Stereopsis is adaptive for the natural environment

William W. Sprague[#1,2,†], **Emily A. Cooper**[#3,†], **Ivana Toši** [4,‡], and **Martin S. Banks**[1,2]

[1]Vision Science Graduate Group, University of California, Berkeley, Berkeley, CA 94720, USA.

[2]School of Optometry, University of California, Berkeley, Berkeley, CA 94720, USA.

[3]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720, USA.

[4]Ricoh Innovations Corp., Menlo Park, CA 94025, USA.

[#] These authors contributed equally to this work.

Abstract

Humans and many animals have forward-facing eyes providing different views of the environment. Precise depth estimates can be derived from the resulting binocular disparities, but determining which parts of the two retinal images correspond to one another is computationally challenging. To aid the computation, the visual system focuses the search on a small range of disparities. We asked whether the disparities encountered in the natural environment match that range. We did this by simultaneously measuring binocular eye position and three-dimensional scene geometry during natural tasks. The natural distribution of disparities is indeed matched to the smaller range of correspondence search. Furthermore, the distribution explains the perception of some ambiguous stereograms. Finally, disparity preferences of macaque cortical neurons are consistent with the natural distribution.

INTRODUCTION

Having two views of the world, as animals with stereopsis do, provides an advantage and a challenge. The advantage is that the differences in the two views can be used to compute very precise depth information about the three-dimensional (3D) scene. The differing viewpoints create binocular disparity, namely, horizontal and vertical shifts between the retinal images generated by the visual scene (Fig. 1). These binocular signals are integrated in the primary visual cortex via cells tuned for the magnitude and direction of disparity (1, 2), and this information is sent to higher visual areas where it is used to compute the 3D geometry of the scene.

The computations involved in stereopsis, however, are very challenging because of the difficulty in determining binocular correspondence: Which points in the two eyes' images came from the same feature in the scene? Consider trying to solve the correspondence problem in a visual environment consisting of sparse small objects uniformly distributed in three dimensions. In every direction, all distances would be equally probable, and therefore, the encountered disparities would be very broadly distributed. However, it is actually more complicated than that: If the viewer fixated a very distant point, all other points in the scene would be closer than where the viewer was fixating, and the disparity distribution would be entirely composed of crossed horizontal disparities (illustrated for one point in Fig. 1, A and B). Similarly, fixation of a very near point would produce a wide distribution of uncrossed horizontal disparities (Fig. 1C). In this environment with fixations of random distance, the search for solutions to binocular correspondence would have to occur over a very large range of disparities.

However, the natural visual environment is not at all like this, and peoples' fixations are not randomly distributed in depth. Instead, the environment contains many extended opaque surfaces, so farther objects are frequently occluded by nearer ones. The environment is also structured by gravity, so many surfaces are earth-horizontal (for example, grounds and floors) or earth-vertical (for example, trees and walls).

Furthermore, viewers do not fixate points randomly, but rather choose behaviorally significant points such as objects they are manipulating and surfaces on which they are walking (3, 4). These constraints are presumably manifested in the brain's search for solutions to binocular correspondence, allowing a much more restrictive and efficient search than would otherwise be needed. In this vein, models of stereopsis invoke constraints based on assumptions about the 3D environment, such as smoothness, uniqueness, and epipolar geometry, to help identify correct correspondence (5). Similarly, probabilistic models make prior assumptions about the most likely disparities in the natural environment. One prior, for example, instantiates the assumption that small disparities are common and large disparities are rare (6). Consistent with this, the preferred disparities of binocular cortical neurons are generally close to zero (1, 2, 7). This type of coding—a bias toward commonly encountered sensory signals—can improve perceptual sensitivity to small changes from likely stimulus values (8). It also naturally leads to reductions in information redundancy, a fundamental principle of efficient coding in sensory and motor systems (9–11).

The study of natural image statistics has yielded substantial advances in the understanding of how the visual system evolved into an efficient information processor (12–15). This work has focused on image statistics such as the distributions of luminance, contrast, and orientation. Less is known about how the visual system exploits natural scene statistics for more complex tasks such as 3D vision because of the technical challenges involved in measuring the relevant statistics (16). Such statistics, however, are particularly important for 3D vision because the visual system cannot measure the third dimension directly from the 2D retinal images. Instead, it must use depth cues that derive from regularities in the relationship between distance and retinal image patterns.

To determine the statistics of natural disparity, one must measure the 3D geometry of visual scenes and binocular eye movements as people engage in natural behavior in those scenes. Some previous studies have attempted this by simulating scenes and eye movements (17–20). Those studies are obviously limited by the ecological validity of the simulated scenes and by the validity of the assumed eye movement strategies. One study went further by measuring fixation distances as people walked through a forest (21). They then used the distribution of those distances to generate simulated fixations in natural scenes recorded by other investigators (22). From there, they calculated the distribution of disparities accumulated across the visual field. They reported that horizontal disparities are centered at zero across the horizontal field of view and that the distribution of these disparities is symmetric with a standard deviation (SD) slightly less than 0.5°. This study and most of the others assumed that the distributions of fixation distances and the distances of visible points in the scene are the same.

Here, we report the properties of binocular eye movements and the resulting binocular visual input directly measured during natural tasks with a unique eye and scene tracking system. We then compare these measurements to known perceptual and neurophysiological properties of stereopsis and consider the implications for efficient coding.

BRIEF METHODS

The eye and scene tracking system simultaneously measured the 3D structure of the visual scene and the gaze directions of the eyes while participants performed everyday tasks. To quantify binocular disparity, we used a retinal coordinate system in which the positions of image points are referenced to the fovea and quantified in angular units. This is appropriate for the study of binocular correspondence because pairs of corresponding points in the two eyes are fixed in retinal coordinates (23) and the properties of binocular neurons in cortical area V1 are invariant when expressed in retinal coordinates (24, 25). We quantify disparities in Helmholtz coordinates (that is, latitudes for azimuth, longitudes for elevation). Consider a point in the world projected into the left and right eyes. Differences in the azimuths of the point images are horizontal disparities; differences in elevation are vertical disparities. These are absolute disparities because the azimuths and elevations of the images in the two eyes are referenced to the fixation point (and hence to the foveas). When fixation changes, the horizontal and vertical disparities created by a given scene can change substantially. One therefore must know scene geometries and fixations within those scenes to understand the diet of binocular disparities experienced by humans and other animals.

The eye and scene tracking system has two outward-facing cameras (Fig. 2A, green box) that capture stereoscopic images of the world in front of the participant. The images are converted to 3D scene geometry using well-established stereo reconstruction techniques. The system also has a binocular eye tracker (orange) that measures binocular fixations. Because the stereo cameras are offset from the participant's eyes, one cannot reconstruct disparities directly from the cameras' 3D maps. To do this, we estimate the translation and rotation of viewpoints from the cameras to the eyes. This is done by displaying a grid of circles with known locations relative to the midpoint of the interocular axis (the cyclopean eye). Mappings between these known 3D locations in cyclopean eye–referenced coordinates

and the corresponding positions of circles in the left camera's images are used to solve for the translation and rotation between the camera and the eyes (Fig. 2B). We then invert this transformation and combine it with the gaze directions measured by the calibrated eye tracker to produce accurate 3D maps and retinal disparities for the participant. Figure 2C outlines the data collection and processing workflow. The system generates a disparity map from the camera images (Fig. 2D) and then reprojects the scenes to each eye, so the images and disparity maps are in eye coordinates (Fig. 2E). The cameras have a limited field of view, so we selected only the parts of the images within 10° of the fovea to minimize the amount of missing data during analysis. From the projection of each point in the world onto each retina, we calculated retinal disparity. Figure 2F shows the horizontal disparities for one participant at one time point. The sign convention for horizontal disparities is that the uncrossed are positive and the crossed are negative. So, when the participant fixates on the sandwich bread (Fig. 2F), points on the table in front of the bread have negative disparity (red) and points behind the bread have positive disparity (blue). For vertical disparities, the convention is positive when elevation is greater in the left than in the right eye, and negative for the opposite.

Three adults participated and performed four tasks. The tasks were selected to represent a broad range of everyday visual experiences. In two tasks, participants navigated a path outside in natural areas ("outside walk") or inside a building ("inside walk"). In another task, involving social interaction, they ordered coffee ("order coffee"). In the fourth task, participants assembled a peanut butter and jelly sandwich ("make sandwich"); this task emphasized near work and has been used in previous eye movement studies (26).

STATISTICS OF BINOCULAR EYE POSITION

We first examined the statistics of binocular eye position as participants engage in natural tasks. It is known that the eyes make binocular-ly coordinated movements to place the point of regard (the fixation point) on the foveas. This is accomplished by disconjugate and conjugate binocular movements, vergence and version, respectively. Vergence and version have vertical and horizontal components. We only looked at the horizontal component of vergence because that is the meaningful component for tracking in depth. Vertical and horizontal components of version movements vary substantially, so we looked at both.

The probability distributions of horizontal vergence are shown in Fig. 3A for each task. We averaged across participants because their distributions were very similar. As one would expect, the distribution of horizontal vergence is quite dependent on the task. Relatively small vergences (that is, far fixation distances) were frequently observed in the two walking tasks, most likely because participants tended to look at least a few steps ahead (27). Larger and more variable vergences were observed in the order coffee and make sandwich tasks as participants looked from objects they were manipulating to other nearby objects. To combine data across tasks, we used a large-population estimate of the proportion of waking time that Americans devote to different activities. The database is the American Time Use Survey (ATUS) (28). We used those activity data to weight our tasks a priori (see Materials and Methods). The ATUS-weighted data for horizontal vergence has a peak at small

vergences and a long tail of larger vergences. The median and SD of the weighted-combination distribution are 3.0° (~114 m) and 2.7°, respectively.

The combined distribution is markedly different from the ones assumed in previous simulations (17–19) and indirect empirical measurements (21). Specifically, our data exhibit more large vergences (near fixations) and greater variance than others do. Read and Cumming (17), for example, assumed a folded Gaussian distribution with a peak at 0° and no negative vergences; this distribution has a central tendency of much smaller vergences (farther fixations) than we observed. Liu and colleagues (21) reported a distribution that is somewhat similar to the one we measured in the outside walk task, but with a much shorter tail; our weighted-combination distribution has a substantially larger median and SD than theirs.

Some researchers have assumed that people choose fixation targets with distances that are statistically the same as the distances of all visible scene points (19, 21). We examined this assumption by comparing the distribution of fixation distances to the distribution of distances of all scene points that are visible in the central 20° of the visual field. The relationship between the two distances varies across tasks (fig. S1), but the overall distribution of actual fixations is significantly biased toward closer points than the distribution of potential fixations (Fig. 3B). Specifically, the median weighted-combination fixation distance is 0.87 diopters (114 cm), whereas the median weighted-combination scene distance is 0.47 diopters (211 cm): that is, they differ by nearly a factor of two. This means that viewers engaged in natural tasks tend to look at scene points that are nearer than the set of all visible points, which contradicts the previous assumption that the distributions of fixation and scene distances are similar to one another (19, 21).

The distributions of horizontal and vertical versions were also similar across participants, but rather task-dependent. Figure 3C shows these as joint probability distributions. The outside walk task yielded a broad, roughly isotropic distribution, whereas the inside walk task yielded a more focused distribution, because participants tended to look straight down the hallway as they walked. The order coffee and make sandwich tasks yielded notably anisotropic distributions; the former spread horizontally and the latter vertically. The weighted combination (Fig. 3D) shows that participants rarely make horizontal or vertical versions greater than ~15°, which is consistent with previous measurements (29, 30). The SDs for the weighted-combination horizontal and vertical versions were 12.1° and 8.0°, respectively. These distributions are reasonably similar to the ones assumed in previous simulations and computational models. For example, Read and Cumming (17) assumed isotropic Gaussian distributions for version with an SD of 10° or 20°. The versions we observed are somewhat concentrated on the cardinal axes, suggesting that viewers tend to look up and down or left and right, but not both simultaneously.

In summary, these results show that the patterns of disconjugate and conjugate eye movements depend significantly on the task and the environment and that the distribution of fixation distances is not the same as the distribution of scene distances because people tend to look at near visible objects.

