



Does depth perception require vertical-disparity detectors?

Jenny C. A. Read

School of Biology and Psychology, Newcastle University, UK  

Bruce G. Cumming

Laboratory of Sensorimotor Research,
National Eye Institute, National Institutes of Health,
Bethesda, MD, USA  

Stereo depth perception depends on the fact that objects project to different positions in the two eyes. Because our eyes are offset horizontally, these retinal disparities are mainly horizontal, and horizontal disparity suffices to give an impression of depth. However, depending on eye position, there may also be small vertical disparities. These are significant because, given both vertical and horizontal disparities, the brain can deduce eye position from purely retinal information and, hence, derive the position of objects in space. However, we show here that, to achieve this, the brain need measure only the *magnitude* of vertical disparity; for physically possible stimuli, the sign then follows from the stereo geometry. The magnitude of vertical disparity—and hence eye position—can be deduced from the response of purely horizontal-disparity sensors because vertical disparity moves corresponding features off the receptive fields, reducing the effective binocular correlation. As proof, we demonstrate an algorithm that can accurately reconstruct gaze and vergence angles from the population activity of pure horizontal-disparity sensors and show that it is subject to the induced effect. Given that disparities experienced during natural viewing are overwhelmingly horizontal and that eye position measures require only horizontal-disparity sensors, this work raises two questions: Does the brain in fact contain sensors tuned to nonzero vertical disparities, and if so, why?

Keywords: binocular vision, vertical disparity, induced effect

Introduction

Because our eyes view the world from slightly different positions, a given object in space does not, in general, project to corresponding locations on the two retinæ. Because each retina is a two-dimensional (2D) surface, the disparity between the two images is, in principle, 2D. However, geometry imposes a significant simplification: for any point in one eye, the set of possible matches defines a line in the other eye. Consider the image at a point P in the left retina (Figure 1). The object that caused this image could lie anywhere along the ray that projects to the point P (red dashed line in Figure 1). The image of this ray in the right eye defines a one-dimensional (1D) line on the right retina. This *epipolar line* is the locus of all possible matches in the right eye for point P in the left eye. Objects at different distances fall at different places along the epipolar line. For any given eye position, therefore, disparity can be described with a purely 1D measure. However, changes in eye position shift the epipolar lines on the retina, making disparity genuinely 2D. The two dimensions of disparity thus carry different information: The component along the epipolar line carries information about the outside world (the location of objects in space), whereas the orientation of epipolar lines carries information about the observer (the current position of the eyes).

In the coordinate system used by Longuet-Higgins (1982) or Read and Cumming (2004), when the eyes are in primary position, all epipolar lines are horizontal, and hence, retinal

disparities are purely horizontal. Changes in gaze angle and vergence away from primary position rotate the epipolar lines on retina, and vertical disparities become possible. We recently investigated the range of horizontal and vertical disparities encountered in typical viewing situations (Read & Cumming, 2004). We found that the frequency distribution was highly elongated: Horizontal disparities are far commoner than vertical disparities of the same magnitude. This, of course, reflects the horizontal offset in the position of the eyes. Vertical disparities do occur, but become large only when the eyes are converged and looking off to one side. Because it seems likely that relatively little time is spent viewing objects obliquely, the disparities encountered by the visual system are overwhelmingly horizontal.

One might therefore expect that, to construct an efficient representation of the visual world (Barlow, 1961; Simoncelli & Olshausen, 2001), the brain should devote resources to encoding horizontal, rather than vertical, disparity. It should contain disparity detectors tuned to a range of horizontal disparities, reflecting those encountered in normal viewing, but they should almost all be tuned to zero vertical disparity because the vertical disparities encountered in real life are almost all always smaller than the range of an individual disparity detector anyway (Figure 5 of Read & Cumming, 2004). Such detectors would resemble the one sketched in Figure 2. The receptive fields of this cell fall at the same vertical positions in the two eyes, which means that the cell is tuned to zero vertical disparity, but at different horizontal

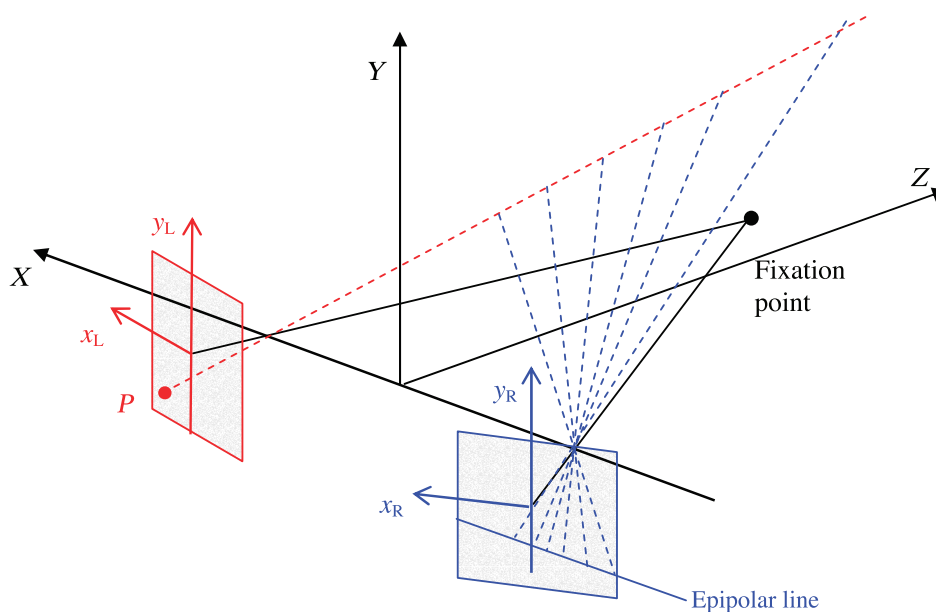


Figure 1. Definition of an epipolar line. The blue epipolar line on the right retina is the locus of all possible matches for the point P in the left retina. On the planar retina used here, the epipolar line is straight; if it were projected onto a curved retina, as in Figure 3, it would be curved.

positions, which means that it is tuned to a nonzero horizontal disparity. The statistics of binocular vision mean that the most efficient encoding is a population consisting almost entirely of pure horizontal-disparity detectors like the one in Figure 2. At least close to the fovea, the physiological evidence supports this expectation (Cumming, 2002; Gonzalez, Justo, Bermudez, & Perez, 2003; Gonzalez, Relova, Perez, Acuna, & Alonso, 1993; Maunsell & Van Essen, 1983; Poggio, 1995). For example, the only study to have systematically probed the response of cortical cells to all combinations of horizontal and vertical disparity (Cumming, 2002) found that the distribution of preferred disparities in parafoveal V1 neurons is nearly four times as wide in the horizontal direction as the vertical and that the spread in the vertical direction is comparable to the uncertainty in the measurement.

This fits with the psychophysical evidence that even small amounts of vertical disparity degrade stereopsis (Duwaer & van den Brink, 1982; Farell, 2003; McKee, Levi, & Bowne, 1990; Westheimer, 1978, 1984) and that stereopsis fails completely for elevated eye positions where the epipolar lines are rotated significantly away from the horizontal (Schreiber, Crawford, Fetter, & Tweed, 2001). All this suggests that the brain, instead of representing all epipolar lines equally, makes the very sensible choice of concentrating on near-horizontal epipolar lines. The finite width of receptive fields means that the epipolar lines are broadened into narrow bands or strips, as suggested by Schreiber et al. (2001), which means that small amounts of vertical disparity can be tolerated even by detectors on horizontal epipolar lines.

In this article, we shall take this one step further and suggest that any variation in vertical-disparity tuning may be

simply noise that is ignored when the population activity is read out. In this picture, the brain assumes that all its disparity detectors lie exactly on the epipolar lines appropriate to primary position, and if any are actually tuned to small nonzero vertical disparities, that is ignored. As a matter of terminology, we shall reserve the term “vertical-disparity detector” for a sensor that is tuned to a nonzero vertical disparity *and* whose vertical-disparity tuning is taken into consideration in the readout. Thus, in this picture, the brain would contain no vertical-disparity detectors. If, as we propose, the precise vertical-disparity tuning of individual neurons is ignored, then any scatter in vertical-disparity

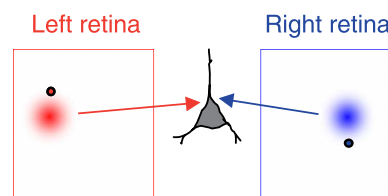


Figure 2. Gray neuron = binocular disparity sensor, receiving input from left- and right-eye receptive fields (colored blobs). The sensor is tuned to a horizontal disparity given by the offset between its left and right receptive fields and is tuned to zero vertical disparity. Small circles show left- and right-eye images of a stimulus with vertical disparity. This sensor is optimally tuned to the horizontal disparity of the stimulus, and it would respond maximally if the stimulus vertical disparity were zero. However, because the images are offset vertically, they cannot both fall on the center of the receptive fields, and thus, the sensor will not respond maximally.

tuning simply represents noise—for a given vertical disparity, the sensor's response is either slightly larger or slightly smaller than the visual system would have expected. Thus, for simplicity, we shall consider a model stereo system in which all disparity detectors are tuned to *exactly* zero vertical disparity. That is, their receptive fields are identical in profile and are located at the same vertical position in both retinæ. Such pure horizontal-disparity neurons (Figure 2) can still sense binocular correlation between the two eyes' images, even when there is a small amount of vertical disparity: They tolerate vertical disparities that are small compared to the receptive-field size. When the vertical disparities are too large, of course, they simply perceive the images in the two eyes as being uncorrelated. This mirrors the psychophysical evidence that stereo performance declines as vertical disparity increases (Duwaer & van den Brink, 1982; Farell, 2003; McKee et al., 1990; Westheimer, 1978, 1984).

