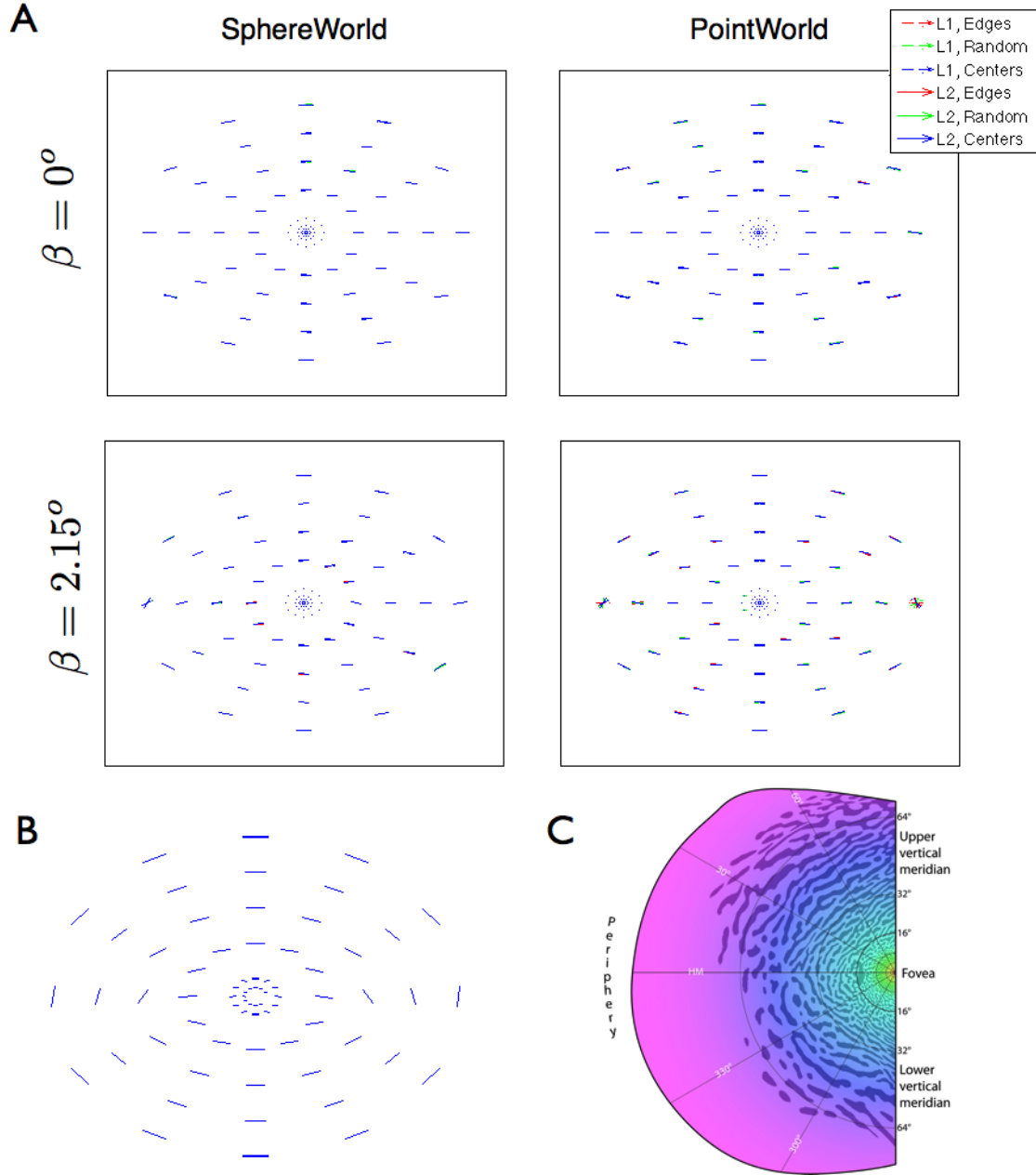
## Figure Legends



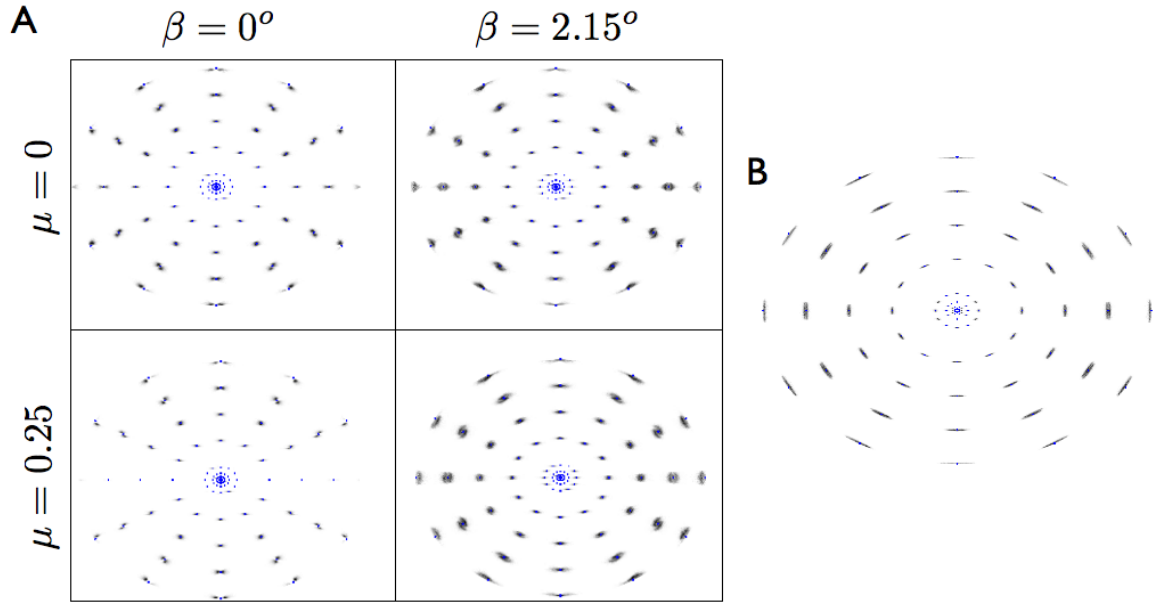
**Figure 1. Images from a model world are mapped to each retina and then mapped back to a flattened V1.** At the top is an artistic representation of the model world generated using Equation 3. Point F denotes the fixation point, and point P is projected onto each of the two retina using the pinhole eye approximation [31], in which rays from a source point pass straight through the nodal point of the spherical eye. As shown, the generalized Listing's planes are rotated by  $\mu\alpha$  from a head-fixed Listing's plane, which we assume is perpendicular to the midsagittal plane, though it varies experimentally from person to person [33–35]. The images of point P on each retina project to V1 neurons whose positions are determined by the retinotopic mapping, which we model using the Schwartz conformal mapping [36], as discussed in the text. The black arrow in the flattened representation of the primary visual cortex represents a connection between the two neurons that are sensitive to the same part of the visual space. This process is repeated for multiple fixation points F and imaged points P. The grey circle where the dotted lines join represents the optic chiasm. The diagram is not to scale.



**Figure 2. Distributions of observed disparities depends on fixation strategy.** Figures 2A and 2B plot vertical and horizontal retinal disparity histograms, respectively, when  $\beta = 2.15^\circ$ . All y-axes are on a log scale. The different line colors, line types, and markers stand for different fixation strategies, eye rotation strategies, and world types, described in detail in the legend at the top right. Figures 2C and 2D plot a normalized histogram of observing a particular horizontal disparity and vertical disparity pair in SphereWorld and PointWorld, respectively, for a simulated user with  $\beta = 2.15^\circ$ ,  $\mu = 0.25$ , and an edge-based fixation strategy. Darker regions are more likely than lighter regions, as indicated by the colorbar at bottom right.