STATISTICS OF NATURAL DISPARITY

We next examined the distributions of horizontal and vertical disparities for each region of the central visual field. We show these data in two ways: In Fig. 4, we show the median horizontal and vertical disparities throughout the visual field, and in Fig. 5, we show the joint and marginal distributions of horizontal and vertical disparities in different field sectors. The tasks are schematized in Fig. 4A. Figure 4B shows the median horizontal disparities for each task. The patterns of disparity were remarkably similar across participants (fig. S2), so we averaged their data. The panels are plotted in retinal coordinates with the fovea at the center. The patterns of horizontal disparity differ noticeably across tasks. For example, in the make sandwich task, in which participants generally adopted short fixation distances, there is much more variation in median disparity across the visual field than in the other tasks, where fixations were gen erally more distant. In the inside walk task, disparities tended to be negative (crossed), but in the order coffee task, disparities tended to be positive (uncrossed). Figure 4C shows the median vertical disparities for each task, which overall are much smaller in magnitude than the horizontal disparities. The left and right panels of Fig. 4D show the median of the weighted-combination horizontal and vertical disparities across the visual field, respectively. To summarize the data, Fig. 5A shows the joint and marginal distributions of horizontal and vertical disparities for nine subregions of the visual field (for the weighted combination only), and Fig. 5B shows these distributions when they are combined over the entire central 20° of the visual field.

Some clear patterns emerge from these data. With respect to horizontal disparities:

1)   The variation of disparity as a function of elevation is much greater than the variation as a function as azimuth. Specifically, there is a consistent positive vertical gradient of horizontal disparity, whereas the horizontal gradient is ~0. This is because crossed disparities are most common in the lower visual field, and uncrossed disparities in the upper field. The pattern is due, in large part, to participants fixating the ground or objects on the ground while walking, and fixating objects on tabletops while manipulating things with their hands. A previous simulation study suggested that the presence of a ground plane in and of itself contributes to this pattern (18).

2)   Horizontal disparity is ~0° near the eyes' horizontal meridians. This means that points to the left and right of fixation tend to be at roughly the same distance as the fixated point. This observation is consistent with a previous study (21).

3)   The SD of horizontal disparity increased slightly with retinal eccentricity (Fig. 5A and fig. S2C). The overall dispersal was greatest in the tasks that contained more near fixations (order coffee and make sandwich). The SD overall was larger than reported by one study (21) and smaller than found in a previous simulation (17).

4)   The distributions of horizontal disparity generally exhibited positive skewness (Fig. 5 and fig. S2E), meaning that there was a longer tail of positive (uncrossed) disparity than of negative (crossed) disparity. This observation is somewhat surprising because there is a geometric reason to expect negative

skewness. The largest uncrossed disparity that one can conceivably observe for a fixation distance of Z_0 is (31):

$$\delta \approx \frac{I}{Z_0} \left(\frac{180}{\pi} \right)$$

where δ is disparity in degrees, I is interocular distance in meters [assumed to be 0.0625 m; (32)], and Z_0 is fixation distance in meters. Thus, when the eyes are fixating at 1, 10, and 100 m, the largest possible uncrossed disparity is, respectively, 3.6, 0.36, and 0.036°. Despite this geometric constraint on large uncrossed disparities, we still observed positive skewness. This is a consequence of people tending to adopt near fixations (keeping Z_0 relatively small) and tending to fixate nearer points in the scene (Fig. 3B). Our observation of positive skewness contradicts earlier reports of symmetry (6, 21, 33) or negative skewness (17, 19).

5)   The kurtosis of the horizontal disparity distributions was markedly high, meaning that there was a high probability of observing disparities close to the peak value with long tails of lower probability. For the weighted combination, kurtosis was 17 to 50, depending on location in the visual field (Fig. 5A and fig. S2G). High kurtosis and low SD mean that horizontal disparities at a particular location in the visual field are quite predictable. For example, to encompass two-thirds of the disparity observations in 3° patches in various positions in the visual field requires a disparity range of only 0.74° on average (see Fig. 5 for details). Previous work observed or assumed much lower kurtosis of less than 0 (6), 0 (17), ~3 (19, 33), and 5.6 (21) and slightly smaller SDs (6, 19, 21), which in combination implies lower predictability than we observed.

With respect to vertical disparities:

1)   Positive vertical disparities were consistently observed in the upper left and lower right visual fields, and negative disparities in the upper right and lower left fields, consistent with expectations from the underlying binocular viewing geometry (34, 35).

2)   The magnitude of variation differed substantially across tasks: There was much greater variation in the two near tasks of order coffee and make sandwich.

3)   Vertical disparities are generally zero near the retinal vertical meridians. When eyes are vertically aligned, the projections of points in the head's median plane yield zero vertical disparity. When eyes are fixating in or near this plane, zero vertical disparities are observed near the vertical meridians in retinal coordinates. The band of zero disparities is shifted rightward for two participants in the make sandwich and order coffee tasks because they often looked left of the head's median plane, so the projection of that plane was shifted rightward in retinal coordinates. Vertical disparities are also almost always zero on the horizontal meridians, which is a consequence of the torsional eye movements that we assumed are made (see Materials and Methods).

4)   The SD of vertical disparity varied systematically across the visual field (Fig. 5A and fig. S2D). It was close to zero near the horizontal meridian and increased roughly linearly with increasing eccentricity above and below that meridian. This pattern is expected from the underlying geometry (34, 35). When the eyes are torsionally aligned, vertical disparity along the horizontal meridian must be zero. We did not observe a vertical band of small SD because as participants looked leftward and rightward (Fig. 3C), that band shifted rightward and leftward in retinal coordinates, thereby adding variation for positions right and left of the vertical meridian.

Lastly, the joint distribution of horizontal and vertical disparities has an aspect ratio of ~20:1 (Fig. 5B; horizontal disparity more spread than vertical). This makes sense because the eyes are horizontally separated, so when they are directed to a fixation point and aligned torsionally, non-fixated points in natural scenes should create a larger range of horizontal disparities than of vertical disparities (1, 17, 36). The joint distribution is diamond-shaped because the marginal distributions of horizontal and vertical disparities are leptokurtic and nearly independent.

The observed distributions of horizontal and vertical disparities are very different from the ones derived from simulations of scenes and fixations (17, 19, 21). In particular, the observed distributions of horizontal disparity have high kurtosis and positive skewness unlike the distributions in previous work (6, 17, 19, 21, 33). They also have different SDs than in previous accounts (6, 17, 19, 21, 33). The observed distributions of vertical disparity have much smaller SDs than those of the distributions derived from one previous study (17) and larger than another (19).

The empirical joint distribution is an estimate of the prior distribution of horizontal and vertical disparities in natural viewing and can be incorporated into probabilistic models of stereopsis. We next examined how these natural disparity statistics help us understand some well-known phenomena in binocular vision: (i) positions of corresponding points and shape of the horopter, (ii) perceived depth and binocular eye movements with ambiguous stereograms, and (iii) disparity preferences among cortical neurons.

CORRESPONDING POINTS AND THE HOROPTER

The visual system allocates resources nonuniformly over the span of possible disparities. As a result, the quality of binocular vision varies substantially with relative position in the retinas. The nonuniform allocation is manifest by pairs of retinal points with special status. These corresponding-point pairs are special because (i) binocular matching solutions between the two eyes' images are biased toward them (37–39); (ii) the region of single, fused binocular vision straddles them (40, 41); and (iii) the precision of stereopsis is greatest for locations in space that project to those points (42–46). The horopter is the set of locations in 3D space that project onto corresponding retinal points and therefore comprise the locations of finest stereopsis (47, 48).

If corresponding points were in the same anatomical locations in the two eyes (that is, zero horizontal and vertical disparities), the horopter would be a circle containing the fixation

point and the centers of the two eyes (the Vieth-Müller Circle) and a vertical line in the head's median plane (Fig. 6A). Empirically measured corresponding points do not have zero disparity. Instead, they have uncrossed (far) horizontal disparities above and to the left and right of fixation, and crossed (near) disparities below fixation (23, 31, 35, 41, 48–51). Thus, the empirical horizontal horopter is less concave than the Vieth-Müller Circle and the empirical vertical horopter is convex and pitched top-back relative to the vertical line. The positions of corresponding points (and the shape of the horopter) are a manifestation of the visual system's allocation of resources for finding binocular correspondence matches. That is, they are the bets the system places on the most likely disparities.

Helmholtz (48) and others hypothesized that the horopter is specialized for placing the region of most precise stereopsis in the ground plane as a standing observer fixates various points on the path ahead (35, 49, 51). However, the horopter cannot be described by one 3D shape in space because its shape changes with changes in fixation. For example, when fixation distance increases, the horizontal horopter changes from concave to planar to convex (35, 41) and the vertical horopter becomes more convex (31) (Fig. 6B). The vertical horopter also becomes more pitched with increasing fixation distance, a property that is by itself consistent with Helmholtz's hypothesis (35). Collectively, these observations mean that the horopter does not in general coincide with the ground plane because it is quite convex at fixation distances greater than 1 m. It seems that the region of most precise stereopsis is adapted for some other visual requirement. We hypothesized that the disparities that define the horopter reflect the most likely disparities seen during a wide range of natural behaviors, rather than simply fixating the ground. For example, convexity may be a likely property of visual experience because many objects are closed and therefore convex on average.

With the distributions of horizontal disparity throughout the visual field in hand (Figs. 4 and 5), we zeroed in on the values along the vertical and horizontal meridians to see if they are consistent with the disparities of the corresponding points. Not surprisingly, the disparities near the meridians vary substantially across tasks. For example, the disparities created when walking through an outdoor scene are qualitatively similar to the disparities associated with the vertical and horizontal horopters (uncrossed upward, leftward, and rightward, and crossed downward), whereas the disparities from inside walking are somewhat dissimilar. Figure 6C shows the 25th, 50th, and 75th percentiles for the disparities at each vertical and horizontal eccentricity. However, the most appropriate comparison is between the disparities of the horopter and the observed natural disparities once weighted according to estimated frequency of occurrence. This is shown in Fig. 6D. The weighted-combination natural disparities are shown in brown. Horopter data from several previous studies are shown in yellow. Both the natural disparity and horopter data along the vertical meridian (left panel) are quite asymmetric: crossed disparity below the fovea and uncrossed above. Both also exhibit convexity (31). The natural disparity and horopter data along the horizontal meridian (right panel) are symmetric with a tendency toward uncrossed disparity. The weighted-combination distribution sometimes has two peaks, usually one near zero disparity and one at a greater crossed or uncrossed disparity (movies S1 and S2). This property is not captured by the percentiles alone, so we also plotted the full distribution as a frequency map along with the percentiles in Fig. 6D. The central tendency of the distribution of natural disparities closely follows the horopter data.

Thus, despite the variation of natural disparities across tasks, the overall agreement is good: The disparities of the horopter tend to have the same sign and similar magnitude as the natural disparities. Recall that the task weighting was done before the data were analyzed, so the agreement between the horopter and natural disparity statistics is not a consequence of adjusting weights to maximize fit. Therefore, the natural disparity statistics near the horizontal and vertical meridians of the visual field are consistent with the positions of corresponding retinal points. This means that the visual system does in fact allocate its limited binocular resources in a way that exploits the distribution of natural disparities and thereby makes depth estimates most precise in the regions in space where surfaces are most likely to be observed.

NATURAL DISPARITY STATISTICS AND PERCEIVED DEPTH FROM DISPARITY

It would be rational for the visual system to incorporate the prior distribution of natural disparities when constructing depth percepts from noisy or ambiguous binocular inputs. In particular, use of an appropriate prior would increase the odds of inferring the true 3D structure of the scene. We examined this by developing a probabilistic model, that incorporates the measured natural disparity statistics to infer depth from ambiguous stereoscopic stimuli.

When identical periodic stimuli are presented to the two eyes (Fig. 7A, upper row), the wallpaper illusion occurs. The viewer perceives a single plane, and the apparent depth of that plane is closely associated with the distance to which the eyes are converged (37, 53). The phenomenon is called an illusion because the stimulus provides multiple correlation solutions, but only one is seen. The upper left panel of Fig. 7B shows that local cross-correlation of the images yields equally high-correlation peaks for numerous horizontal disparities (39). The same behavior is observed in disparity-selective cortical neurons: When presented with periodic stimuli, V1 neurons exhibit multiple response peaks (54). Therefore, if the disparity estimates obtained by local correlation or V1 responses were the sole determinant of perceived depth, the wallpaper stimulus should look like multiple parallel planes stacked in depth. The fact that the stimulus looks like one plane at the same depth as the fixation point means that more than inter-ocular correlation or responses of individual V1 neurons are involved in constructing the final percept.

When the spatial frequencies of periodic stimuli differ in the two eyes (Fig. 7A, lower row), a different illusion called the venetian blinds effect occurs. Although cross-correlating the images again yields numerous peaks (Fig. 7B, lower left), one sees a series of small slanted planes [vertical blinds; (55–57)]. If the disparity estimates obtained by correlation were the sole determinant of perceived depth, the venetian blinds stimulus would appear to have multiple slanted planes stacked in depth. Again, this is not what one perceives.