Most visual scientists would immediately dismiss this simple model as a model of human stereopsis. They would point to the mountain of psychophysical evidence demonstrating that vertical disparity profoundly influences both eye movements and depth perception. These effects are of two main types: (1) Appropriate patterns of vertical disparity influence the depth perception caused by horizontal disparity (Backus, Banks, van Ee, & Crowell, 1999; Banks & Backus, 1998; Banks, Hooge, & Backus, 2001; Berends & Erkelens, 2001; Berends, van Ee, & Erkelens, 2002; Brenner, Smeets, & Landy, 2001; Clement, 1992; Duke & Howard, 2005; Friedman, Kaye, & Richards, 1978; Frisby et al., 1999; Gillam, Chambers, & Lawergren, 1988; Gillam & Lawergren, 1983; Helmholtz, 1925; Ito, 2005; Kaneko & Howard, 1996; Ogle, 1952, 1953; Pettet, 1997; Pierce & Howard, 1997; Pierce, Howard, & Feresin, 1998; Rogers & Bradshaw, 1993, 1995; Stenton, Frisby, & Mayhew, 1984; Wei, DeAngelis, & Angelaki, 2003; Westheimer, 1984; Westheimer & Pettet, 1992; Williams, 1970). (2) Uniform vertical disparity evokes corrective vertical vergence movements, even at short latencies, in the direction that reduces the vertical disparity (Allison, Howard, & Fang, 2000; Busetini, Fitzgibbon, & Miles, 2001; Howard, Allison, & Zacher, 1997; Howard, Fang, Allison, & Zacher, 2000; Yang, FitzGibbon, & Miles, 2003). Such phenomena are evidence that vertical disparity is not simply "tolerated" because of the finite width of horizontal epipolar bands; it is actively detected and used in perception. To all previous workers, it has seemed obvious that the stereo system must therefore include true vertical-disparity detectors: That is, the early visual system must contain neurons tuned to a range of vertical disparities, and the vertical-disparity tuning of each detector must be taken into account when decoding its population activity. This expectation has motivated several physiological studies that have looked for disparity-tuned neurons with vertical-disparity tuning clearly different from zero (Durand, Zhu, Celebrini, & Trotter, 2002; Gonzalez et al., 2003; Trotter, Celebrini, & Durand, 2004).

However, in this article, we shall demonstrate that this expectation is not correct. Our simplified model visual system, even containing no vertical-disparity sensors at all, is surprisingly powerful. Because vertical disparity reduces effective binocular correlation, sensors that measure binocular correlation, even if their receptive fields have zero vertical-disparity tuning, can sense the magnitude of vertical disparity in the stimulus. True, individual detectors are not sensitive to the sign of vertical disparity (i.e., whether corresponding features are higher in the left or right eye¹). At first sight, this appears a fatal flaw, ruling out almost all the well-known illusions of vertical disparity, such as the induced effect. But in fact, these illusions depend on the interaction between vertical disparity applied to the stimulus and vertical disparity due to eye position. These reinforce or cancel, depending on their sign, outside the organism, resulting in a characteristic pattern of vertical disparity magnitude—and hence binocular correlation—across the visual field. A visual system containing only horizontal-disparity sensors can deduce gaze angle and vergence from this pattern, and we demonstrate that such a system is subject to the induced effect. Thus, in fact, our model visual system can experience all the illusions of vertical disparity demonstrated to date. Furthermore, the fact that such information has to be derived from the pattern of sensor response across large regions of the visual field provides a possible reason why vertical disparities, unlike horizontal ones, are pooled over large regions of the visual field (Adams et al., 1996; Howard & Pierce, 1998; Kaneko & Howard, 1996; Stenton et al., 1984). Hence, both classes of psychophysical phenomena could potentially be mediated solely by activity in horizontal-disparity sensors.

In this article, we address the following question: Could all of the perceptual effects of vertical disparity be mediated solely through its effect on *horizontal*-disparity detectors? We shall show that, for all experiments published to date, the answer seems to be yes. Vertical disparity in the stimulus reduces the effective binocular correlation sensed by a population of horizontal-disparity detectors like the one sketched in Figure 2. This allows one to deduce a local map of the unsigned magnitude of vertical disparity. Given this map of magnitudes, we show that the global constraints on stereo geometry make it possible to infer the signs; thus, the full vertical-disparity field—at least for disparities generated by physically possible stimuli—could potentially be deduced from the activity of purely horizontal-disparity detectors. Hence, both classes of psychophysical phenomena could potentially be mediated solely by activity in horizontal-disparity sensors. We conclude that the existing evidence does not conclusively demonstrate that the visual system contains detectors tuned to nonzero vertical disparity. However, without such detectors, then it should be possible to recreate the effects of vertical disparity by suitably manipulating the binocular correlation. So far, we have not been able to achieve this. We suggest that this failure is the most compelling evidence to date that the visual system really does encode vertical disparity.

Methods

Definitions of retinal coordinates, correspondence, and disparity

As the eyes move, epipolar lines move on the retina. A visual system that was capable of a full 2D solution of the correspondence problem at all eye positions would therefore have to include disparity sensors whose receptive fields were located on all possible sets of epipolar lines. In this article, we instead consider a model visual system all of whose disparity sensors lie on the epipolar lines for a single eye position. For simplicity, we choose this reference eye position to be primary position and we choose a coordinate system in which the epipolar lines at primary position are horizontal. This is mathematically convenient, as it means that—irrespective of their position on the retina—our disparity sensors never have any vertical disparity, whereas if we chose the epipolar lines appropriate to a vergence of 5° , our sensors would have vertical disparity depending on their cyclopean position on the retina. This choice is also consistent with psychophysical evidence that the stereo system cannot solve the correspondence problem if the epipolar lines are too far from horizontal (Schreiber et al., 2001).

The literature contains several different definitions of vertical disparity. The following paragraphs define how it is used here. First, we shall need a coordinate system on each retina, as well as a way of bringing the two eyes' coordinate systems into register by defining which points are in *anatomical correspondence*. We define anatomical correspondence such that, when the eyes are in primary position, objects at infinity project to anatomically corresponding points in the two eyes. We define our retinal coordinate frame as drawn in Figure 3A. This employs a Cartesian coordinate system (x, y) on an imaginary plane tangent to the retina at the fovea. Any point P on the retina can be mapped onto this planar retina by drawing a line from the nodal point through the point P and seeing where it intersects the plane (red line in Figure 3A). To describe the position of any point on the retina in angular coordinates, we use (\hat{x}, \hat{y}) , related to (x, y) , as shown in Figure 3B. For example, the blue lines in Figure 3A shows $\hat{y} = -35^\circ$, that is, all points that are 35° below the horizontal meridian. The pink lines in Figure 3A shows $\hat{x} = -35^\circ$, that is, all points that are 35° to the left of the vertical meridian. Anatomically corresponding points in the two eyes have the same coordinates: $\hat{x}_L = \hat{x}_R$, $\hat{y}_L = \hat{y}_R$. We are now in a position to define disparity. Points that are in *stereo correspondence* are viewing the same object in space. The retinal disparity is the difference between the retinal coordinates of stereoscopically corresponding points. For example, if an object projects to (\hat{x}_L, \hat{y}_L) in the left retina and to (\hat{x}_R, \hat{y}_R) in the right, then its horizontal angular disparity is $\Delta\hat{x} = \hat{x}_R - \hat{x}_L$ and its vertical angular disparity is $\Delta\hat{y} = \hat{y}_R - \hat{y}_L$.

Note that, in this coordinate system, when the eyes are in primary position, there is no vertical disparity. Because the eyes are displaced horizontally, an object closer than infinity in general has images at different angles \hat{x} from the vertical meridian in the two eyes: It thus has a horizontal disparity depending on its distance from the observer. However, all objects project to the same angle \hat{y} above the horizontal meridian: There is no vertical disparity when the eyes are in primary position. Once the eyes move away from primary position, objects have, in general, both horizontal disparity and vertical disparity. Roughly speaking, horizontal disparity reflects the position of an object in space, but vertical disparity reflects the alignment of the eyes (Garding, Porrill, Mayhew, & Frisby, 1995; Longuet-Higgins, 1982; Mayhew, 1982; cf. also Figure 9B and 9C). This is why vertical disparity can be used to recover eye position.

Simulations

The details of a mathematically precise description obscure the essential simplicity of the study. We therefore save the equations for the Appendix and here give a conceptual overview of the three steps in our simulations: (1) Generate an example of a three-dimensional visual scene. (2) Calculate the effective binocular correlation sensed by a population of disparity detectors responding to this scene, given that all the detectors are tuned to zero vertical disparity. (3) Estimate eye position from the pattern of variation in this effective correlation across the visual field (the main challenge). This third step exploits the fact that the effect of vertical disparity is to reduce the effective binocular correlation experienced by horizontal-disparity detectors. Roughly speaking, gaze angle can be deduced from the horizontal position at which binocular correlation is maximal, whereas vergence can be deduced from the rapidity with which binocular correlation declines away from this maximum. Note that the symbols used throughout this article are listed for reference in Table 1.

Visual scene

For the simulations shown in Figures 9 and 10, we first generated a visual scene made up of a random set of surfaces. For purposes of illustration, we wanted to choose a complex depth structure (to demonstrate that our approach is not restricted to simple cases like a frontoparallel surface) while setting the depths such that the horizontal disparity would remain detectable ($\pm 1^\circ$ or so) across most of the visual field. To achieve this, we started off with a sphere centered on the midpoint between the eyes. The radius of the sphere was chosen to be close to the fixation distance. Then, we divided the visual field up along polar coordinates, like a dartboard. Points within each segment were brought closer or moved further away along the radius of the sphere, by the same random factor for each area. We then

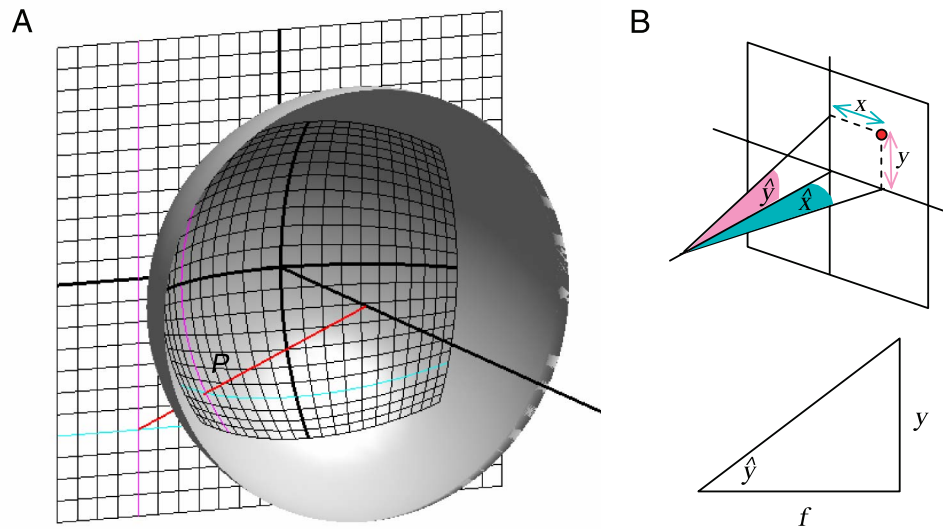


Figure 3. Representing the retinae by planes. (A) Mapping from a planar to hemispherical retina. The red line shows how the point ($\hat{x} = -35^\circ$, $\hat{y} = -35^\circ$) is mapped from the plane onto the hemisphere, by drawing a ray from the nodal point to the plane. The lines $\hat{x} = -35^\circ$ and $\hat{y} = -35^\circ$ are drawn on both the plane and the hemisphere, in pink and cyan, respectively. (B) Converting from retinal position coordinates to angular coordinates. The point (x, y) is shown on the planar retina. Its angular \hat{x} coordinate is the angle defined by the fovea, the nodal point, and the point $(x, 0)$: $\tan \hat{x} = x/f$, where f is the distance from fovea to nodal point; the \hat{y} coordinate can be described in a similar manner.

placed 10,000 dots at random over these surfaces and colored them either black or white at random (the remainder of each surface was gray). The resulting “exploded sphere” is shown in Figure 8. We stress that the precise details of this stimulus are not important; it was merely a simple way of generating a complex visual scene containing many different disparities within a detectable range.