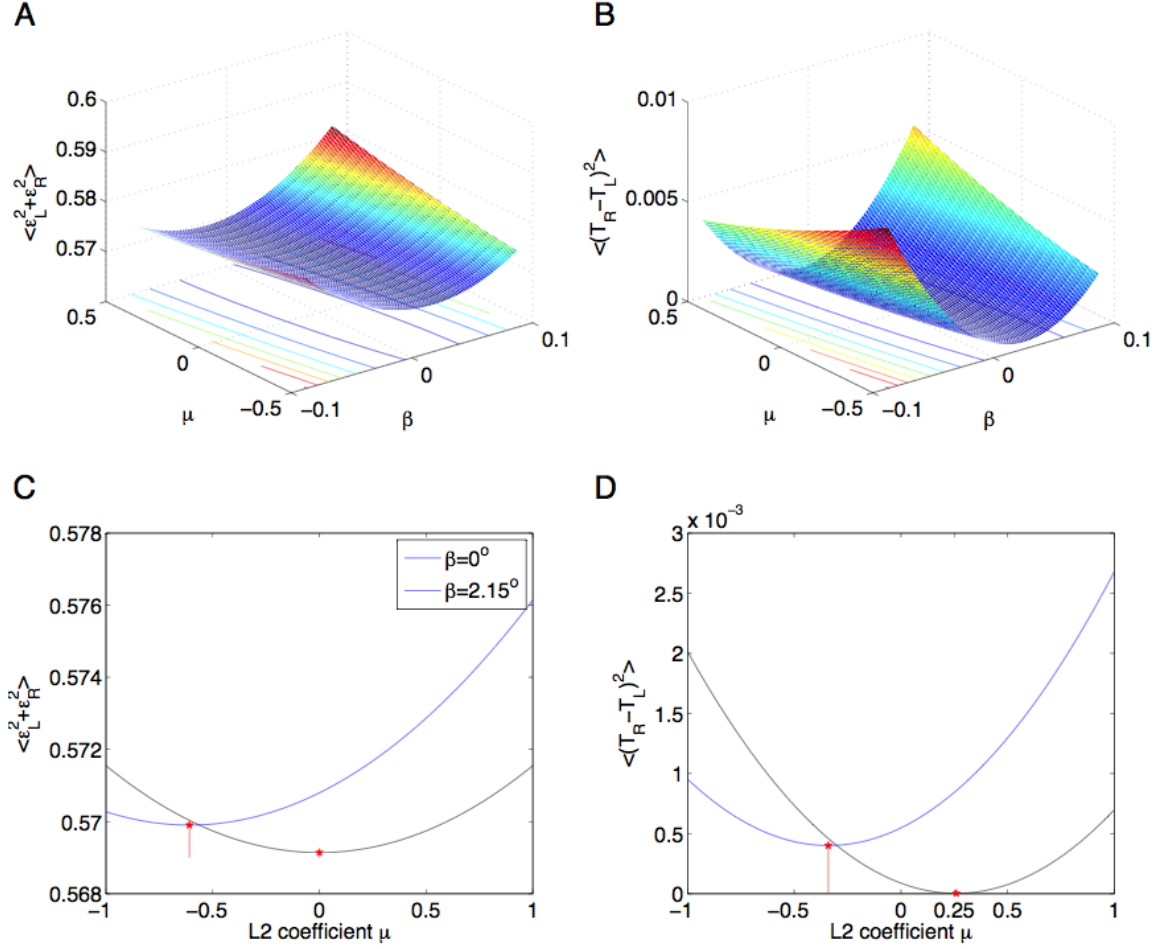


**Figure 3. Simulated binocular disparity maps have higher switching eccentricities than those in physiological OD maps.** Figure 3 shows the inferred OD maps for various combinations of simulation parameters. Rows of the table correspond to two different values of the primary Listing’s planes exorotation angle,  $\beta = 0^\circ$  and  $\beta = 2.15^\circ$ ; columns of the table of the table correspond to binocular disparities in SphereWorld with parameters shown in Figure 6 and disparities in PointWorld with the equivalent observer-object distance distribution. Each map includes the three fixation strategies (edges, random, centers) and two eye rotation strategies (L1 and L2) as shown in the legend at top right. Figure 3B shows an inferred PointWorld OD map using simulation parameters that are designed to minimize switching eccentricity within bounds for what values of simulation parameters are realistic, using  $D_{max} = 3\text{ m}$ ,  $\beta = 2.65^\circ$ , and  $\mu = 0.25$ . Figure 3C shows a physiological OD map of a human, generously supplied by Horton and Adams [8]. All inferred OD maps show OD stripes starting at  $90^\circ$  eccentricity and at eccentricities decreasing successively by  $20^\circ$ .