These effects have been attributed to smoothness constraints embodied in the neural computation of disparity (5, 58). However, the percepts with venetian blinds stimuli are definitely not smooth (39), which suggests that another principle is involved. We hypothesized that prior assumptions about the most likely disparities significantly influence

the percepts in these illusions. To test this hypothesis, we formulated the problem of disparity estimation from binocular stimuli as a Bayesian inference problem. Our model (see Materials and Methods) multiplies the stereo matching likelihood and the prior on disparity statistics to generate a posterior distribution of disparity given the binocular stimuli (Fig. 7B). For visualization purposes, we set elevation to zero and display 2D cross sections of these distributions. The stereo matching likelihood is a sigmoidal function of the normalized local cross-correlation between the left and right images. The finite size of receptive fields is taken into account by windowing the stimuli with a Gaussian (59). The likelihood, based on the cross-correlation of the images, yields numerous peaks for both types of binocular images. Therefore, the maximum of the likelihood yields an ambiguous solution. The prior distribution is a kernel-density fitted version of the weighted-combination distribution in Fig. 5B. Estimated disparity is inferred as the maximum of the posterior [that is, maximum a posteriori (MAP) solution]. This MAP solution (Fig. 7C) gives a disparity estimate that matches the perceived solution for both effects, that is, a plane at the same distance as the fixation point for the wallpaper stimulus (upper panel) and a row of slanted surfaces for the venetian blinds stimulus (lower panel).

The natural distribution of disparity can also help explain binocular eye movements made to fuse ambiguous stereoscopic stimuli. When a stimulus with nonzero disparity is presented, binocular eye movements are made to eliminate the disparity and thereby align the images in the two eyes. The movements are combinations of horizontal and vertical vergence. An interesting problem arises when the stimulus is one-dimensional, such as an extended edge. The disparity component that is parallel to the edge has no effect on the retinal images and hence cannot be measured. As a consequence, the disparity direction and magnitude cannot be determined. This is the stereo aperture problem (60). Disparity energy units in area V1 are subject to this ambiguity because they see only a local region of an edge or grating (54). When presented a disparate edge, the oculomotor system could, in principle, align the two images by executing a vergence movement in one of many directions. Rambold and Miles (61) showed that the direction of vergence response is heavily biased toward horizontal, a bias that is not observed when the stimulus is a random dot stereogram.

In the motion equivalent to the stereo aperture problem, the ambiguous motion of an edge is perceived unambiguously as moving orthogonal to the edge. This is well modeled as a computation involving the observed motion and a prior for slow speed (62, 63). The motion prior is isotropic, that is, spread equally for all motion directions. We observed that the disparity prior is very anisotropic (Fig. 5): the spread for horizontal disparity is ~20 times greater than for vertical disparity. Application of this horizontally elongated prior to the stereo aperture stimulus would predict that the horizontal solutions will be strongly favored, as Rambold and Miles observed for vergence eye movements (61).

We conclude that the prior distribution of naturally occurring disparities reported here determines to a large degree the manner in which ambiguous stereoscopic stimuli (that is, those yielding multiple solutions with equally supported local correlation between the two eyes' images) are reduced to one percept or to one signal that drives vergence eye movements.

DISPARITY PREFERENCES IN THE VISUAL CORTEX

Information theoretic analyses predict that efficient neuronal populations will have more neurons tuned for more likely inputs (8, 64). Following this reasoning, one would expect disparity preferences among cortical neurons to reflect the statistics of natural disparity. Most binocular cortical neurons respond preferentially to a limited range of disparity. There is a consensus that most are tuned to crossed disparities (that is, points gen erally closer than fixation). This has been reported for areas V1 (1, 2, 65–67), V2 (65, 68), V3 (69), V4 (70–72), MT (73–75), and MST (76). Some have argued that the overall preference for crossed disparity reflects the behavioral significance of objects near the viewer or that the preference is somehow related to figure-ground segregation in natural vision (70). However, on the basis of our measured distributions of natural disparities, we propose an alternative explanation. If disparity-tuned neurons are distributed efficiently, we predict that preferences should only exhibit the bias toward crossed disparity in the lower visual field, and a bias toward uncrossed disparity in the upper field (Figs. 4 to 6).

To examine this hypothesis, we computed the correlation between preferred disparities and receptive-field locations of V1 neurons in five single-unit studies (1, 2, 65–67). These investigators recorded preferred horizontal disparity and receptive-field position for 820 cells in macaque V1. They also reported preferred vertical disparity in 188 cells. The groups used different experimental setups and coordinate systems (Materials and Methods and table S1), so we first took into account the geometry of their setups and converted their data into the Helmholtz coordinate system that we used to quantify the natural disparity statistics. We assumed an interocular distance of 3 cm for macaques (2).

Receptive-field locations were not uniformly sampled across the visual field (Fig. 8A). Eighty-five percent of the sampled neurons had receptive fields in the lower field. This sampling bias is undoubtedly due to the fact that the upper field is represented in ventral V1, which is more difficult to access than dorsal V1.

Our analysis confirmed the oft-reported bias toward crossed horizontal disparity: Of the 816 neurons with nonzero disparity preferences, 55 and 45% had crossed and uncrossed preferences, respectively. However, this bias depends on receptive-field location. Figure 8B shows that neurons representing the upper visual field actually tended to prefer uncrossed horizontal disparity (mean preferred disparity is +0.14°), whereas neurons representing the lower field preferred crossed disparity (mean preferred disparity is −0.09°). The difference between the two distributions is statistically significant (Mann-Whitney U test, P = 0.007). This upper/uncrossed, lower/crossed pattern is qualitatively similar to the natural disparity statistics that we measured in humans (Figs. 4 to 6) despite the species' differences in height and eye separation. We also computed the distributions of preferred horizontal disparity for the left and right visual fields and observed no systematic difference (P = 0.3), which is also consistent with natural disparity statistics.

We also examined preferred vertical disparity as a function of field location. From the natural disparity statistics (Figs. 4 and 5), one would predict a bias toward positive vertical disparity in the upper left and lower right quadrants and a bias toward negative disparity in

the other quadrants (Fig. 8C). Although there is a tendency toward positive vertical disparity in the upper left and lower right quadrants, the difference in preferred disparity among the sampled neurons is not significantly different from the lower left and upper right quadrants ($P = 0.2$). The lack of significance may be because so few neurons were characterized in the upper right and lower left quadrants [34 (18%) and 5 (3%), respectively].

Figure 8D plots the joint and marginal distributions of preferred horizontal and vertical disparities for the 158 neurons for which both were measured. The joint distribution, like the natural disparity statistics, is quite anisotropic with greater spread for horizontal disparity (1). However, the anisotropy is not nearly as great as we observed in the statistics (Fig. 5B). The greater spread of preferred vertical disparity may be a consequence of the need to measure large vertical disparities to guide vertical vergence eye movements or of measurement error in the single-unit studies.

We conclude that horizontal disparity preferences in macaque V1 are biased toward crossed (near) disparities in the lower visual field and un-crossed (far) disparities in the upper field. Therefore, the oft-cited bias toward crossed disparity does not generalize to the whole visual field; it is instead caused by oversampling of the lower field. The reports of a crossed disparity bias in areas V2, V3, V4, MT, and MST (65, 68–76) may also be due to biased sampling of the visual field: In every study in which the anatomical or field positions of the samples were reported (65, 68–73, 76), all or most were from the lower visual field.

The elevation-dependent change in disparity preferences among macaque V1 neurons is compatible with the distribution of horizontal disparity encountered by humans engaged in natural tasks and therefore may provide an efficient means of allocating neural resources for the computation of depth.

CONCLUSION

We find that the distribution of natural disparities is consistent with (i) the positions of corresponding points across the visual field and therefore with the regions in space where stereopsis is most precise, (ii) the ways in which ambiguous stereograms are perceived and drive vergence eye movements, and (iii) disparity preferences across the visual field in macaque visual cortex. A similar relationship between disparity statistics and allocation of neural resources is implied by observations in other species. In domestic cats and burrowing owls, for example, receptive-field positions are offset systematically in the two eyes: offsets of crossed disparity above fixation and uncrossed disparity below fixation (77). Under the plausible assumption that these interocular offsets are a manifestation of offsets in corresponding points, the crossed-below, uncrossed-above pattern is similar to the one observed in humans (Fig. 6). The offsets in cats and owls are much larger than those in humans, consistent with the idea that they reflect the most common disparities encountered by animals whose eyes are close to the ground (77).

An important question remains: how do these neural processes come to reflect the statistical regularities? By learning? Or are they hardwired as a result of evolution? The positions of corresponding points in humans do not change with 1 week of exposure to altered disparity

distributions (31), which is inconsistent with learning over that period; perhaps learning occurs over longer periods. Consistent with long-term learning, preferred orientation among binocular neurons in cat visual cortexes can be altered by extensive early visual experience. When the images delivered to the two eyes are rotated in opposite directions, cortical neurons develop orientation preferences that differ in the two eyes by corresponding amounts (78). However, the positions of corresponding points are presumably reflected by changes in receptive-field locations, not orientation preferences. Thus, it remains to be determined how the striking relationships between binocular visual experience and binocular neural processes come to be. By whatever means they materialize, the end product is well adapted to the regularities of the 3D environment.

MATERIALS AND METHODS

Participants and procedure

Three male adults (21 to 27 years old) participated in the study. The human subjects protocol was approved by the Institutional Review Board at the University of California, Berkeley, and all participants gave informed consent before starting the experiment. The participants were tested for normal stereopsis using a Titmus test and were screened with optometric measurements for normal binocular eye movements and fixation disparity. In separate experimental sessions, participants performed four different tasks while wearing a custom mobile tracking system described below. These tasks were as follows: (i) walk a path inside a building (inside walk), (ii) walk a series of paths outside on the Berkeley campus (outside walk), (iii) go to a café and place an order (order coffee), and (iv) make a peanut butter and jelly sandwich (make sandwich). For the two walking tasks, participants were shown the paths ahead of time by walking them once while guided by the experimenter. For the outside walk task, the participants walked in a natural wooded area, a natural open area, and along manmade footpaths on campus.

After each session, the calibration measurements were used to assess the integrity of the data before further analysis. Once all sessions were completed for each participant and each activity, the data from the sessions were trimmed to 3-min clips. The clips were selected by determining the frame at which the desired task was initiated (for example, upon entering the campus park during outside walk, upon entering the hallway during inside walk) and then taking three continuous minutes of subsequent activity. Because participants walked in a variety of outdoor areas, the outside walk was a total of 3 min randomly selected from the different environments. This resulted in 12 clips overall (three participants, four tasks).

Notation

We use the following notation repeatedly. Each term represents a scalar value in units of visual angle (in degrees): δ, binocular disparity; β, azimuth latitude; η, elevation longitude.

Subscripts R, L, and C indicate that the quantity is for the left, right, or cyclopean eye, respectively. β values are positive in the left visual field, and h values are positive in the upper field.

System specifications

The eye tracking component of the apparatus was a modified EyeLink II system (SR Research). The EyeLink II is a head-mounted, video-based binocular eye tracking system. The eye tracker captures a video of both eyes using two infrared (IR) cameras on adjustable arms attached to a padded headband (Fig. 2A). The headband contains two IR lights that illuminate the pupil and sclera. They also create reflections off the corneas, and those reflections are tracked along with the pupil center to estimate gaze direction for each eye at 250 Hz. The manufacturer reports a root mean square (RMS) error of 0.022° and a field of view of 40° horizontally and 36° vertically. The EyeLink Host PC was replaced with a custom mobile host computer that was used to acquire and store the eye tracking data. This machine had an Asus AT3GC-I motherboard and a 32-GB solid- state hard drive (Corsair CSSD-V32GB2-BRKT) inside a mini-ITX enclosure. It was powered by an external Novuscell universal lithium battery pack. The display computer was replaced with a Lenovo ThinkPad X220 laptop that was connected to the mobile host computer via ethernet cable. During calibration, this laptop was connected via VGA (video graphics array) to a 121×68– cm LG display (model no. 55LW6500-UA). This laptop drove the display with specific patterns for calibration, and acquired and stored the stereo rig images. Both computers were stored in a backpack and ran on battery power to allow participants full mobility while performing the experimental tasks.

The stereo cameras were two Sony XCD-MV6 digital video cameras ($29 \times 29 \times 19$ mm; 1.3 oz). Each camera contained a 640×480 monochrome CMOS (complementary metal oxide semiconductor) sensor. The cameras had fixed focal length (3.5 mm) Kowa lenses with a field of view of 69° horizontally and 54° vertically. Focus was set at infinity. Manufacturer-reported resolution for the lenses is 100 line pairs per millimeter (lp/mm) at the center and 60 lp/mm in the corners, with a barrel distortion of −28%. The resulting resolution was 1.5 lines per pixel (l/p) centrally and 0.8 l/p in the corner. Frame captures were triggered in a master-slave relationship, with the master camera triggered at 30 Hz by the display laptop using a programmable trigger via a Koutech 3-Port FireWire (IEEE 1394b) ExpressCard. Once triggered, the master camera sent the signal to the slave camera and to the host computer for the EyeLink. This was done via a custom cable connected to the digital I/O (input/output) interfaces of the two cameras and to the parallel port of the host computer. This input signal was recorded on the host computer in the EyeLink Data File for temporal synchronization of eye tracking and camera data. Image files were transferred from the cameras to the laptop via FireWire. The two cameras were attached to a custom aluminum mount with a 6.5-cm fixed baseline. The mount was affixed just above the IR lights on the eye tracker. The optical axes of the cameras were parallel and pitched 10° downward to maximize the field of view that was common to the tracker and cameras.