Neuronal response

We calculated the response of a population of model disparity detectors to the visual scene. We used binocular energy-model units (Ohzawa, DeAngelis, & Freeman, 1990). All the neurons used in our simulations have receptive fields with identical profiles in the two eyes (no phase disparity) and located at identical vertical positions in both eyes (no vertical disparity). We did not include any variation in vertical-disparity tuning, although this probably exists in the real visual system. The effect of including this would simply have been to add some random noise to the model. The receptive fields were (in general) located at different horizontal positions in the two eyes, meaning that the neuron was tuned to a nonzero horizontal disparity. The mean position of the receptive field in the two eyes defined the preferred visual direction of the cell, which can be thought of as its location on a notional *cyclopean retina*. To begin with (Figure 9), we consider simplified units with Gaussian receptive fields, representing the overall activity of

neurons with many different orientations and phases. Later (Figures 10, 11, and 12), we consider more realistic units with Gabor receptive fields, in which different orientations and phases are explicitly represented.

We wanted to obtain an estimate of binocular correlation from the activity of these neurons, in normalized units going from 1 (images in the two eyes’ receptive fields are identical) to 0 (images are uncorrelated) to -1 (images are anticorrelated, i.e., identical after polarity inversion). If the neuron is tuned to the disparity of the stimulus, then the binocular correlation it sees is just the binocular correlation of the stimulus. For example, suppose a random-dot stereogram is generated, in which stereoscopically corresponding dots have probability p of being the same contrast (both black or both white) and probability $1 - p$ of being opposite contrasts (one black, one white). The binocular correlation of the stimulus is $C_{\text{stim}} = 2p - 1$. To see how C_{stim} can be estimated from the output of an energy-model neuron, recall that the response of an energy-model unit is $(L + R)^2$, where L and R are the outputs from the left- and right-eye receptive fields, respectively (see the Appendix for details). This can be divided into two components: a sum of monocular terms $M = L^2 + R^2$ and a binocular component $B = 2LR$. We assume that the visual system is able to keep track of both these components separately. This could be done, for example, by differencing the outputs of matched tuned-excitatory and tuned-inhibitory neurons to estimate B and summing the same outputs to estimate M .

Symbol	Description	Application
$C_{\text{stim}}(x_c, y_c)$	Binocular correlation of the stimulus, as a function of position on the cyclopean retina	Equations 17 and 19
C	Effective binocular correlation sensed on average by a cell	Equations 18 and 19
D	Vergence angle, $H_R - H_L$	Figure A1(B) and Equation 1
$D_{1/2}$	Half the vergence angle, $(H_R - H_L)/2$	
Δx	Horizontal position disparity, in distance on a planar retina, $x_R - x_L$	Equation 13
$\Delta \hat{x}$	Horizontal angular disparity, in degrees, $\hat{x}_R - \hat{x}_L$	Equation 10
Δy	Horizontal position disparity, in distance on a planar retina, $y_R - y_L$	
$\Delta \hat{y}$	Horizontal angular disparity, in degrees, $\hat{y}_R - \hat{y}_L$	Equation 10
f	Focal length of eyes	Figure A1(B) and Equation 7
H_c	Cyclopean gaze direction, $(H_R + H_L)/2$	Equation 2
H, H_L, H_R	Helmholtz azimuthal angle, Helmholtz azimuthal angle of the left eye, Helmholtz azimuthal angle of the right eye, respectively, in degrees to the left	Figure A1(B) and Equation 3
$l_{1/2}$	Half the interocular distance	Figure A1(B)
V, V_L, V_R	Helmholtz elevation, Helmholtz elevation of the left eye, Helmholtz elevation of the right eye, respectively, in degrees downward	Equation 3
X	Horizontal position in head-centered space, in Cartesian coordinates	Figure A1(A) and Equation 8
\hat{X}	Horizontal position in head-centered space, in degrees to the left	Figure A1(A) and Equation 8
x	Horizontal retinal position, in distance on a planar retina	Figures 3, A1(B), and Equations 4 and 7
x_c	Horizontal cyclopean location, in distance on a planar retina, $(x_R + x_L)/2$	
\hat{x}	Angular vertical retinal position, in degrees	Figures 3, A1(B), and Equation 7
\hat{x}_c	Horizontal angular cyclopean location, in degrees, $(\hat{x}_R + \hat{x}_L)/2$	Equation 11
Y	Vertical position in head-centered space, in Cartesian coordinates	Figure A1(A) and Equation 8
\hat{Y}	Vertical position in head-centered space, in degrees above the horizontal	Figure A1(A) and Equation 8
y	Vertical retinal position, in distance on a planar retina	Figures 3, A1(B), and Equations 4 and 7
\hat{y}_c	Vertical angular cyclopean location, in degrees, $(\hat{y}_R + \hat{y}_L)/2$	Equation 11
Y	Vertical position in head-centered space, in Cartesian coordinates	Figures 3, A1(B), and Equation 7
\hat{Y}	Vertical position in head-centered space, in degrees above the horizontal	Figure A1(A) and Equation 8
Z	Distance in front of observer, in Cartesian head-centered coordinates	Figure A1(A)

Table 1. Symbols used in this paper, with brief descriptions and where they are defined.

The ratio B/M provides a measure of the binocular correlation sensed by the neuron. For example, if $\langle B \rangle$ and $\langle M \rangle$ represent, respectively, the expected value of the binocular and monocular components, averaged over many random-dot patterns with the cell's preferred disparity, then the ratio $\langle B \rangle / \langle M \rangle$ will be equal to the binocular correlation C_{stim} of the stimulus. In this article, we consider only stimuli with 100% correlation. In this case, if the disparity of the stimulus perfectly matches the disparity tuning of the cell, $\langle B \rangle / \langle M \rangle$ will be 1. If there is a mismatch between the cell's preferred disparity and the disparity of the stimulus, then the value of $\langle B \rangle / \langle M \rangle$ will be smaller, reflecting the smaller effective binocular correlation of the stimulus within the receptive field (although the stimulus correlation at the correct disparity is still 100%). For a sensor with Gaussian receptive fields, like that shown in Figure 2, the value of $\langle B \rangle / \langle M \rangle$ falls off as a Gaussian function of the difference between the cell's preferred disparity and that of the stimulus, with a standard deviation equal to $\sqrt{2}$ times the

standard deviation of the receptive field (Equation 19; Figure 9D).

To reduce the run time, the simulations presented in this article include only sensors tuned to the horizontal disparity of the stimulus at the center of their receptive field because these are the most informative in constraining eye position. (We have also looked at recovering eye position using a full population tuned to many different horizontal disparities and verified that this works in essentially the same way.) In this subpopulation of optimal sensors, if there were no vertical disparity, the effective correlation would always be 1 (apart from small reductions at occluding edges, where the horizontal disparity changes abruptly within a single receptive field). However, because the neurons in our simulations are all tuned to zero vertical disparity, any vertical disparity in the stimulus will reduce the effective binocular correlation that they experience. The amount of the reduction depends on the magnitude of vertical disparity relative to the receptive-field size (Equation 20). Thus, the effective

binocular correlation reported by a population of purely horizontal-disparity detectors reflects the magnitude, but not the sign, of vertical disparity. As we shall see in the [Results](#) section, the reduction in binocular correlation occurs in a characteristic way across the retina, reflecting the position of the eyes. It is this that makes it possible to recover eye position from this population.

So far, we have considered the ratio $\langle B \rangle / \langle M \rangle$, where $\langle B \rangle$ and $\langle M \rangle$ represent the *expected* value of the binocular and monocular components, respectively, averaged over all possible random-dot stimuli. Obviously, this is not available to the brain when it views a single stimulus. For any individual neuron responding to a single random-dot image, the value of B/M is extremely “noisy,” reflecting random variations in the pattern of black and white dots. This means that an estimate of eye position that uses only one neuron at each point in the visual field is noisy and unreliable. However, at each position in the visual field, the brain contains a multitude of neurons tuned to a range of orientations, spatial frequencies, pattern of ON/OFF regions, and so forth. Combining information from all these neurons greatly improves the estimate of binocular correlation and, hence, of eye position. To demonstrate this, in our later simulations ([Figures 10](#), [11](#), and [12](#)), we calculate the responses of 30 neurons at each point on the cyclopean retina, covering three preferred orientations and 10 preferred phases (see the [Appendix](#) for details). We calculate the total binocular component, $\Sigma_n B_n$, and monocular component, $\Sigma_n M_n$, for these neurons and estimate the binocular correlation from their ratio, $(\Sigma_n B_n) / (\Sigma_n M_n)$. This is far less noisy than the ratio B_n / M_n for any one neuron ([Figure 10](#)) and approximates the expected value $\langle \Sigma_n B_n \rangle / \langle \Sigma_n M_n \rangle$.