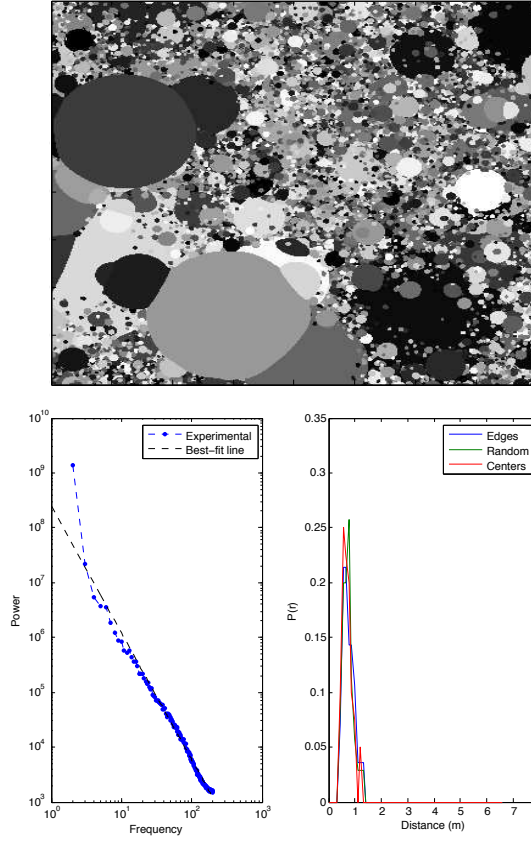


**Figure 4. L2 does not reduce stereoscopic search zone area when primary Listing's planes are exorotated.** Figure 4A shows stereoscopic search zones for various combinations of  $\beta$  and  $\mu$  in a PointWorld environment when  $D_{max} = 1$  m. Figure 4B shows stereoscopic search zones when  $\beta = 2.65^\circ$ ,  $\mu = 0.25$ , and  $D_{max} = 3$  m, as in Figure 3B. The blue points denote the point on the retina for which we calculated the stereoscopic search zone, and the darkness is proportional to the density of wires that fall in that region of the stereoscopic search zone.





**Figure 5. Exorotated primary Listing's planes alter the optimal eye rotation strategy, according to Tweed's visuomotor optimization theory.** Tweed's visuomotor optimization theory includes two terms, one motor ( $\langle \epsilon_L^2 + \epsilon_R^2 \rangle$ ) and one visual ( $\langle (T_R - T_L)^2 \rangle$ ), as described in the text. In Figure 5A and Figure 5B,  $\langle \epsilon_L^2 + \epsilon_R^2 \rangle$  and  $\langle (T_R - T_L)^2 \rangle$ , respectively, are plotted as a function of  $\beta$  and  $\mu$ , with contour lines below the three-dimensional plot. Figure 5C and Figure 5D show how  $\langle \epsilon_L^2 + \epsilon_R^2 \rangle$  and  $\langle (T_R - T_L)^2 \rangle$  vary as a function of L2 coefficient  $\mu$  for two values of  $\beta$ :  $\beta = 0^\circ$  and  $\beta = 2.15^\circ$ , the latter of which corresponds to the physiologically observed primary Listing's plane exorotation angle. The value of  $\mu$  which minimizes each curve is marked with a red pentagon and a red line to the y-axis on each plot for each value of  $\beta$ . Note that when  $\beta = 0^\circ$ , the optimal  $\mu$  from a motor perspective is  $\mu = 0$  and the optimal  $\mu$  from a visual perspective is the physiologically observed  $\mu \simeq 0.25$ , as noted by Tweed in Reference [25].



**Figure 6. SphereWorld statistics.** At top, an image of SphereWorld. At bottom left, the power spectrum of the image above, with a best fit line of  $S(k) \sim \frac{1}{k^{2.35}}$ , which agrees well enough with the predicted scaling of  $S(k) \sim \frac{1}{k^{3.2}}$ . At bottom right,  $P(r)$  for the edges, random, and centers, showing that  $D_{max} \simeq 1$  m for all three fixation strategies.