All system calibration and 3D scene reconstruction were performed using Python 2.7.6, with display and event loops driven from the Pygame library (v1.9.1; pygame.org), the EyeLink bindings (Pylink v1.0.0.37; pypi. python.org/pypi/PyLink), and the Open Source Computer Vision module (OpenCV v2.3.3; opencv.org).

Stereo rig

Calibration—The intrinsic parameters of the individual cameras and their relative positions in the stereo rig were estimated before each experimental session, using camera calibration routines from OpenCV. The cameras were securely fixed to the rig, and the intrinsic parameters were in principle known. Calibration was performed before each session to correct small deviations that could occur when using the apparatus. In the calibration, images of a printout of the OpenCV circle grid pattern (11×4 circles) were captured on 8.5 $\times$ 11–inch paper, covering a range of distances and orientations across each camera's field of view. Between 100 and 200 frames were used for each calibration. These frames were processed offline by first identifying the projections of the circle grid in each image (findCirclesGrid-Default) and then using the known grid geometry to estimate the focal length, focal center, and four lens distortion coefficients that minimized the reprojection error to each camera using the global Levenberg-Marquardt optimization algorithm (calibrateCamera2). The rotation and translation between the two cameras were then estimated, and the intrinsic estimates were further refined by applying the same principle and method to minimize the reprojection error for both cameras simultaneously (stereoCalibrate). Experimental sessions for which the above parameter estimates resulted in an RMS error of predicted versus actual pixel positions of the circle grid greater than 0.3 pixels were discarded because errors larger than that resulted in poor 3D reconstructions. During the subsequent eye tracker calibration and for the duration of the experimental session, the stereo rig captured and stored synchronized stereo pairs at 30 Hz.

3D scene reconstruction—Once a session was completed, the intrinsic and extrinsic parameters of the stereo rig were used to calculate the rectification transforms and to remap the pixels in the stereo image pairs for each frame into rectified images with parallel epipolar lines (stereoRectify, InitUndistortRectifyMap, and remap). Disparities between each pair of images were estimated using a modified implementation of the Semi-Global Block Matching algorithm from OpenCV (stereoSGBM and stereoSGBM.compute) (79, 80). This method of disparity estimation finds the disparity that minimizes a cost function on intensity differences over a region ("block") surrounding each pixel in the stereo pair. The blocks are treated as continuous functions, and thus, subpixel disparities are possible. Speckle filtering was disabled and the prefilter cap was set to zero to minimize regularizations based on the assumptions made about disparity patterns. Smoothness parameters were enforced in the cost function, so preference was given for solutions that resulted in small differences in disparity between nearby pixels. This minimum disparity was set to zero, and the maximum disparity was set to 496, which corresponds to points 12 cm from the cameras or closer. Disparity estimation was performed on both the left and right camera images, and pixels with a difference in disparity greater than two pixels or for which the best-guess disparity was less than 10% better than the second best were discarded. Stereo matching was initially performed with block sizes of 3×3 pixels. It was repeated with block sizes of 5×5 and 17×17, and discarded or missing disparity values were filled in if valid disparities were present at coarser scales. False matches were reduced by removing larger disparities from the finer-scale maps that were not at all present in the coarser maps. Final disparity maps were converted to 3D coordinates using the disparity-to-depth output matrix from stereoRectify (reprojectImageTo3D). The resulting matrices contained the grayscale

intensity and 3D coordinates of each point in the scene in a coordinate system with the left camera of the stereo rig at the origin.

Eye tracker

Calibration—The EyeLink II eye tracker was calibrated using the standard SR Research Tracker Application software. Once the stereo rig calibration described above was completed, we secured the eye tracker and stereo rig to the participant's head. To calibrate the tracker, the participant fixated nine small targets presented sequentially on a screen in front of them. These fixations were then used to map the eye data (location of pupil center and corneal reflection relative to the center) into a direction of gaze for each eye in a head-referenced coordinate system with arbitrary units (*HREF*). To convert *HREF* coordinates into a real-world coordinate system, the participant performed the calibration while biting on a custom bite bar. The vertical axis of the bite bar was on the horizontal midpoint between the rotational centers of the eyes (cyclopean eye) 100 cm in front of the screen's midpoint (22). The location of the cyclopean eye relative to the screen was therefore known, and this was used to translate and scale *HREF* coordinates into real-world, cyclopean eye–referenced coordinates. This transformation was applied to the gaze coordinates for each eye. Gaze coordinates at each time point were the intersection of the eye's gaze direction onto a plane 100 cm in front of the cyclopean eye. To do this conversion, the head tracking functionality of the EyeLink II had to be disabled.

3D fixation estimation—To estimate the 3D coordinates of fixation, a second point along each eye's direction of gaze was needed. Those points were the eyes' rotational centers. Because the cyclopean eye is the midpoint between the two rotation centers, the points could be calculated by measuring the interocular distance for each participant. For each time point, the 3D fixation point was thus the intersection of the two vectors specified by the rotational centers and gaze coordinates. Because of measurement noise in the eye tracker, the two vectors did not necessarily intersect. To ensure intersection, we projected the two vectors at each time point onto a common epipolar plane positioned midway between the two vectors and intersecting the eye's rotational centers. The intersection in this plane yielded a 3D fixation point in head-centered coordinates with its origin at the participant's cyclopean eye. We then calculated the version required to fixate this point.

Because the distances of the fixation points are much greater than the interocular distance, estimates of binocular gaze are more sensitive to eye tracker error for vergence than for version. Our participants all had normal binocular eye movements, so we assumed that they correctly converged on surfaces at the center of gaze. That is, we used the 3D scene data to calculate the vergence required to fixate the surface point along the line of sight specified by each version estimate.

Ocular torsion estimation—Listing's law specifies the 3D orientation of the eyes (that is, vertical, horizontal, and torsional) for different gaze directions (47). According to Listing's law, the final orientation of an eye is determined by rotation from primary gaze position (essentially straight ahead at long distance) about an axis in Listing's plane (roughly the plane of the forehead). When the eyes are converged to a near distance, Listing's law is

not obeyed precisely, and the deviation is expressed by Listing's extended law. In the extended law, there are two different planes of rotation from primary position, one for each eye (81). Listing's extended law is not followed exactly by most human viewers. The deviation from that law is quantified by a gain from 0 to 1, where 0 means that the rotation planes are coplanar as specified by Listing's law, and 1 would mean that for every increase in convergence by 1°, the rotation planes for each eye would rotate by ±1°. We modeled ocular torsion by applying Listing's extended law with a gain of 0.8 to each estimate of eye position. The gain of 0.8 is typical for adults with normal binocular vision (82). According to the extended law:

$$\begin{aligned} \psi_{\mathrm{L}} &= -GH_\Delta \frac{\eta_{\mathrm{L}}}{4} - \beta_{\mathrm{L}} \frac{\eta_{\mathrm{L}}}{2} \\ \psi_{\mathrm{R}} &= -GH_\Delta \frac{\eta_{\mathrm{R}}}{4} - \beta_{\mathrm{R}} \frac{\eta_{\mathrm{R}}}{2} \end{aligned}$$

where ψ_{L} and ψ_{R} are torsion for the left and right eyes, respectively; H is the horizontal vergence angle; β is the Helmholtz azimuth of each eye; η is the Helmholtz elevation of each eye; and G is the gain (set to 0.8).

Registration of 3D scene to fixation

Our goal was to generate retinal images of the viewed scene for each period. To do this, we needed to project the scenes captured by the stereo rig into each eye. To register these coordinate systems, the participant was positioned on the bite bar precisely at three distances from the screen: 50, 100, and 450 cm. At each distance, the OpenCV circle grid pattern was displayed and an image was captured with the left stereo camera. These images were used offline to estimate the rotation and translation between the stereo rig centers of projection and the head-referenced coordinate system using OpenCV routines. The projection of the circle grid in each image was identified (findCirclesGridDefault), and then the pose of the camera in the head-referenced coordinate system was estimated (solvePnP). The estimated pose was the rotation and translation that minimized the reprojection error of the 3D coordinates of the circle grid (in head-referenced coordinates) and their imaged locations on the left camera's sensor. The average and SD reprojection errors were 3.75 and 1.68 pixels, respectively.

Measurement of bias and precision

We measured system error immediately before and after each participant performed a task. While positioned on the bite bar at three screen distances (50, 100, and 450 cm), participants fixated a series of targets in a manner similar to the standard EyeLink calibration. At 50 and 100 cm, the targets were a radial pattern of points at eccentricities of 3°, 6°, 12°, and 16° along axes to the left, right, up, down, right and up, and left and down. At 450 cm, the eccentricities were 3° and 6° because greater eccentricities could not be created with the display screen at that distance. Each target pattern contained a letter "E" of size 0.25° surrounded by a circle of diameter 0.75°. We instructed participants to fixate the center of the "E" and focus sufficiently accurately to identify the "E." The participant indicated with a button press when they were fixating the pattern, and eye tracking data were gathered for the next 500 ms. At the same time, the stereo cameras captured images of each target pattern.

We measured the total error of the reconstruction by comparing our estimate of where the participant was fixating with the expected location of each "E." The median error magnitude was 0.29° across all participants, distances, and eccentricities. Ninety-five percent of the errors were less than 0.82° (fig. S3). The minimum resolvable disparity is limited largely by the precision of the depth estimates from the cameras and depends on the interpupillary distance of the participant.

Weighting data from different experimental activities

We used data from the ATUS to weigh the four activities in our experiment according to the time most people spend on those activities. The ATUS data were obtained from the U.S. Bureau of Labor Statistics (27). ATUS provides data from a large and broad sample of the U.S. population. It indicates how people spend their time on an average day. We used Table A-1, which includes survey results for noninstitutional civilians who are 15 years or older. Data from years 2003 through 2012 were included in this analysis. These data are broken down into the total time spent in 12 primary activities. The primary activities are further categorized into 45 secondary activities. We assigned a set of weights to each secondary activity corresponding to the four tasks in our experiment (time asleep was given zero weight). It is important to note that those weight assignments were done before we analyzed our eye tracking and scene data. Thus, there were no free parameters in weight assignment. The weights were summed and then normalized so that they added to one. The resulting values represented an estimate of the percentage of awake time that an average person spends doing activities similar to each of our four tasks. We used these values to compute weighted-combination distributions of disparity, vergence angles, and so forth. The resulting weights for our tasks were as follows: 0.16 for outside walk, 0.10 for inside walk, 0.53 for order coffee, and 0.21 for make sandwich.

Horopter analysis

For this analysis, we used data points within ±2° of the vertical median and within ±2° of the horizontal meridian. Horizontal disparity percentiles were computed at each eccentricity in steps of 0.1°. For the weighted-combination data, we also computed a density distribution of disparities for each eccentricity with a Gaussian kernel of bandwidth 4.5 arc min. These were normalized by the maximum density at that eccentricity. Thus, the heat maps in Fig. 6D illustrate the relative frequency of disparities at each eccentricity rather than the frequency across all eccentricities.

For comparison with the natural disparities data along the meridians, we also replotted the disparities of the horopter from several previous perceptual studies (Fig. 6D). For the vertical horopter (30), the disparities of the horopter over ±8° of elevation were converted into Helmholtz retinal coordinates using the conversion equations for a Wheatstone haploscope (after the removal of fixation and cyclovergence misalignments). For the horizontal horopter (22, 34, 41, 51), the disparities of the horopter were computed from the Hering-Hillebrand deviation value (H) reported for each observer:

$$\delta = \beta_{\mathrm{L}} - tan^{-1}\left(\frac{tan\left(\beta_{\mathrm{L}}\right)}{1+Htan\left(\beta_{\mathrm{L}}\right)}\right)$$

where δ is the disparity of the horopter at a given azimuth in the left eye (β_{L}) and positive azimuths are leftward in the visual field (34). To convert left eye azimuths to cyclopean azimuths, half of δ was then subtracted from β_{L}. One observer's data were excluded from study (34) because she had atypical binocular vision.

For the illustrations in Fig. 6 (A and B), we projected the horopter disparities out from the eyes and found their intersection in space. The vertical horopter disparities were fit with a second-order polynomial, and all disparities were magnified before projecting to make the pattern of differences between the geometric and empirical horopters clear.