Estimating eye position from the response of correlation sensors

We assume that the brain has been able to solve the stereo correspondence problem to arrive at an accurate map of horizontal disparity at each point in the image. Note that, even for stimuli with vertical disparity, the *horizontal* correspondence problem can still be solved from a population of purely horizontal-disparity detectors. Roughly speaking (ignoring the problem of false matches, which arises when the stimulus disparity is not constant), the horizontal disparity of the stimulus can be deduced from the preferred horizontal disparity of the units that are reporting the largest binocular correlation. Any vertical disparity in the stimulus will reduce the size of this peak binocular correlation but will not affect which sensor is reporting the peak. In practice, for a realistic visual scene containing objects at different depths, the false-matching problem is nontrivial and requires additional constraints such as a prior preference for smooth surfaces. However, this need not concern us here. It is clear that the brain is able to solve this correspondence problem with great accuracy, and the important point is that any vertical disparity in the stimulus need not affect the solution of the horizontal correspondence problem. Thus,

we can assume that the brain has access to the horizontal-disparity field of the stimulus.

Now, if both the horizontal-disparity field and the eye position are known, the vertical disparity at any retinal location can be calculated ([Equation 16](#)). This vertical-disparity field predicts the effective correlation reported by the corresponding horizontal-disparity detectors: Larger vertical disparity at a particular region of the visual field reduces the effective correlation reported there. Thus, given the 2D disparity field, we can predict the *expected* value of $\langle \Sigma_n B_n \rangle / \langle \Sigma_n M_n \rangle$, where the angle brackets represent averaging over all possible random-dot patterns with the given disparity field, and compare this to the actual value $(\Sigma_n B_n) / (\Sigma_n M_n)$, which our neuronal population gave us for the particular random-dot pattern to which it was exposed. Our fitting routine searches for the eye position that best predicts the observed pattern of response magnitudes.

We used the MATLAB routine `fminsearch`, adjusting gaze angle and vergence to minimize the sum of the squared errors between the predicted and actual correlation at each point in the visual field. Calculating the expected correlation exactly is prohibitively slow, although we restrict ourselves only to the best matching sensor at each position, because for each sensor, we must integrate the stimulus disparity across its receptive field ([Equation 18](#)) to obtain its expected response. To speed up the fitting procedure, we therefore did the fitting under the approximation that stimulus disparity is constant across the receptive field ([Equation 21](#)). The main effect of this approximation was to ignore the lower effective correlation sensed by our V1-like model neurons when there was a depth discontinuity within the receptive field of a neuron (compare [Figures 10C](#) and [10D](#)). Tests indicated that this did not significantly affect our estimates of gaze angle and vergence.

Results

The induced effect does not prove that vertical disparity is encoded

The idea that vertical disparity plays a role in perception was first introduced by Helmholtz (1925). However, Helmholtz’s conclusions were later challenged (Hering, 1942; Hillebrand, 1893), and the accepted view, summarized by Ogle (1954) and Westheimer (1978), became that (1) vertical disparities made no contribution to depth perception and (2) their sign could not be discriminated. This orthodoxy was overturned with Ogle’s (1938) demonstration that vertical magnification of one eye’s image produces a sensation of slant about a vertical axis, with a frontoparallel stimulus appearing closer to the observer on the side of the magnified eye. The only effect of changing the eye that receives the magnification is to invert the sign of the vertical-disparity field, but this also inverts the direction of the perceived slant. Thus, this illusion is compelling evidence

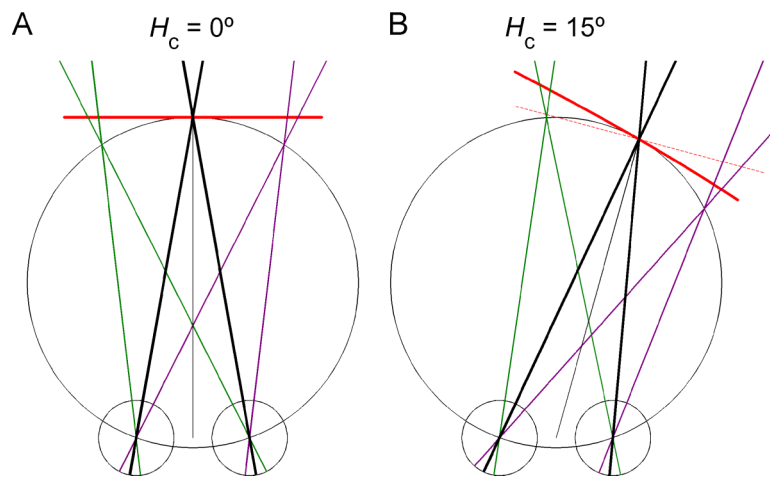


Figure 4. Sketch of how a gaze misestimate produces a percept of slant. The heavy black rays mark the fixation point in both panels, whereas the lighter black line is the cyclopean gaze direction. The purple and green rays mark two additional points with zero horizontal disparity, respectively, to the left and right of fixation. The black circle is the Vieth–Mueller circle of all points with zero horizontal disparity; this is a circle through both eyes and the fixation point. (A) A frontoparallel plane viewed straight on (red) subtends uncrossed disparities that are symmetric on either side of fixation. (B) To obtain the same pattern of horizontal disparities when the eyes are looking off to the side requires the plane to be tilted (thick red line) away from the gaze-normal (dashed red line). For illustrative purposes, this figure uses a large value of vergence: 20° .

that the perceptual system also makes use of information that depends on the sign of vertical disparities in the stimulus. However, we shall show below that it is possible to detect the sign of an induced effect without using sensors tuned to a range of vertical disparities. The induced effect is equally compatible with a visual system that contains only pure horizontal-disparity detectors like that sketched in Figure 2. Thus, the induced effect is *not* evidence that the visual system contains a population of vertical-disparity detectors. Before we demonstrate this, it will be helpful to review the current literature on how the induced effect produces its depth illusion.

Gaze direction can be deduced from vertical disparity

Probably the most widely accepted explanation of the induced effect is that it reproduces the 2D disparity field which would be produced if the eyes were looking off to one side and if the object were slanted about a vertical axis (Backus et al., 1999; Gillam & Lawergren, 1983; Mayhew, 1982; Mayhew & Longuet-Higgins, 1982; Petrov, 1980; Porrill, Frisby, Adams, & Buckley, 1999). The vertical-disparity field indicates that the gaze must be oblique, and the horizontal field indicates that the surface must be slanted toward the magnified eye (Gillam & Lawergren, 1983; Howard & Rogers, 1995; see Figure 4). This, of course, does not explain why the surface is perceived as directly ahead, rather than off to one side (Banks, Backus, & Banks, 2002, but see Berends et al., 2002); the assumption is that the visual system does not construct a single, internally consistent global model of scene and eye position but uses different (and possibly inconsistent) heuristics to estimate visual parameters such as slant/slant, distance, and so forth.

The induced effect, then, arises because the vertical-disparity field produced by vertically magnifying one eye's image is almost identical to that produced by an oblique gaze angle. This fact can be appreciated very simply by considering how a single square projects onto the retina. We begin by considering the vertical disparities produced by oblique gaze (Figures 5A and 5B). The perspective diagrams in the upper row show a frontoparallel square directly in front of the observer, drawn in green, projected onto the two planar retinæ, drawn in red for the left eye and blue for the right. The fixation point is indicated in black. In Figure 5A, the observer is fixating on the midline; in Figure 5B, the observer is looking off 5° to the left. In the bottom row, the two planar retinæ are shown face-on and superimposed, with the left eye's image again drawn in red and the right eye's image in blue. Points on the square have, in general, both horizontal disparity and vertical disparity. For both gaze angles, there is one horizontal position where the vertical disparity is zero. This is where the left and right images superimpose, so that the red and blue lines cross over. When the eyes are fixating the middle of the square, this locus of zero vertical disparity is on the vertical meridian of the retina (Figure 5A). When the eyes are fixating the square 5° from its midline, the locus is 5° away from the vertical meridian (Figure 5B).

Figure 6 examines this in more detail, showing how vertical disparity varies across the retina when the eyes view a frontoparallel plane either straight on (Figure 6A) or obliquely (Figure 6B). At each location, pixel color represents the vertical disparity at the corresponding point on a cyclopean retina. To generate these plots, take a point on a frontoparallel plane, say a corner of the green square in Figure 5, and work out where its image would strike each

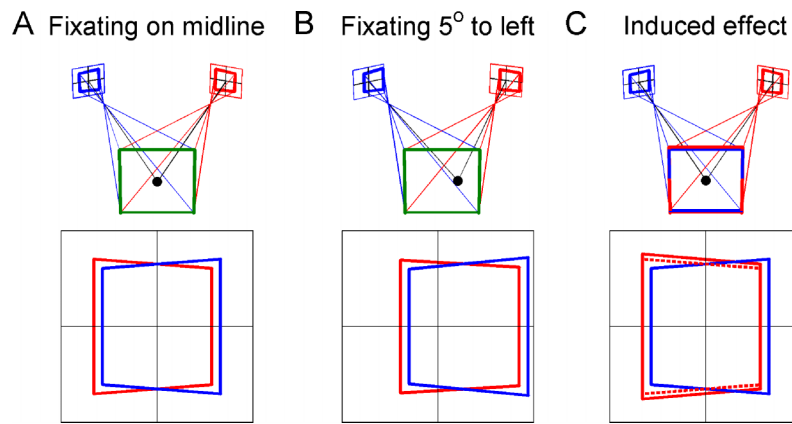


Figure 5. Retinal images of a frontoparallel square, viewed straight on (A), obliquely (B), and with an induced-effect vertical magnification (C). For clarity, in this example, we chose a very large vergence angle, $D = 40^\circ$. The eyes are fixating the plane of the square. The distance of the plane from the eyes is 1.4 times the interocular distance.

retina. The difference between the two \hat{y} coordinates gives the vertical disparity, used to pick a pseudocolor, and the mean of the two \hat{x} and \hat{y} coordinates gives the position on the cyclopean retina, specifying where to plot this pseudocolor. These more detailed maps show the same features that were already visible in Figure 5. It is clear from the retinal diagrams in Figure 5A that the vertical disparity varies in sign across the image, whereas the horizontal disparity is the same for all corners of the square. At the top left and bottom right of the retina, the vertical disparity is negative (left image above right); at the top right and bottom left, it is positive. On the vertical and horizontal meridians of the retina, the vertical disparity is zero. The same pattern is visible in Figure 6A, where the eyes are fixating the midline as in Figure 5A. When the eyes move 5° to the left (Figures 5B and 6B), the whole vertical-disparity field shifts 5° across the retina. The locus of zero vertical disparity is no longer the vertical meridian but the line 5° to the right of the meridian. The vertical-disparity fields here were calculated for a frontoparallel plane. However, the vertical-disparity field is actually rather insensitive to the position of objects in space (this is one of the advantages of our retinal coordinate frame). If we calculated the vertical-disparity field for objects at different distances, the horizontal-disparity field would obviously reflect the depth of the objects, but the vertical-disparity field would be very similar to that shown here. This is clear, for example, in Figure 9C, where the vertical-disparity field varies smoothly, showing none of the “dartboard” structure of the visual scene, in contrast to the horizontal-disparity field (Figure 9B). As noted by numerous previous workers, the vertical-disparity field largely reflects eye position, rather than stimulus location (Garding et al., 1995; Longuet-Higgins, 1982; Mayhew & Longuet-Higgins, 1982). Thus, eye position can be recovered from the vertical-disparity field. As is apparent from Figure 6B, the gaze angle can be read off from the locus of zero vertical disparity.² The vergence can also be deduced, from the rate at which vertical disparity increases away from this locus.