Bayesian model for disparity estimation

To predict the 3D percepts of a Bayesian observer from ambiguous stereograms, we performed the following calculations. Let us denote the left image as L and the right image as R, which are functions of Helmholtz coordinates (η,β), where η and β denote elevation and azimuth in the visual field, respectively. Further, let us denote disparity across the visual field as D, which is also a function of (η,β). If we denote the prior distribution on disparity as P(D), then the estimation of the disparity map D can be formulated as a MAP estimation problem, where the posterior is:

$$P\left(D|L,R\right) \propto P\left(L,R|D\right)P\left(D\right).$$

In this formulation, P(L,R|D) is the likelihood of the binocular images given the disparity, and P(D) is the prior on disparity. For simplicity, we show the development for horizontal disparity only, but the model was also generalized to estimating a two-dimensional disparity vector.

Although natural disparities are likely to exhibit correlations, we asked if the stereoscopic effects could be explained using only zero-order disparity statistics. Therefore, we assume an independent distribution over the disparities in different spatial positions (η,β) for the likelihood and prior. Under these assumptions, we can get a MAP estimate for the whole visual field by maximizing the posterior at each point in the visual field independently. If we denote δ_p to be a random variable that represents disparity at a point p in the visual field, with coordinates (η_p,β_p), then our posterior at that point is:

$$P\left(\delta_p|L,R\right) \propto P\left(L,R|\delta_p\right)P\left(\delta_p\right).$$

We fit a function to the prior P(δ_p), at all points (η_p,β_p) in the visual field, using kernel density estimation from the Python Statsmodels module.

The likelihood P(L,R| δ_p) represents the probability of matches between the left and right eye images, given the disparity value. This probability is modeled using the normalized local

cross-correlation, a good approximation to the disparity energy calculation performed in visual cortex (58). Each image is sampled by a Gaussian window with an SD of 10 arc min (58) and centered at point p, denoted as $w_p(\eta,\beta)$. The local cross-correlation between the windowed left and right image is a function of disparity:

$$c(\delta_p) = \int_\eta \int_\beta \left[w_p(\eta,\beta) \, L\left(\eta, \beta + \frac{\delta_p}{2}\right) \right] \left[w_p(\eta,\beta) \, R\left(\eta, \beta - \frac{\delta_p}{2}\right) \right] d\eta d\beta$$

To normalize the correlation values to have maximum value 1, we divide it with its maximum value. Furthermore, we convert the cross-correlation into a likelihood using a sigmoidal function:

$$P(L, R | \delta_p) = \frac{1}{1 + e^{-ac(\delta_p)}}$$

where a determines the slope of the sigmoid and can thus change the shape of the likelihood. The sigmoid was chosen because it is commonly used as a nonlinearity in converting stimulus inputs into neural responses.

We examined the influence of a on the MAP solution for the venetian blinds effect. Smaller values of a yield smoother solutions and smaller estimated disparities; larger values yield more sawtooth-like solutions and larger estimated disparities. Values of 2 to 10 yielded MAP solutions that were consistent with perceived depth in the venetian blinds effect, so we chose a value of 3.

Physiology coordinate system conversion

To describe how we convert from positions on the screens to retinal disparities in Helmholtz coordinates, we use the following notation: x, y: horizontal and vertical screen coordinates relative to the fixation point; a: azimuth longitude (positive leftward, like β); H_δ: half vergence angle; d: screen distance.

We consider three types of apparatus for measuring disparity selectivity in cortical neurons: single-screen (2, 65) and two-screen haploscopes with lines of sight normal to the screens (63), and two-screen haploscope with lines of sight angled to the screens in the same way they would be with a single-screen setup (1, 64). Table S1 provides the details of each of the setups used in these studies.

With a single-screen setup, the position of each eye's image in screen coordinates can be converted to Helmholtz retinal coordinates with the following equations. These equations assume that the animal is fixating the center of the screen and that positive x and positive y are right and up on the screen, respectively. Screen coordinates are measured in the same units as the screen distance (typically centimeters). Azimuths and elevations in longitude for the left and right eyes can be computed from screen coordinates as follows:

$$\alpha_L = -arctan\left(\frac{x_L \quad cos \quad H_\delta}{d+x_L \quad sin \quad H_\delta \quad cos \quad H_\delta}\right)$$

$$\alpha_R = -arctan\left(\frac{x_R \quad cos \quad H_\delta}{d+x_R \quad sin \quad H_\delta \quad cos \quad H_\delta}\right)$$

$$\eta_L = arctan\left(\frac{y_L}{d+x_L \quad sin \quad H_\delta}\right)$$

$$\eta_R = arctan\left(\frac{y_R}{d+x_R \quad sin \quad H_\delta}\right)$$

We want horizontal disparities to be defined in terms of azimuth latitude, so we convert each eye's azimuth longitude:

$$\beta = arctan\left(tan \quad \alpha \quad cos \quad \eta\right)$$

Disparity is then:

$$\delta_\beta = \beta_L - \beta_R$$
$$\delta_\eta = \eta_L - \eta_R$$

where d_b is horizontal disparity and d_h is vertical disparity. Finally, position in the cyclopean visual field for that disparity is:

$$\beta_C = \frac{\beta_L + \beta_R}{2}$$
$$\eta_C = \frac{\eta_L + \eta_R}{2}$$

With a Wheatstone haploscope in which the optic axis of each eye is normal to the screen regardless of the vergence angle, the position of each eye's image in Helmholtz coordinates is:

$$\beta_L = arctan\left(\frac{x_L}{\sqrt{d^2+y_L^2}}\right)$$

$$\beta_R = arctan\left(\frac{x_R}{\sqrt{d^2+y_R^2}}\right)$$

$$\eta_L = arctan\frac{y_L}{d}$$

$$\eta_R = arctan\frac{y_R}{d}$$

Note, however, that having an odd number of mirrors in each eye's optical path flips the sign of the x coordinate in each eye. Disparity and visual field location then proceed as before.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

REFERENCES AND NOTES

1. Cumming BG. An unexpected specialization for horizontal disparity in primate primary visual cortex. Nature. 2002; 418:633–636. [PubMed: 12167860]

2. Durand J-B, Celebrini S, Trotter Y. Neural bases of stereopsis across visual field of the alert macaque monkey. Cereb. Cortex. 2007; 17:1260–1273. [PubMed: 16908495]

3. Land MF, Hayhoe M. In what ways do eye movements contribute to everyday activities? Vision Res. 2001; 41:3559–3565. [PubMed: 11718795]

4. Hollands MA, Patla AE, Vickers JN. "Look where you're going!": Gaze behaviour associated with maintaining and changing the direction of locomotion. Exp. Brain Res. 2002; 143:221–230. [PubMed: 11880898]

5. Marr D, Poggio T. A computational theory of human stereo vision. Proc. R. Soc. Lond. B. Biol. Sci. 1979; 204:301–328. [PubMed: 37518]

6. Read JCA. A Bayesian model of stereopsis depth and motion direction discrimination. Biol. Cybern. 2002; 86:117–136. [PubMed: 11911114]

7. Prince SJD, Pointon AD, Cumming BG, Parker AJ. Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. J. Neurophysiol. 2002; 87:191–208. [PubMed: 11784742]

8. Ganguli D, Simoncelli EP. Efficient sensory encoding and Bayesian inference with heterogeneous neural populations. Neural Comput. 2014; 26:2103–2134. [PubMed: 25058702]

9. Barlow, H. Sensory Communication. Rosenblith, W., editor. MIT Press; Cambridge, MA: 1961. p. 217-234.

10. Atick JJ, Redlich AN. What does the retina know about natural scenes? Neural Comput. 1992; 4:196–210.

11. Olshausen BA, Field DJ. Natural image statistics and efficient coding. Network. 1996; 7:333–339. [PubMed: 16754394]

12. Schwartz O, Simoncelli EP. Natural signal statistics and sensory gain control. Nat. Neurosci. 2001; 4:819–825. [PubMed: 11477428]

13. Karklin Y, Lewicki MS. Emergence of complex cell properties by learning to generalize in natural scenes. Nature. 2009; 457:83–86. [PubMed: 19020501]

14. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996; 381:607–609. [PubMed: 8637596]

15. Girshick AR, Landy MS, Simoncelli EP. Cardinal rules: Visual orientation perception reflects knowledge of environmental statistics. Nat. Neurosci. 2011; 14:926–932. [PubMed: 21642976]

16. Geisler WS. Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol. 2008; 59:167–192. [PubMed: 17705683]

17. Read JCA, Cumming BG. Understanding the cortical specialization for horizontal disparity. Neural Comput. 2004; 16:1983–2020. [PubMed: 15333204]

18. Hibbard PB, Bouzit S. Stereoscopic correspondence for ambiguous targets is affected by elevation and fixation distance. Spat. Vis. 2005; 18:399–411. [PubMed: 16167773]

19. Hibbard PB. A statistical model of binocular disparity. Vis. Cogn. 2007; 15:149–165.

20. Burge J, Geisler WS. Optimal disparity estimation in natural stereo images. J. Vis. 2014; 14:1–18. [PubMed: 24492596]

21. Liu Y, Bovik AC, Cormack LK. Disparity statistics in natural scenes. J. Vis. 2008; 8:1–14.

22. Yang Z, Purves D. A statistical explanation of visual space. Nat. Neurosci. 2003; 6:632–640. [PubMed: 12754512]

23. Hillis JM, Banks MS. Are corresponding points fixed? Vision Res. 2001; 41:2457–2473. [PubMed: 11483177]

24. Gur M, Snoddderly DM. Visual receptive fields of neurons in primary visual cortex (V1) move in space with the eye movements of fixation. Vision Res. 1997; 37:257–265. [PubMed: 9135859]

25. Read JCA, Cumming BG. Measuring V1 receptive fields despite eye movements in awake monkeys. J. Neurophysiol. 2003; 90:946–960. [PubMed: 12711706]

26. Hayhoe M. Vision using routines: A functional account of vision. Vis. Cogn. 2000; 7:43–64.

27. Patla AE, Vickers JN. How far ahead do we look when required to step on specific locations in the travel path during locomotion? Exp. Brain Res. 2003; 148:133–138. [PubMed: 12478404]

28. Time spent in detailed primary activities, and percent of the civilian population engaging in each detailed activity category, averages per day by sex.. Bureau of Labor Statistics American Time Use Survey. Table A-1available at http://www.bls.gov/tus/tables.htm

29. Bahill AT, Adler D, Stark L. Most naturally occurring human saccades have magnitudes of 15 degrees or less. Invest. Ophthalmol. 1975; 14:468–469. [PubMed: 1132942]

30. Stahl JS. Eye-head coordination and the variation of eye-movement accuracy with orbital eccentricity. Exp. Brain Res. 2001; 136:200–210. [PubMed: 11206282]

31. Cooper EA, Burge J, Banks MS. The vertical horopter is not adaptable, but it may be adaptive. J. Vis. 2011; 11:1–19.

32. Dodgson N. Variation and extrema of human interpupillary distance. Proc. SPIE. 2004; 5291:36–46.

33. Prince SJD, Eagle RA. Weighted directional energy model of human stereo correspondence. Vision Res. 2000; 40:1143–1155. [PubMed: 10738073]

34. Read JCA, Phillipson GP, Glennerster A. Latitude and longitude vertical disparities. J. Vis. 2009; 9:1–37.

35. Schreiber KM, Hillis JM, Filippini H, Schor C, Banks M. The surface of the empirical horopter. J. Vis. 2008; 8:1–20.

36. Gårding J, Porrill J, Mayhew J, Frisby J. Stereopsis, vertical disparity and relief transformations. Vision Res. 1995; 35:703–722. [PubMed: 7900308]

37. Brewster D. On the knowledge of distance given by binocular vision. Trans. R. Soc. Edinburgh. 1844; 15:663–675.

38. Prince SJD, Eagle RA. Stereo correspondence in one-dimensional Gabor stimuli. Vision Res. 2000; 40:913–924. [PubMed: 10720662]

39. Vlaskamp BNS, Guan P, Banks MS. The venetian-blind effect: A preference for zero disparity or zero slant? Front. Psychol. 2013; 4:836. [PubMed: 24273523]

40. Fischer FP. Fortgesetze Studien über Binokularsehen. III. Experimentelle Beiträge zum Begriff der Seerichtungsgemeinschaft der Netzhaute auf Grund der binokularen Noniusmethode. Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere. 1924; 204:247–260.