Numerous psychophysical studies show that the brain makes some use of the vertical-disparity field in calibrating the information available from horizontal disparity.

The induced effect mimics oblique gaze direction

At first, it is not obvious why the induced effect should mimic the effect of a shift in gaze angle. After all, oblique gaze shifts the images horizontally across the retina (Figures 5A and 5B), whereas in the induced effect, one eye’s image is magnified vertically. The key is that the vertical magnification is simply what is applied to the *stimulus*. This combines with the vertical-disparity field caused by the viewing geometry—if the eyes are not in primary position—to yield the vertical disparity actually experienced on the retina. Once this is realized, the similarities between the induced effect and oblique gaze become clear. This is illustrated first of all in Figure 5C. Here, as in Figure 5A, the eyes are fixating the center of the square, on the midline. But now, the square presented to the left eye has been magnified vertically: Each Y coordinate has been multiplied by 1.08. The plot at the bottom of Figure 5C shows the retinal image in the two eyes; the red dotted lines show the original, unmagnified image for comparison. Note that the vertical magnification has shifted the locus of zero vertical disparity. Whereas before, the red and blue lines crossed on the vertical meridian, now they cross to the right of the vertical meridian, just as if the eyes were gazing to the left (Figure 5B). Thus, it is already clear that vertically magnifying one eye’s image, as in the induced effect, mimics oblique gaze (Mayhew, 1982; Ogle, 1964).

Figures 6C and 6D show these results more formally. Figure 6C shows the induced-effect vertical-disparity field as it would be experienced if the eyes were in primary position. In primary position, the viewing geometry produces no vertical disparity, and the vertical disparity experienced on the retina is just that applied to the stimulus. The disparity field is zero along the horizontal meridian, and its

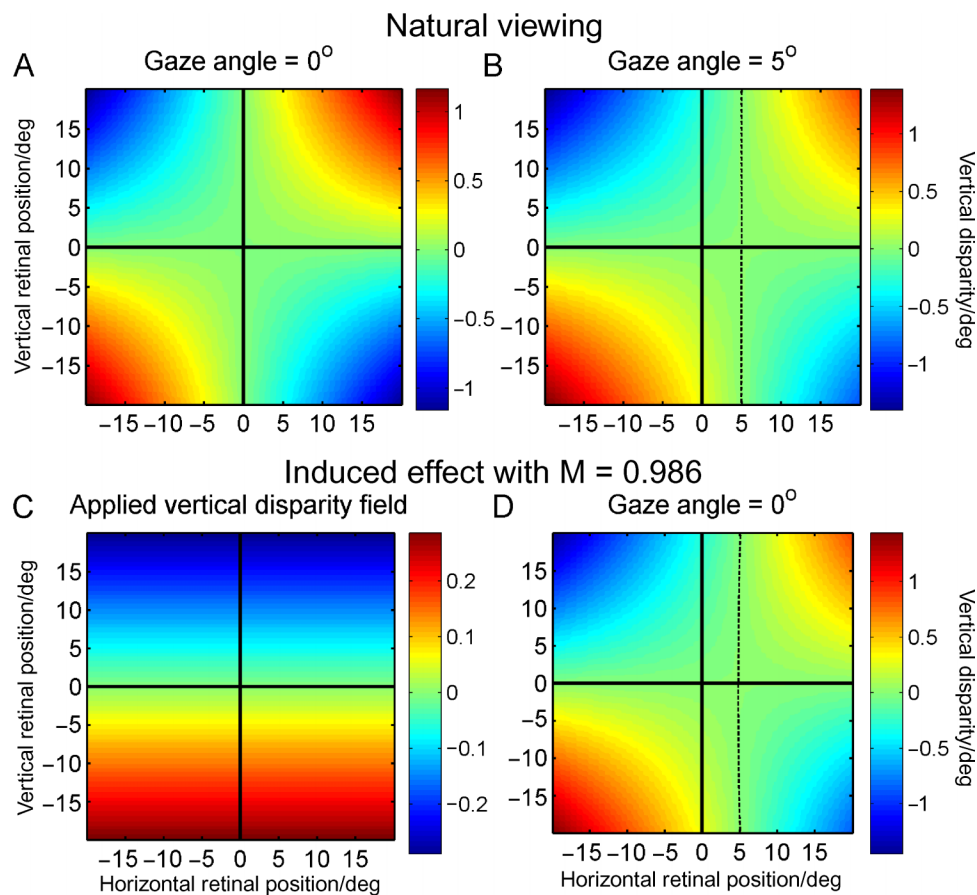


Figure 6. Vertical magnification mimics the vertical-disparity field produced by oblique gaze angle. Panels A and B show the vertical-disparity field of a frontoparallel plane under natural viewing, when the eyes are either (A) fixating the midline or (B) looking 5° to the left of the midline, with a vergence angle of 10° . Panels C and D show the effect of vertical magnification. Here, the right eye's image has been shrunk vertically and the left eye's image expanded vertically. Panel C shows the *applied* vertical-disparity field in the induced effect, that is, what would be experienced on the retina if the eyes were in primary position. Panel D shows the vertical disparity actually produced on the retina by this vertical scaling when the eyes are viewing the midline with a vergence of 10° . Retinal vertical-disparity field produced by the induced effect (D) is almost indistinguishable from that produced by oblique viewing (B). As in Figure 5, interocular distance $l = 6.3$ cm; plane is at $Z = 8.65$ cm. Vergence angle $D = 10^\circ$ in Panels A, B, and D; $D = 0^\circ$ in Panel C. Gaze angle $H_c = 0^\circ$ in Panel A, C, and D; $H_c = 5^\circ$ in Panel B. The induced effect was applied symmetrically: Y coordinates in the left eye were divided by \sqrt{M} , whereas those in the right eye were multiplied by \sqrt{M} , where the magnification factor $M = 0.94$. Solid black lines show the horizontal and vertical retinal meridians; dashed line in Panels B and D shows locus of zero vertical disparity.

magnitude increases with vertical position, with opposite signs in the upper and lower halves of the retina. However, if the eyes are not in primary position, but are converged, then the vertical-disparity field experienced at the retina reflects not only the vertical disparity artificially applied to the stimulus (Figure 6C) but also the vertical-disparity field due to the geometry (Figure 6A). The result is shown in Figure 6D. To a good approximation, it is simply the sum of the disparity fields in Figures 6A and 6C. The positive vertical disparity artificially applied to the bottom half of the retina reinforces the positive vertical disparity that the bottom left of the retina already experiences due to the viewing geometry, while it counteracts the negative vertical disparity on the bottom right of the retina. The effect is to shift the whole pattern over to the right—exactly as would occur

if the eyes moved to the left. Thus, as recognized by Longuet-Higgins (1982), Mayhew (1982), and Mayhew and Longuet-Higgins (1982), the vertical-disparity field produced when the eyes fixate the midline and view an induced-effect stimulus (Figure 6D) is indistinguishable from that produced when the eyes view a normal, nonmagnified stimulus while gazing off to one side (Figure 6B). However, if the horizontal-disparity field produced by a frontoparallel plane is interpreted assuming an oblique gaze angle, the plane is perceived as slanted away from gaze-normal, as shown in Figure 4. Current explanations of the induced effect argue that this is why vertical magnification leads to the perception of slant (Backus et al., 1999; Berends et al., 2002; Gillam & Lawergren, 1983; Mayhew, 1982; Mayhew & Longuet-Higgins, 1982; Petrov, 1980; Porrill et al., 1999).

Vertical-disparity detectors are not needed to recover gaze direction

To summarize, gaze angle can be recovered from the vertical-disparity field. The induced effect misleads the brain by causing a vertical-disparity field that is usually associated with oblique gaze. The induced effect and similar perceptual consequences of vertical disparity have therefore been accepted as proof that the brain detects and uses vertical disparity. This has, for example, motivated physiological studies searching for vertical-disparity detectors (Durand et al., 2002; Gonzalez et al., 2003, 1993). Notice, however, that the sign of vertical disparity is quite unnecessary for the extraction of gaze parameters. Figure 7 shows the magnitude of the absolute value of the vertical-disparity field shown in Figure 6B; it is still easy to see that the gaze angle is 5° . We now show that the magnitude of vertical disparity can be deduced from the activity of purely horizontal-disparity detectors, without the need for any specific vertical-disparity detectors. Suppose the brain's disparity sensors measure binocular correlation at different disparities. Suppose further that the entire population is tuned to zero vertical disparity (although to a range of horizontal disparities), so that the effect of a vertical disparity is to reduce the binocular correlation sensed by horizontal-disparity detectors (Figure 2). Provided it could solve the correspondence problem for horizontal disparity, the brain could still deduce the magnitude of vertical disparity, from the reduced response of the optimally responding disparity sensors. For example, the sensor shown in Figure 2 is the optimally responsive sensor for the stimulus shown because no members of the population have the correct vertical disparity and because this sensor's receptive fields are appropriate for the horizontal disparity of the stimulus. However, even this optimal response would be less than maximal because the vertical disparity means that the stimulus is not identical in both receptive fields. Because the disparity sensors measure correlation, this reduction is not confounded with variations in luminance and so forth. The brain could deduce the magnitude—but not the sign—of the vertical disparity from this reduced response (quantified in Equation 21). As we have seen (Figure 7), this is sufficient to recover eye position.