41. Ogle, KN. Researches in Binocular Vision. Saunders; Philadelphia: 1950.

42. Ogle KN. Precision and validity of stereoscopic depth perception from double images. J. Opt. Soc. Am. 1953; 43:907–913. [PubMed: 13097263]

43. Blakemore C. The range and scope of binocular depth discrimination in man. J. Physiol. 1970; 211:599–622. [PubMed: 5501054]

44. Julesz B, Breitmeyer B, Kropfl W. Binocular-disparity-dependent upper-lower hemifield anisotropy and left-right hemifield isotropy as revealed by dynamic random-dot stereograms. Perception. 1976; 5:129–141. [PubMed: 951161]

45. Schumer RA, Julesz B. Binocular disparity modulation sensitivity to disparities offset from the plane of fixation. Vision Res. 1984; 24:533–542. [PubMed: 6740973]

46. Badcock DR, Schor CM. Depth-increment detection function for individual spatial channels. J. Opt. Soc. Am. A. 1985; 2:1211–1216. [PubMed: 4020518]

47. Tyler, CW. Vision and Visual Dysfunction. Regan, D., editor. Macmillan; London: 1991. p. 38-74.

48. von Helmholtz, H. Handbuch der Physiologischen Optik. ed. 1. Optical Society of America; Rochester NY: p. 1867

49. Nakayama K. Benton S. Geometrical and physiological aspects of depth perception. 3-D Image Processing. Society of Photo-Optical Instrumentation Engineers Proceedings. 1977; 20:1–8.

50. Siderov J, Harwerth RS, Bedell HE. Stereopsis, cyclovergence and the backwards tilt of the vertical horopter. Vision Res. 1999; 39:1347–1357. [PubMed: 10343847]

51. Grove PM, Kaneko H, Ono H. The backward inclination of a surface defined by empirical corresponding points. Perception. 2001; 30:411–429. [PubMed: 11383190]

52. Amigo G. The stereoscopic frame of reference in asymmetric convergence of the eyes. Vision Res. 1967; 7:785–799. [PubMed: 5608667]

53. McKee SP, Verghese P, Ma-Wyatt A, Petrov Y. The wallpaper illusion explained. J. Vis. 2007; 7:1–11.

54. Cumming BG, Parker AJ. Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. J. Neurosci. 2000; 20:4758–4767. [PubMed: 10844045]

55. Cibis PA, Haber H. Anisopia and perception of space. J. Opt. Soc. Am. 1951; 41:676. [PubMed: 14880996]

56. Tyler CW, Sutter EE. Depth from spatial frequency difference: An old kind of stereopsis? Vision Res. 1979; 19:859–865. [PubMed: 516457]

57. Halpern DL, Wilson HR, Blake R. Stereopsis from interocular spatial frequency differences is not robust. Vision Res. 1996; 36:2263–2270. [PubMed: 8776491]

58. Qian N, Zhu Y. Physiological computation of binocular disparity. Vision Res. 1997; 37:1811–1827. [PubMed: 9274767]

59. Filippini HR, Banks MS. Limits of stereopsis explained by local cross-correlation. J. Vis. 2009; 9:8.1–8.18. [PubMed: 19271878]

60. Morgan MJ, Castet E. The aperture problem in stereopsis. Vision Res. 1997; 37:2737–2744. [PubMed: 9373672]

61. Rambold HA, Miles FA. Human vergence eye movements to oblique disparity stimuli: Evidence for an anisotropy favoring horizontal disparities. Vision Res. 2008; 48:2006–2019. [PubMed: 18675438]

62. Weiss Y, Simoncelli EP, Adelson EH. Motion illusions as optimal percepts. Nat. Neurosci. 2002; 5:598–604. [PubMed: 12021763]

63. Hedges JH, Stocker AA, Simoncelli EP. Optimal inference explains the perceptual coherence of visual motion stimuli. J. Vis. 2011; 11:1–16.

64. Brunel N, Nadal J-P. Mutual information, Fisher information, and population coding. Neural Comput. 1998; 10:1731–1757. [PubMed: 9744895]

65. Gonzalez F, Bermudez MA, Vicente AF, Romero MC. Orientation preference and horizontal disparity sensitivity in the monkey visual cortex. Ophthalmic Physiol. Opt. 2010; 30:824–833. [PubMed: 21205269]

66. Prince S, Cumming B, Parker A. Range and mechanism of encoding of horizontal disparity in macaque V1. J. Neurophysiol. 2002; 87:209–221. [PubMed: 11784743]

67. Samonds JM, Potetz BR, Lee TS. Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. Proc. Natl. Acad. Sci. U.S.A. 2012; 109:6313–6318. [PubMed: 22474369]

68. Chen G, Lu HD, Roe AW. A map for horizontal disparity in monkey V2. Neuron. 2008; 58:442–450. [PubMed: 18466753]

69. Adams D, Zeki S. Functional organization of macaque V3 for stereoscopic depth. J. Neurophysiol. 2001; 86:2195–2203. [PubMed: 11698511]

70. Hinkle DA, Connor CE. Quantitative characterization of disparity tuning in ventral pathway area V4. J. Neurophysiol. 2005; 94:2726–2737. [PubMed: 15987762]

71. Tanabe S, Doi T, Umeda K, Fujita I. Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area V4. J. Neurophysiol. 2005; 94:2683–2699. [PubMed: 16000525]

72. Watanabe M, Tanaka H, Uka T, Fujita I. Disparity-selective neurons in area V4 of macaque monkeys. J. Neurophysiol. 2002; 87:1960–1973. [PubMed: 11929915]

73. Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. J. Neurophysiol. 1983; 49:1148–1167. [PubMed: 6864243]

74. Bradley DC, Andersen RA. Center-surround antagonism based on disparity in primate area MT. J. Neurosci. 1998; 18:7552–7565. [PubMed: 9736673]

75. DeAngelis GC, Uka T. Coding of horizontal disparity and velocity by MT neurons in the alert macaque. J. Neurophysiol. 2003; 89:1094–1111. [PubMed: 12574483]

76. Gonzalez F, Perez R, Justo MS, Bermudez MA. Receptive field organization of disparity-sensitive cells in macaque medial superior temporal cortex. Eur. J. Neurosci. 2001; 14:167–173. [PubMed: 11488961]

77. Cooper ML, Pettigrew JD. A neurophysiological determination of the vertical horopter in the cat and owl. J. Comp. Neurol. 1979; 184:1–26. [PubMed: 762277]

78. Shinkman PG, Bruce CJ. Binocular differences in cortical receptive fields of kittens after rotational disparate binocular experience. Science. 1977; 197:285–287. [PubMed: 877554]

79. Hirschmüller H. Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. 2008; 30:328–341. [PubMed: 18084062]

80. Birchfield S, Tomasi C. A pixel dissimilarity measure that is insensitive to image sampling. IEEE Trans. Pattern Anal. Mach. Intell. 1998; 20:401–406.

81. Mok D, Ro A, Cadera W, Crawford JD, Vilis T. Rotation of Listing's plane during vergence. Vision Res. 1992; 32:2055–2064. [PubMed: 1304083]

82. Somani RA, DeSouza JF, Tweed D, Vilis T. Visual test of Listing's law during vergence. Vision Res. 1998; 38:911–923. [PubMed: 9624440]
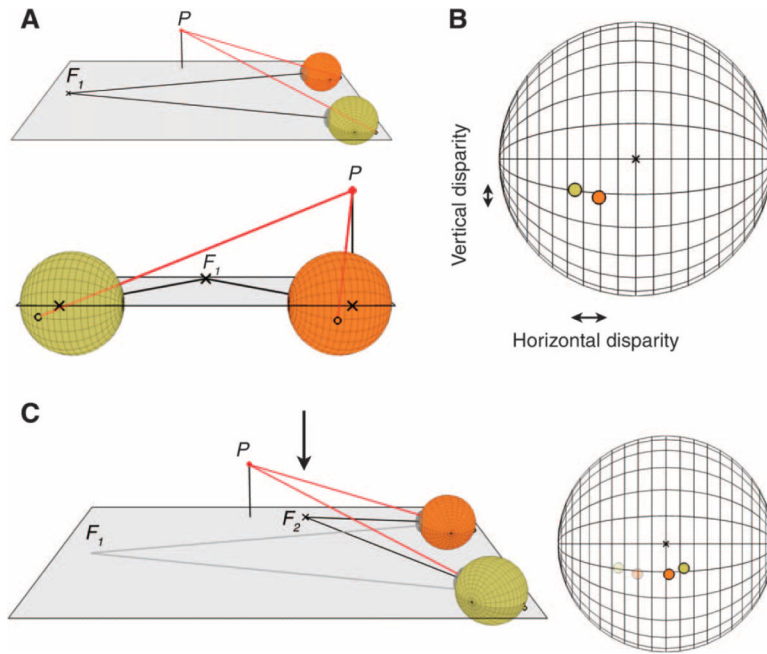
Fig. 1. Binocular disparities used for stereopsis

(**A**) Two views of a simple 3D scene. The eyes are fixating point F_1, which lies straight ahead. Point P is positioned above and to the right of the viewer's face, and is closer in depth than F_1. The upper panel shows a side view and the lower panel a view from behind the eyes. Lines of equal azimuth and elevation in Helmholtz coordinates are drawn on each eye. (**B**) Retinal projections of P from the viewing geometry in (A). The yellow and orange dots correspond to the projections in the left and right eyes, respectively. The difference between the left and right eye projections is binocular disparity. The difference in azimuth is horizontal disparity, and the difference in elevation is vertical disparity. In this example, point P has crossed horizontal disparity, because it is closer than point F_1 and the image of P is therefore shifted leftward in the left eye and rightward in the right eye. (**C**) For a given point in the scene, the disparity at the retinas can change substantially depending on where the viewer is fixating. In the left panel, the same point P is observed, but with a different fixation point F_2 that is now closer to the viewer than P (indicated by the arrow). The original fixation point F_1 is overlaid in gray. In the right panel, the retinal disparities projected by P are shown for both fixations [disparities from (B) are semitransparent]. For this viewing geometry, point P now has uncrossed horizontal disparity: that is, the image of P is shifted rightward in the left eye and leftward in the right eye.
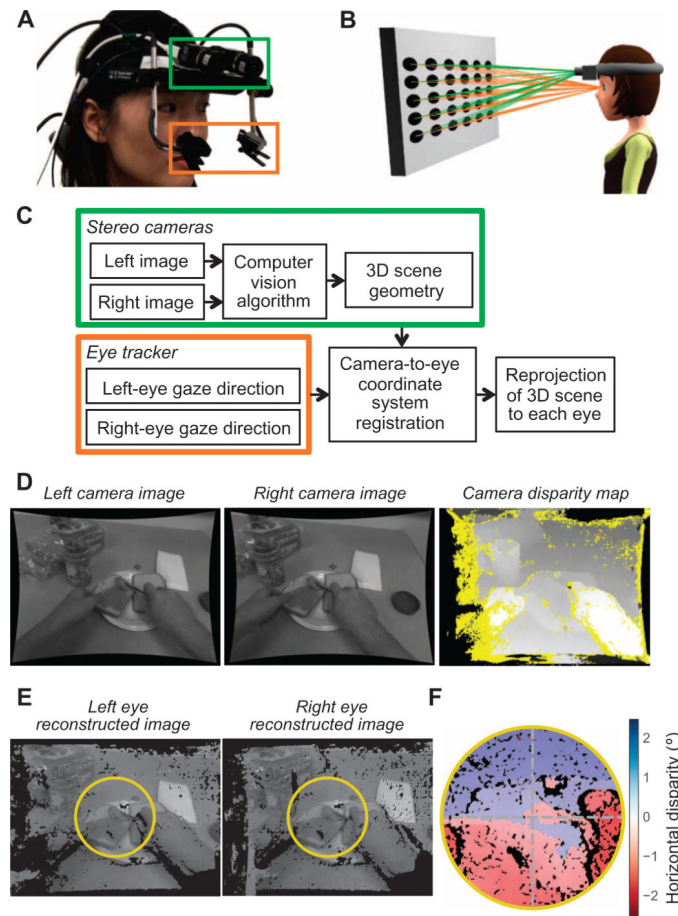
Fig. 2. System and method for determining retinal disparities when participants are engaged in everyday tasks