Figure 9 shows two quantitative examples of how binocular correlation is affected by stimulus disparity. To generate a complex depth scene to use as an example, we divided the visual field radially and azimuthally into small surfaces with random depths, as shown in Figure 8. Figure 9A shows a horizontal slice through this visual scene, for two different eye positions. In the upper row, the eyes are looking 2° off to the left, with a vergence angle of 3.5° . In the bottom row, the eyes are looking 5° to the right, with a vergence angle of 8° . Figures 9B and 9C show the horizontal and vertical disparity maps for the whole visual field, for the two different eye positions shown in Figure 9A. The axes show cyclopean retinal location, that is, mean position in the two retinæ. Both maps reflect eye position: Obviously, the horizontal disparity of each surface segment depends on whether the

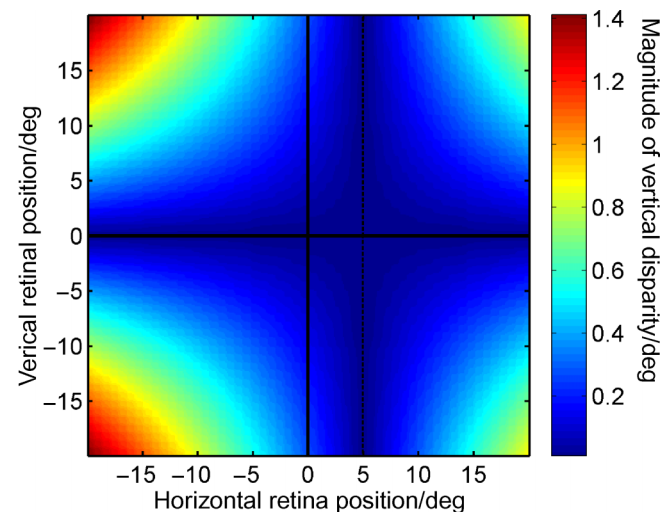


Figure 7. Magnitude of vertical disparity, for natural viewing with gaze angle $H_c = 5^\circ$ and vergence angle = 10° . This vertical-disparity field was shown in Figure 6B. The heavy black lines show the horizontal and vertical meridians; the lighter line shows the locus of zero vertical disparity.

horopter is in front of or behind the segment, while the center of the pattern reflects whether the eyes are looking left or right. In addition, the horizontal disparity map reflects the visual scene: The dartboard structure of the visual scene is clearly visible. The vertical-disparity field, on the other hand, essentially depends only on eye position.

Figure 9D shows the expected binocular correlation sensed by units like that illustrated in Figure 2, averaging over all patterns of black and white dots on the exploded-sphere surface (Figure 8). See the Appendix for an explanation of how this correlation measure is obtained (Equation 18). We restrict ourselves to considering only those neurons that are tuned to the horizontal disparity of the stimulus. This is possible because we assume the brain has been able to solve the horizontal correspondence problem. The pseudocolor at each point represents the binocular correlation sensed by a neuron with Gaussian receptive fields centered on this cyclopean position, whose preferred horizontal disparity is the actual horizontal disparity of the stimulus at this cyclopean position. Both receptive fields are at the same vertical position in the retina, which means that all the neurons in this simulation are tuned to zero vertical disparity. If the stimulus had a constant horizontal disparity and no vertical disparity, then these neurons, being tuned to the stimulus disparity, would view corresponding regions of the visual field and would so report 100% binocular correlation. The correlation field in Figure 9D would therefore simply be 1 everywhere. In practice, two effects reduce the sensed correlation below 1: (1) At depth boundaries, there are discontinuities in stimulus horizontal disparity; hence, the simple sensor shown in Figure 2 cannot be perfectly matched to the stimulus horizontal disparity across its receptive field. This reduces the correlation below 1. (2) Where there is vertical disparity,

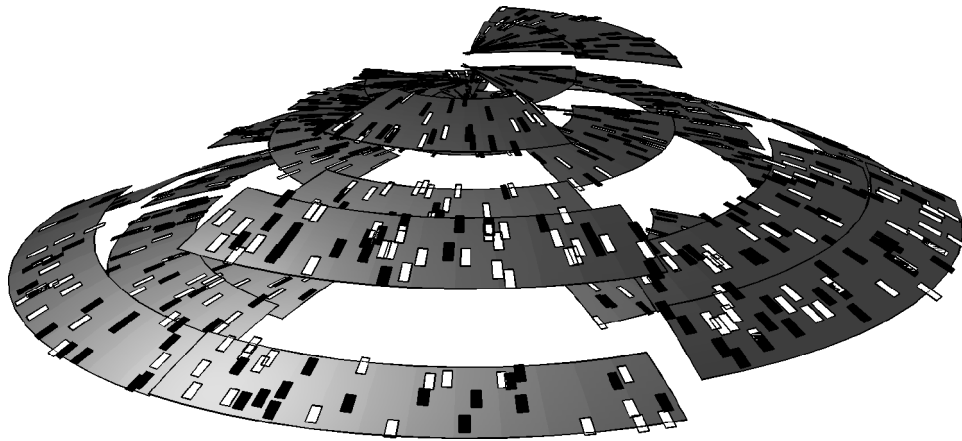


Figure 8. Sketch of visual scene used for simulations in Figure 9. To generate a complex visual scene, a spherical surface is cut up into segments, which are randomly moved nearer or further from the observer, who is at the center of the sphere (Figure 9A). For illustration, the surface segments are shown in gray; only the dots are relevant for the simulations. In the simulations, 50,000 infinitesimal dots were used; for illustration, 1,000 large dots are shown.

the receptive fields in the two retinae are viewing different regions of the image, and thus, again, the correlation falls below 1. The first effect depends on the details of the visual scene. It is responsible for the thin lines of reduced correlation along depth boundaries in Figure 9D but, otherwise, has little effect on the correlation field. The second effect reflects eye position and is much more significant. It imposes a global structure on the correlation field. The correlation reaches 1 only along a cross-shaped region reflecting the locus of zero vertical disparity; away from this cross-shaped locus, the correlation decays away as vertical disparity becomes progressively larger. The details of this global pattern depend on eye position. The horizontal bar of the cross is always along the horizontal retinal meridian because (with no elevation) vertical disparity is always zero for $y = 0$. However, where the vertical bar crosses the X -axis depends on gaze angle. As we saw in Figures 5 and 6, the horizontal location of this locus of zero vertical disparity reveals whether the eyes are looking to right or left of center. This is clearly visible in Figure 9—compare the location of the peak response in the top row, where the eyes are looking left, with that in the bottom row, where they are looking right. The convergence state is also encoded in this correlation field. When the eyes are strongly converged, as in the bottom of Figure 9, the correlation falls off steeply from its peak; where the convergence is less, the rate of falloff is slower. Note that in Figure 9D, the color scales are different for the two rows. However, the contour lines marking vertical disparity are drawn at the same values (multiples of 0.1°) in both cases. The fact that the contour lines are much closer in the bottom row shows that the rate of change is steeper where the eyes are more converged. The rate of falloff depends both on vergence and receptive-field size: Vergence determines the rate of increase of vertical disparity (Figure 9C), whereas receptive-field size determines how much a particular vertical disparity reduces correlation (Equation 20). However, if we know the sensors' receptive-field size, we can

read off both gaze angle and vergence state from the correlation field in Figure 9D.

Figure 9 serves to illustrate the basic ideas. However, it falls short of being a realistic physiological model in two respects. First, the correlation fields plotted in Figure 9D were obtained with binocular energy-model units with Gaussian receptive fields, whereas disparity sensors in the real early visual system have bandpass orientation and spatial frequency tuning. More seriously, the correlation fields in Figure 9D were calculated from theoretical expressions (Equation 18) representing the *average* response over many random-dot patterns, of which Figure 8 is just one example. In reality, the visual system usually has only one stimulus available from which to deduce eye position. Thus, we have yet to demonstrate that eye position can be reliably recovered under these circumstances. In practice, neither of these shortcomings is serious. The Gaussian receptive fields used in Figure 9D can be regarded as representing the sum of receptive fields tuned to many different orientations and phases. Rather than averaging the response of a single sensor over many images, the visual system can reduce variation by averaging the response of many sensors to a single image. Hence, including a realistic range of neuronal receptive fields also solves the problem of noise.

To quantify these ideas and to prove that eye position can still be recovered from the outputs realistically available from the early visual system, we show results with more realistic model neurons in Figure 10. These are binocular simple cells with Gabor receptive fields, constructed in the same way as subunits of the binocular energy model (Ohzawa et al., 1990). We include neurons tuned to three different orientations and 10 different receptive-field phases (although the phase *disparity* was in every case zero). As explained in the Appendix, the response of energy-model units can be divided into a “binocular” component B and a “monocular” component M . To obtain a measure of binocular correlation corresponding to that shown in Figure 9D,

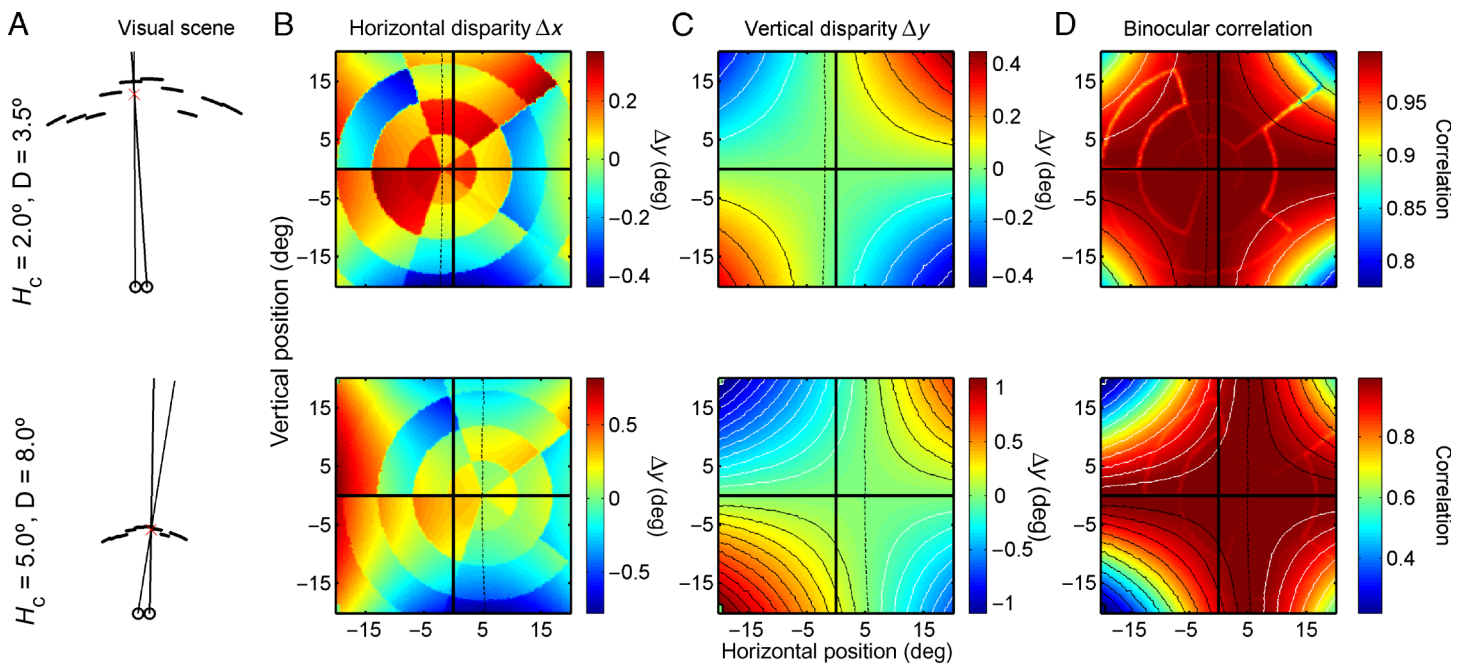


Figure 9. How binocular correlation reflects eye position. (A) Visual scene and eye position viewed from below. Red cross marks fixation. (B and C) Horizontal- and vertical-disparity fields for the stimulus, as a function of horizontal and vertical cyclopean location. Note that the horizontal-disparity field reflects the dartboard depth structure of the visual scene (Figure 8), whereas the vertical-disparity field varies smoothly, reflecting eye position but not the details of the visual scene. (D) Expected value of the binocular correlation (Equation 18) sensed by neurons like that shown in Figure 2, with receptive fields that are isotropic Gaussians ($SD = 0.5^\circ$) and horizontal position disparity equal to the horizontal disparity of the stimulus at that point in the visual field. This expected value requires averaging over all random-dot patterns with the disparity fields shown in Panels B and C. The visual scene is an exploded sphere (Figure 8). The two rows are for two different eye positions. To keep the horizontal disparity of the stimulus mostly within a range that can be detected by human observers, the visual objects are presented close to fixation in both cases, which means they are at different physical distances (the distance scale is the same for both parts of Panel A). Top row: vergence, $D = 3.5^\circ$; gaze direction, $H_c = -2.0^\circ$. Bottom row: vergence, $D = 8.0^\circ$; gaze direction, $H_c = 5.0^\circ$. In Panels B, C, and D, solid black lines mark the vertical and horizontal retinal meridians; the dashed lines mark the locus of zero vertical disparity. Note that this is to the left of the vertical meridian in the top row and to the right in the bottom row, reflecting the different directions of gaze. The contour lines in Panels C and D show vertical disparity, spaced 0.1° apart. Black contour lines are for positive values, and white ones are for negative values. Note that the response falls off much more rapidly in the bottom row, reflecting the larger vergence.

but for a single random-dot pattern, we calculate the values of B and M for every neuron in the population and then divide the sum of all the B s by the sum of all the M s (see the Appendix, Equation 24). The result is shown in Figure 10B. As in Figure 9, the top row is for a gaze angle of -2° and a vergence angle of 3.5° , whereas the bottom row is for a gaze angle of 5° and a vergence angle of 8° . The color scales are the same for all panels in the same row. For comparison, Figure 10A shows the B/M ratio for a single neuron in the population. Because of the chance pattern of black and white dots in the stimulus, this is so noisy that it carries very little information about the eye posture. In contrast, Figure 10C shows the value we would expect to obtain if we averaged over all possible random-dot patterns (Equation 17), completely removing all stimulus-related variation. Clearly, summing over 30 neurons (Figure 10B) has greatly reduced the variability experienced with just 1 neuron (Figure 10A). The response to a single random-dot pattern (Figure 10B) is now very similar to the expected result of averaging the responses

to all possible random-dot patterns (Figure 10C)—and what is important is that it allows us to deduce gaze direction and vergence.

We fed the population response (Figure 10B) into a fitting routine that searched for the gaze angle and vergence that produced the closest match to the observed population response, given the actual stimulus horizontal disparities (assumed to be available to the visual system from an accurate solution of the correspondence problem, not included in this simulation). This procedure is described in the Methods section and in the Appendix. Note that to speed up the algorithm, we calculated the expected response using approximate expressions that ignore the variation in stimulus disparity within a receptive field (Equation 23) rather than the full expressions used in Figure 10C. The best matching approximate response is shown in Figure 10D, together with the fitted eye positions. Clearly, this is very similar to the exact expression (Figure 10C) except that it does not reproduce the lines of low effective correlation along depth discontinuities.

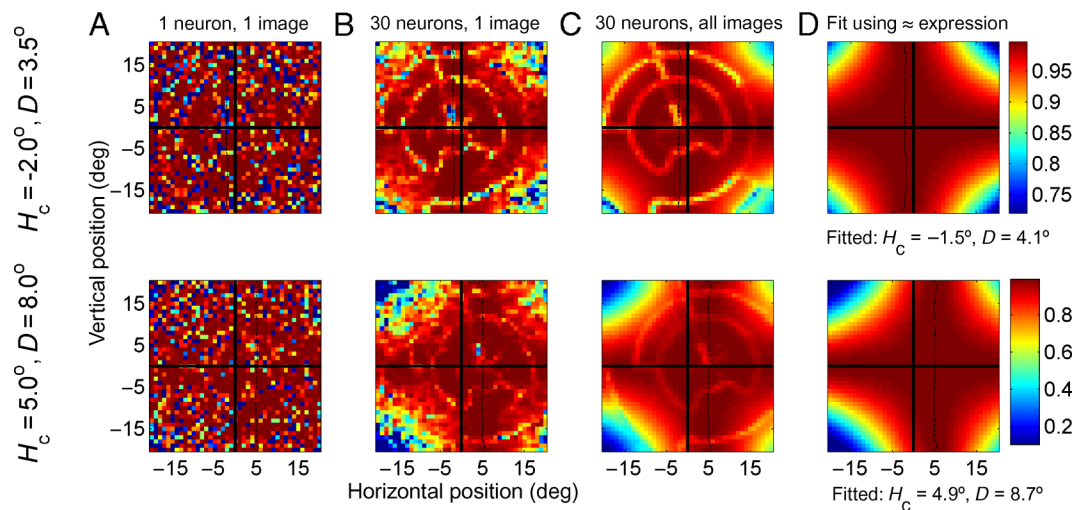


Figure 10. Estimating binocular correlation with real neurons. (A) Binocular correlation field estimated with one neuron, response to a single random-dot image. (B) Binocular correlation field estimated with 30 neurons, response to a single random-dot image. (C) Binocular correlation field expected from 30 neurons, averaging over all possible random-dot images and using the true gaze angle and vergence. (D) Best matching correlation field, using fitted gaze angle and vergence, using an approximation to the value expected from averaging all possible random-dot images. See the [Appendix](#) for a detailed description of how each panel was generated. The cyclopean retina is sampled more coarsely in this figure than in [Figure 9](#), and a larger receptive-field size was used (SD of Gaussian envelope = 1° , instead of 0.5° in [Figure 9](#)). With 30 times as many neurons simulated, this was necessary to reduce the run time to a reasonable duration. The values quoted in the text for quantitative fitting of estimated eye position used the sampling shown here; finer sampling might have produced small improvements in accuracy. A further small technical point is that the sampling actually used a grid on a planar retina. The planar coordinates have been converted to angles for the axis labels in these graphs, although the grid is not strictly uniform on a hemispherical retina. See the [Appendix](#) and [Figure 3](#) for the difference between these coordinates. The neurons' receptive fields are Gabor functions of three different orientations and 10 different phases, with an isotropic Gaussian envelope of $SD = 1^\circ$. As before, they have zero phase disparity, zero vertical disparity, and horizontal position disparity equal to the horizontal disparity of the stimulus at the center of their receptive field.

We repeated each of the simulations shown in [Figure 10](#) for 10 different random-dot images. For the example where the true gaze angle and vergence were -2.0° and 3.5° , the fitted values were $-1.6^\circ \pm 0.2^\circ$ and $3.9^\circ \pm 0.1^\circ$ (mean $\pm SEM$), respectively. For the example where the true values were 5.0° and 8.0° , the fitted values were $5.0^\circ \pm 0.2^\circ$ and $8.5^\circ \pm 0.1^\circ$, respectively. The accuracy of the gaze angle measurement was largely limited by the receptive-field size (SD of Gabor envelope was 1°).

[Figure 11](#) shows the results of repeating this procedure with 11 more eye postures, yielding four values of gaze angle and three values of vergence. For each vergence, a new set of “exploded-sphere” surfaces was generated, placing the sphere roughly at the fixation distance, so that the horizontal disparities close to the fovea were within the detectable human range. For each fit, the set of surfaces was then covered with a new pattern of black and white dots, and eye posture was estimated by fitting the effective correlation derived from the responses of 30 neurons, as in [Figure 10B](#). In [Figure 11](#) (left panel), the fitted gaze angles are shown as a function of the actual gaze angle for three different vergences; the error bars represent the standard deviation over 10 different random-dot patterns. Gaze angle is reconstructed most accurately for large vergence angles because, here, the decay in correlation is fastest: When $D = 15^\circ$, gaze angle is recovered to better than 0.5° . With small

vergence angles, small gaze angles can still be recovered accurately: When $D = 3.5^\circ$, the gaze angles of -2° is recovered with a mean absolute error of 0.7° . However, for large gaze angles, there are significant errors: The two gaze angles $>10^\circ$ are recovered with a mean error of 4° for this small vergence. This is because, as the vergence approaches zero with a large gaze angle, the locus of zero vertical disparity no longer falls within the central 20° simulated here. Vertical disparity and, hence, effective correlation vary progressively less as a function of horizontal position on the retina, and therefore, the fit becomes less and less constrained. However, there is no evidence that the visual system can recover large gaze angles with this accuracy from retinal information; hence, this way of extracting gaze parameters is certainly accurate enough to explain the available psychophysics. In [Figure 11](#) (right panel), the fitted vergence is shown as a function of the actual vergence for four different gaze angles. Vergence is recovered to within 0.5° or so. There is a slight bias: Vergence is systematically overestimated. This may reflect inaccuracies in the fitting assumptions (the least squares fit assumes that errors above and below the expected value are equally likely, which is not the case), as well as the deficiencies of the approximate expression used in the fitting algorithm ([Equation 23](#) in place of the correct expression, [Equation 17](#)). Nevertheless, these results clearly demonstrate that both gaze angle and vergence