(**A**) The device. A head-mounted binocular eye tracker (EyeLink II) was modified to include outward-facing stereo cameras. The eye tracker measured the gaze direction of each eye using a video-based pupil and corneal reflection tracking algorithm. The stereo cameras captured uncompressed images at 30 Hz and were synchronized to one another with a hardware trigger. The cameras were also synchronized to the eye tracker using digital inputs. Data from the tracker and cameras were stored on mobile computers in a backpack worn by the participant. Depth maps were computed from the stereo images offline. (**B**) Calibration. The cameras were offset from the participant's eyes by several centimeters. To reconstruct the disparities from the eyes' viewpoints, we performed a calibration procedure for determining the translation and rotation between the 3D locations. The participant was positioned with a bite bar to place the eyes at known positions relative to a large display. A calibration pattern was then displayed for the cameras and used to determine the transformation between the cameras' viewpoints and the eyes' known positions. (**C**) Schematic of the entire data collection and processing workflow. (**D**) Two example images from the stereo cameras. Images were warped to remove lens distortion, resulting in bowed edges. The disparities between these two images were used to reconstruct the 3D geometry of the scene. These disparities are illustrated as a grayscale image with bright points representing near pixels and dark points representing far pixels. Yellow indicates regions in

which disparity could not be computed due to occlusions and lack of texture. (**E**) The 3D points from the cameras were reprojected to the two eyes. The yellow circles (20° diameter) indicate the areas over which statistical analyses were performed. (**F**) Horizontal disparities for this scene at one time point. Different disparities are represented by different colors as indicated by the color map on the right.
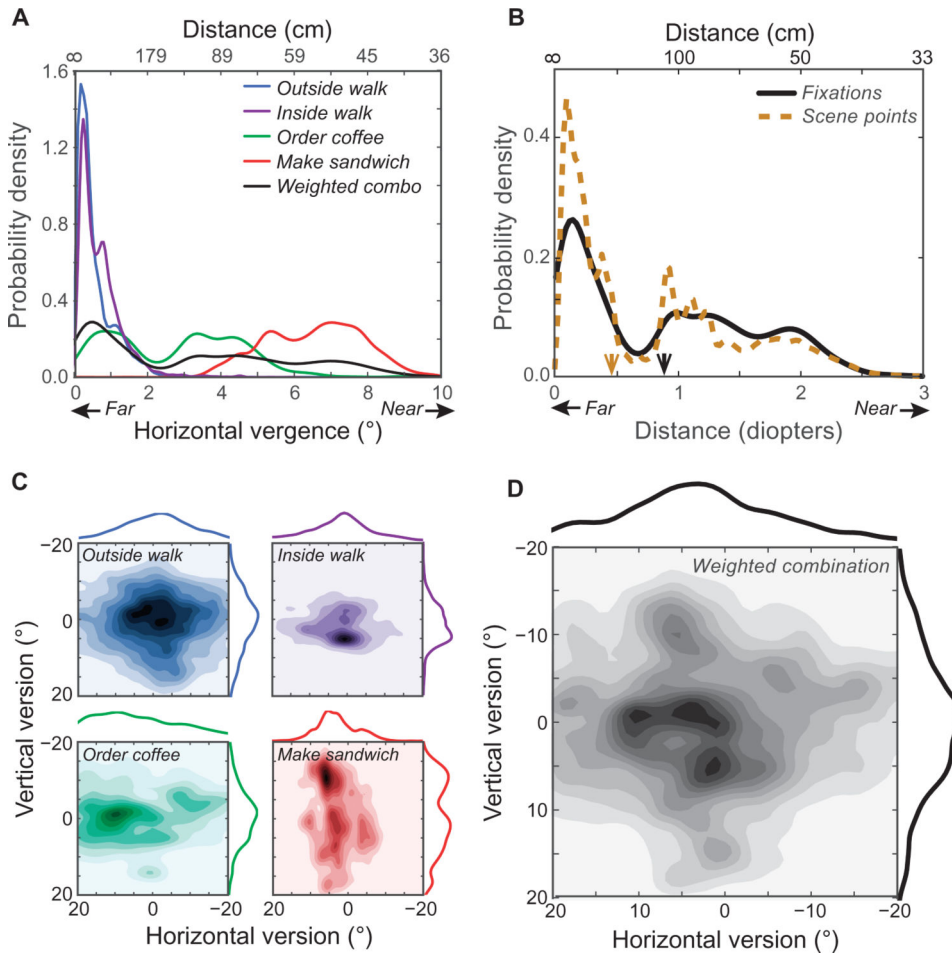
Fig. 3. Distributions of binocular eye position for the four tasks and weighted combination of tasks

(**A**) Distributions of horizontal vergence for the tasks and the weighted combination. Probability density is plotted as a function of vergence angle. Small and large vergences correspond to far and near fixations, respectively. For reference, vergence distance (assuming an interocular distance of 6.25 cm and version of 0°) is plotted on the abscissa at the top of the figure. Different colors represent the data from different tasks; the thick black contour represents the weighted-combination data. (**B**) Distances of fixation points and visible scene points. The probability of different fixation distances and scene distances is plotted as a function of distance in diopters (where diopters are inverse meters). The upper abscissa shows those distances in centimeters for comparison. The black solid curve represents the distribution of weighted-combination fixation distances. The orange dashed curve represents the distribution of distances of visible scene points, whether fixated or not, falling within 10° of the fovea. All distances are measured from the midpoint on the interocular axis (the cyclopean eye). The median distances are indicated by the arrows: the orange arrow for scene points and the black one for fixation points. (**C**) Distributions of version for the four tasks. Horizontal version and vertical version are plotted, respectively, on the abscissa and ordinate of each panel. By convention, positive horizontal and vertical

versions are leftward and downward, respectively. The data are, of course, plotted in head-centered coordinates. Colors represent probability (darker being more probable).
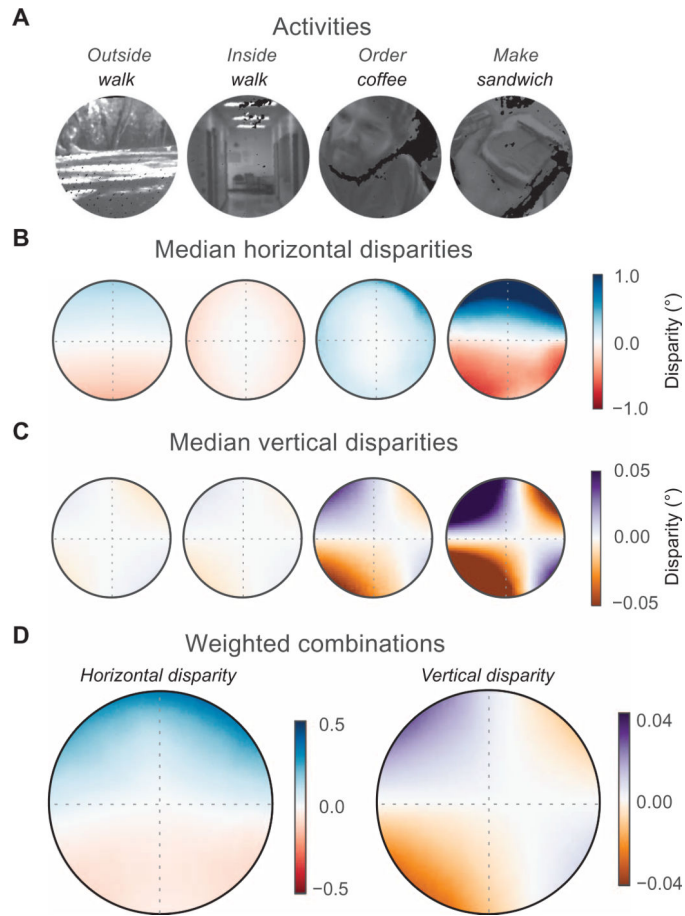
Fig. 4. Median disparities in different parts of visual field for different tasks
(**A**) Example frames from the four activities. (**B**) Median horizontal disparities and visual field position. The circular plots each represent data from one of the four activities averaged across participants. The fovea is at the center of each panel. Radius is 10°, so each panel shows medians for the central 20° of the visual field. Dashed gray lines represent the horizontal and vertical meridians. Uncrossed disparities are blue, crossed disparities red, and zero disparity white. (**C**) Median vertical disparities and visual field position. Positive disparities are purple, negative disparities orange, and zero disparity white. The panels show the medians for each activity averaged across participants. (**D**) Weighted combinations. The left panel shows the median horizontal disparities when the four activities are weighted according to ATUS. The right panel shows the combined median vertical disparities. The scales of the color maps differ from the scales for the other panels.
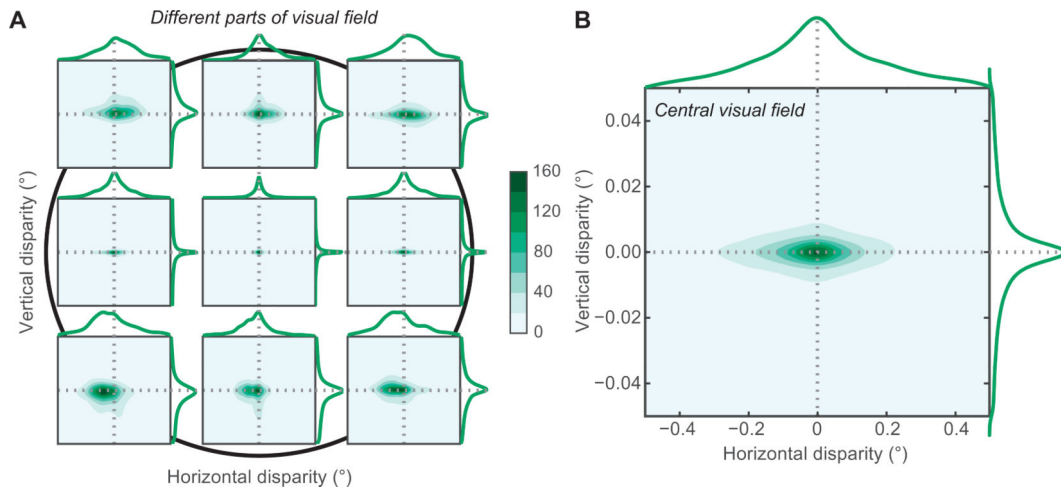
Fig. 5. Distributions of horizontal and vertical disparities

(**A**) Joint distributions of horizontal and vertical disparities in different regions of the visual field. The black circle in the background represents the central 20° of the visual field (10° radius) with the fovea at the center. Each subplot shows the joint disparity distribution in a different region of the visual field. Horizontal disparity is plotted on the abscissa of each subplot and vertical disparity on the ordinate. The scales for the abscissa and ordinates differ by a factor of 10, the abscissa ranging from −0.5° to +0.5° and the ordinate from −0.05° to +0.05°. The centers of the represented regions are either at the fovea or 4° from the fovea. The regions are all 3° in diameter. Frequency of occurrence is represented by color, darker corresponding to more frequent disparities. The curves at the top and right of each subplot are the marginal distributions. Horizontal and vertical disparities are positively correlated in the upper left and lower right parts of the visual field and negatively correlated in the upper right and lower left parts of the field (absolute r values greater than 0.16, P values less than 0.001). One can see in these panels the shift from uncrossed (positive) horizontal disparity in the upper visual field to crossed (negative) disparity in the lower field. We calculated the range of horizontal disparities needed to encompass two-thirds of the observations in each of the nonfoveal positions in the figure. For the upper middle, upper right, mid right, lower right, lower middle, lower left, mid left, and upper left regions, the disparity ranges were respectively 0.69°, 1.37°, 0.53°, 0.73°, 0.49°, 0.72°, 0.25°, and 0.43°; the average across regions was 0.74°. (**B**) Joint distribution of horizontal and vertical disparities in the central 20° (10° radius) of the visual field. The format is the same as the plots in A. Movies S1 and S2 show how the marginal distribution of horizontal disparity changes as a function of elevation and azimuth.
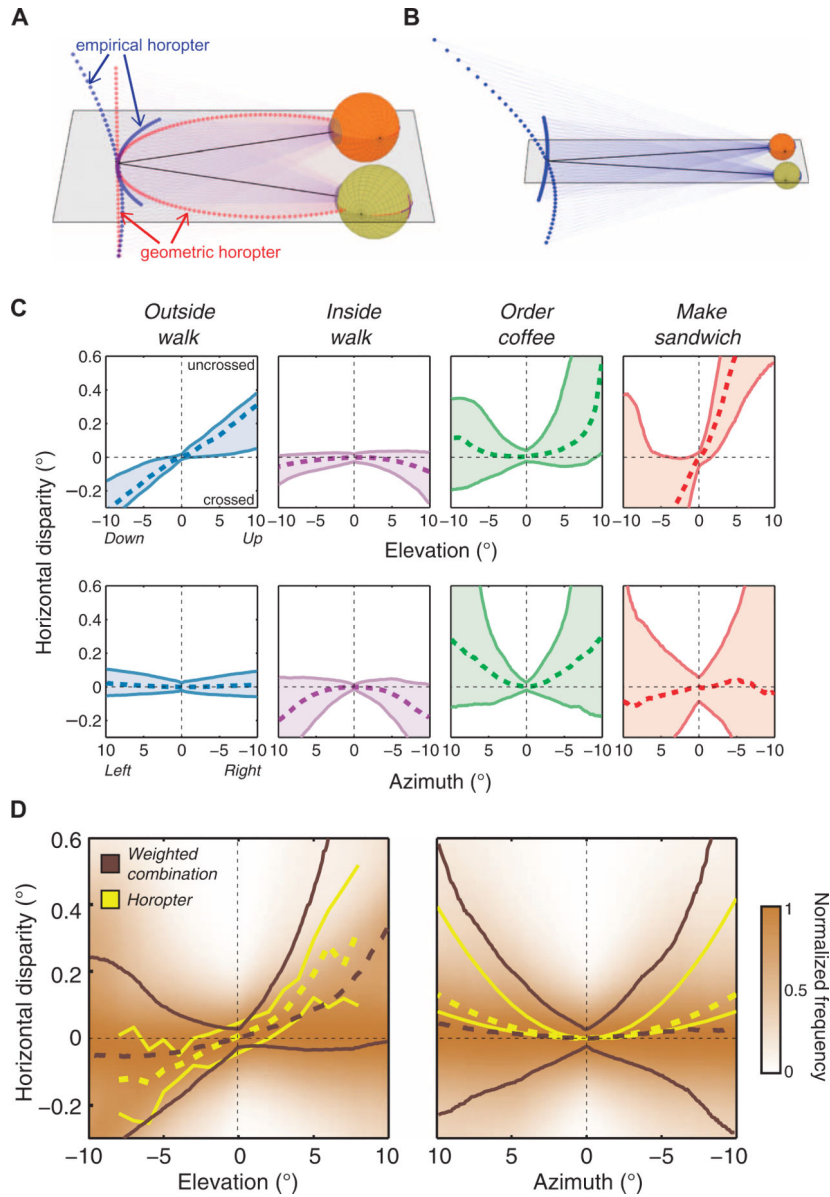
Fig. 6. Natural disparity statistics and the horopter
(**A**) Geometric and empirical horopters. Points in 3D space that project to zero horizontal and vertical disparities are shown as a set of red points. The lines of sight are represented by the black lines, and their intersection is the fixation point. The geometric horopter contains points that lie on a circle that runs through the fixation point and the two eye centers (the horizontal horopter or Vieth-Müller Circle) and a vertical line in the head's median plane (the vertical horopter). Points that empirically correspond are not in the same positions in the two eyes. The projections of those points define the empirical horopter, which is shown in blue. The horizontal part is less concave than the Vieth-Müller Circle, and the vertical part is convex and pitched top-back relative to the vertical line. The disparities associated with the empirical horopter have been magnified relative to their true values to make the differences between the empirical and geometric horopters evident. (**B**) Corresponding points are fixed

in the retinas, so when the eyes move, the empirical horopter changes. Here, the eyes have diverged relative to (A) to fixate a farther point. The horizontal part of the empirical horopter becomes convex, and the vertical part becomes more convex and more pitched. (**C**) Distributions of horizontal disparity near the vertical and horizontal meridians of the visual field. The upper row plots horizontal disparity as a function of vertical eccentricity near the vertical meridian. The lower row plots horizontal disparity as a function of horizontal eccentricity near the horizontal meridian. The columns show the data from each of the four tasks averaged across participants. The colored regions represent the data between the 25th and 75th percentiles. The dashed lines represent the median disparities. (**D**) Weighted-combination data and the empirical horopter. The left panel shows the weighted-combination data and the horopter near the vertical meridian; horizontal disparity is plotted as a function of vertical eccentricity. The right panel shows the data and horopter near the horizontal meridian; horizontal disparity is plotted as a function of horizontal eccentricity. The brown lines represent percentiles of the natural disparity data: dashed lines are the medians, and solid lines are the 25th and 75th percentiles. The yellow lines represent the same percentiles for the horopter. The horopter data in the left panel come from 28 observers in (31). The horopter data in the right panel come from 15 observers in (23, 35, 42, 52). We computed the probability density of natural disparities at each eccentricity, and these distributions are underlaid in brown, with the color scale normalized to the peak disparity at each eccentricity.
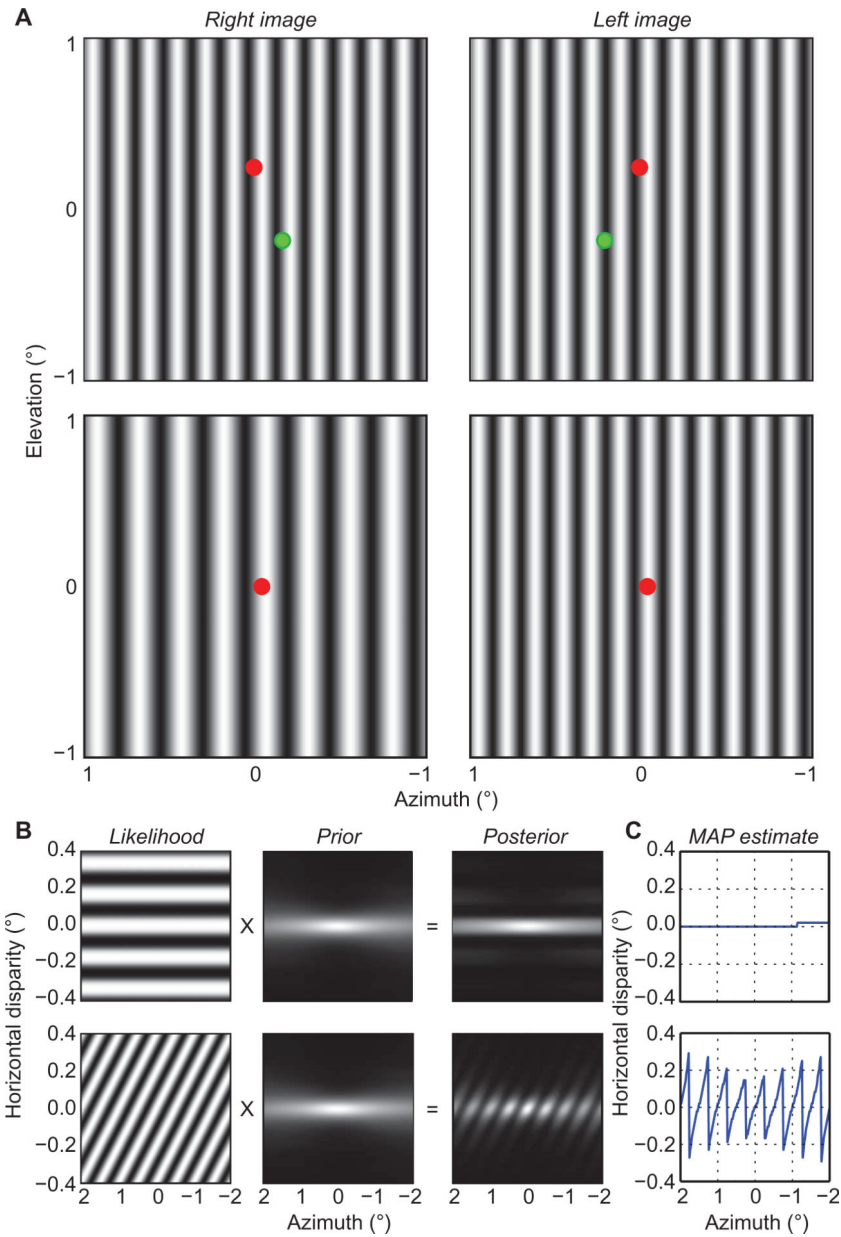
Fig. 7. Wallpaper illusion and venetian blinds effect
(**A**) Upper part shows the wallpaper illusion. Vertical sinewave gratings of the same spatial frequency and contrast are presented to the two eyes. Cross-fuse the red dots to see the stimulus stereoscopically. Notice that the stimulus appears to lie in the same plane as the fused red dot. Now cross-fuse the green dots and notice that the stimulus now appears to lie in the same plane as the fused green dot. The lower part shows the venetian blinds effect. Vertical sinewave gratings of different spatial frequencies are presented to the two eyes ($f_R/f_L = 0.67$). Cross-fuse the red dots to see the stimulus stereoscopically. Hold fixation steady on the fused red dot and notice that the stimulus appears to be a series of slanted planes. (**B**) Likelihood, prior, and posterior distributions as a function of azimuth; elevation is zero. The upper row shows the distributions for the wallpaper stimulus and the lower row

the distributions for the venetian blinds stimulus. The grayscale represents probability, brighter values having greater probability. (**C**) MAP estimates for wallpaper (upper) and venetian blinds (lower). Note the change in the scale of the ordinate relative to (B). If one wanted to visually match the stimuli in A to the units in B and C, the correct viewing distance is 148 cm, so that the stimuli subtend approximately 2×2°. Note that A shows only a portion of the stimuli used to compute the results.
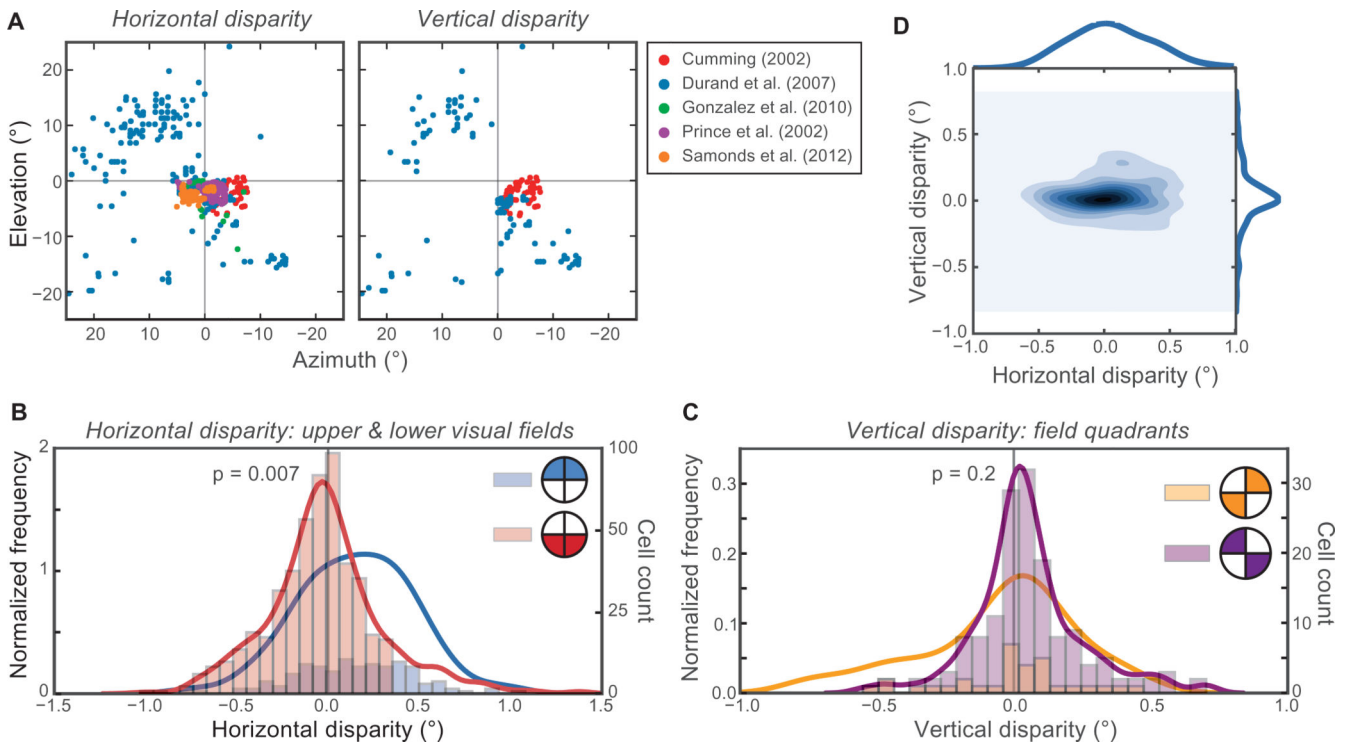
Fig. 8. Receptive-field locations and preferred-disparity distributions among 973 disparity-sensitive neurons from macaque V1

(**A**) Receptive-field location of each neuron. The left and right subplots show neurons for which preferred horizontal disparity and preferred vertical disparity were measured, respectively. The colors represent the study from which each neuron came. Most neurons had receptive fields in the lower visual field. Very few were in the upper right quadrant. (**B**) Distributions of preferred horizontal disparity grouped by upper and lower visual field. The histograms represent the number of cells observed with each preferred disparity. The curves represent those histograms normalized to constant area. The preferred disparities of neurons from the upper visual field (blue) are biased toward uncrossed disparities, whereas the preferences of neurons from the lower visual field (red) are biased toward crossed disparities. (**C**) Distributions of preferred vertical disparities grouped by quadrants with the same expected sign of vertical disparity. Again, the histograms represent the actual cell counts, and the curves represent those histograms normalized for constant area. (**D**) Joint distribution of preferred horizontal and vertical disparities from the sampled neurons. Darker color represents more frequent observations. The marginal distributions for horizontal and vertical disparities are shown on the top and right, respectively.