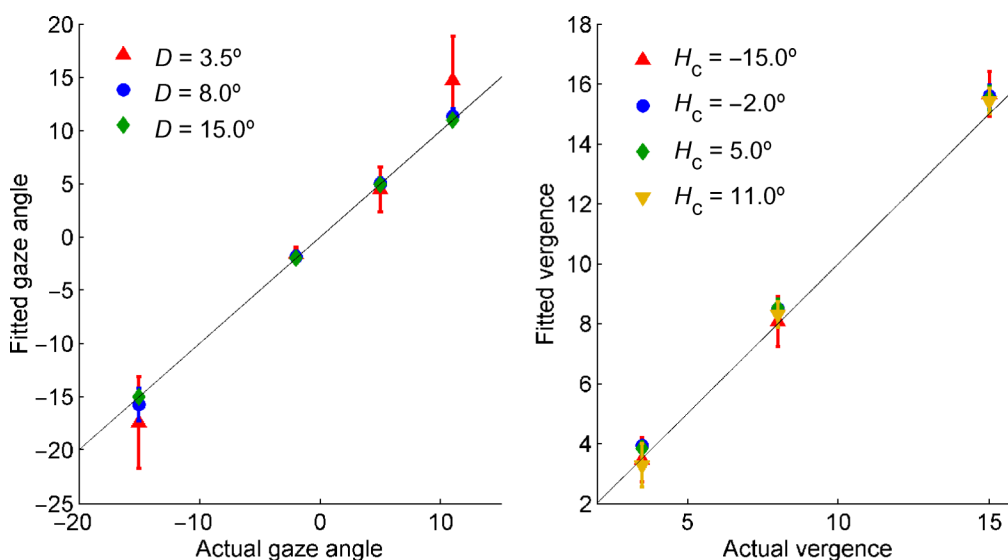


Figure 11. Results of fitting gaze angle and vergence. Symbols and error bars show mean fitted value and standard deviation for 10 different random-dot patterns. For each random-dot pattern, the gaze angle and vergence were estimated from the activity of a population of energy-model simple cells (see example in Figure 10B). At each cyclopean position, only cells tuned to the horizontal disparity of the stimulus were used, but cells with three different orientations and 10 different phases were used. The black line marks the identity, where points would fall if the fits were perfect. The mean absolute error in fitted gaze angle is 2.5° for $D = 3.5^\circ$, 0.7° for $D = 8^\circ$, and 0.3° for $D = 15^\circ$. The mean absolute error in fitted vergence is 0.6° , independent of gaze angle.

can be accurately estimated from the activity in a realistic population of neurons, all tuned to zero vertical disparity.

A stereo system containing only horizontal-disparity detectors could experience the induced effect

We have demonstrated that a visual system containing only pure horizontal-disparity detectors could still accurately deduce gaze angle and vergence from retinal information. It remains to be confirmed that our artificial system is subject to the induced effect. We therefore ran simulations with the conventional induced-effect stimulus. The basic stimulus was a field of black and white dots scattered at random on a frontoparallel screen in front of the simulated observer. The model observer fixated the center of the screen. The vertical coordinate (Y) of the dots on the screen was then multiplied by \sqrt{M} in the stimulus presented to the right eye and divided by \sqrt{M} in the left eye. We then calculated the response of the sensor population to this stimulus and passed this to the fitting algorithm. Sample results are shown in Figures 12A and 12B, where $M = 1.01$. As in Figure 10, at each point on the cyclopean retina, the color shows the response of the sensor that is tuned to the horizontal disparity of the stimulus (although the stimulus here is frontoparallel, its disparity is nonzero in the periphery due to the curvature of the horopter). The heavy black lines show the retinal horizontal and vertical meridians, whereas the dashed line marks the locus of zero vertical disparity on the retina. Figure 12A shows the correlation calculated from the response of 30 neurons, tuned to different orientations and phases, to a single random-dot pattern. Figure 12B

shows the expected correlation that would be obtained if we averaged over all random-dot induced-effect stimuli. Because of the magnification, the region of peak response is shifted away from the vertical meridian, mimicking the effect of oblique gaze. Accordingly, given the population response shown in Figure 12A, our fitting algorithm returned a value of $H_c = -6.9^\circ$, although the actual gaze angle was zero.

The consequences of this erroneous gaze estimate are shown in Figures 12C and 12D. Figure 12C shows the visual scene reconstructed according to Equation 12 from the position of the images in the left and right retinæ, using the correct gaze angle ($H_c = 0^\circ$). Of course, this gives the actual location of the simulated dots in space: on a frontoparallel screen. Figure 12D, on the other hand, shows the visual scene reconstructed using the erroneous estimated gaze angle, $H_c = -6.9^\circ$. The dots now lie on a plane that is slanted away from frontoparallel. This explains the slanted percept experienced in the induced effect.

Corrective vertical vergence movements do not prove that vertical disparity is encoded

We have now confirmed that our model visual system containing only pure horizontal-disparity detectors is still subject to the induced effect, despite the fact that the induced effect is often regarded as evidence for vertical-disparity detectors. But for many vision scientists, the strongest evidence that the human visual system must possess dedicated vertical-disparity detectors is our ability to make corrective

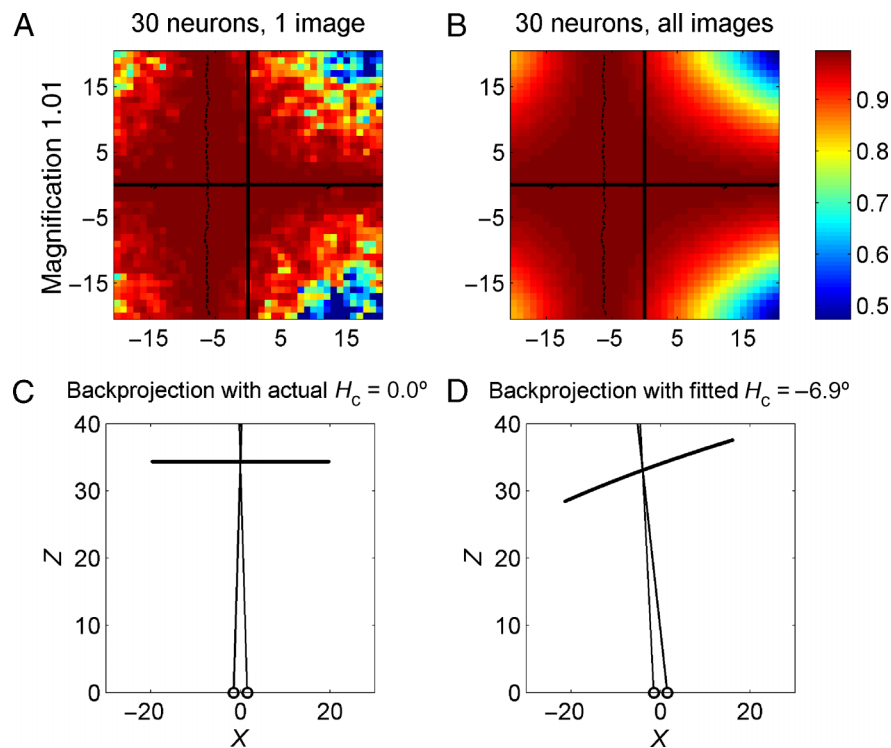


Figure 12. The induced effect. (A and B) Effective binocular correlation with an induced effect stimulus. As in Figure 10, the axes in each plot are angular horizontal and vertical position on the cyclopean retina (in degrees). As before, at each cyclopean position, only the response of the sensor tuned to the horizontal disparity of the stimulus is shown. (A) Response of the sensor population to one particular random-dot pattern. At each cyclopean position, the response reflects the total output of 30 sensors tuned to a range of orientations and phases. (B) Response averaged over all random-dot patterns, thus removing stimulus-dependent “noise.” The color scale is the same for both panels. (C) The actual visual scene and eye position viewed from above. Vergence angle was 5° . The stimulus was made up of dots scattered at random over a frontoparallel screen at the fixation distance, and the gaze angle was zero; hence, the simulated observer was fixating the center of the screen. The right eye’s image was magnified vertically, whereas that of the left eye was shrunk (overall magnification factor, 1.01), although this is not visible because the scene is viewed from above. Estimates of gaze angle and vergence were obtained by fitting the single-image response shown in Panel A (Equation 25). This yielded a vergence angle of 5.1° (true value, 5.0°) and a gaze angle of -6.9° (true value, 0°); that is, the induced effect causes a misestimate of gaze angle. Panel D shows the fitted eye position and the visual scene reconstructed from the retinal stimulus using the misestimated gaze angle. The resulting surface is slanted away from frontoparallel. The neurons’ receptive fields are Gabor functions of varying orientations and phases, with an isotropic Gaussian envelope of $SD = 1^\circ$.

vertical vergence movements. Stimuli in which the two eyes’ images are displaced uniformly across the visual field in opposite vertical directions elicit vertical vergence movements that eliminate the vertical disparity on the retina. This is, presumably, a dynamic mechanism for keeping the eyes correctly aligned and fixated on a single location in space: The stimulus fools the visual system into believing that the eyes are misaligned, and it acts to correct this. We have seen that a population of pure horizontal-disparity detectors also encodes the magnitude of vertical disparity; hence, clearly, such a population could detect the presence of a vertical misalignment. However, it seems obvious that it would be blind as to the direction of the misalignment—whether it was the left eye or the right eye that was too high. The system could certainly find its way to perfect alignment by trial and error, but this cannot explain human performance. Corrective vergence movements always move in the direction that will decrease the error, even if they are short

latency responses (Busetini et al., 2001), showing that the visual system measures not only the magnitude but also the sign of the vertical vergence error. Surely, *this* ability demonstrates that the visual system contains a significant population of vertical-disparity detectors, tuned to a range of vertical disparities.

In fact, vertical-disparity detectors are not necessary even for correction of vertical vergence. Surprisingly, the population of pure horizontal-disparity detectors considered in this article enables one to deduce not only the magnitude but also the sign of vertical vergence error, given the constraints of stereo geometry. To see why, consider the sketch in Figure 13, showing the images of a square as they appear on the two retinae. The first panel just reproduces the situation of Figure 5A, in which the eyes fixate a point on the midline. The Helmholtz elevation is zero for both eyes; thus, there is no vergence error. In the next two panels, the eyes have a vertical vergence error of magnitude 1° . In Figure 13B, the

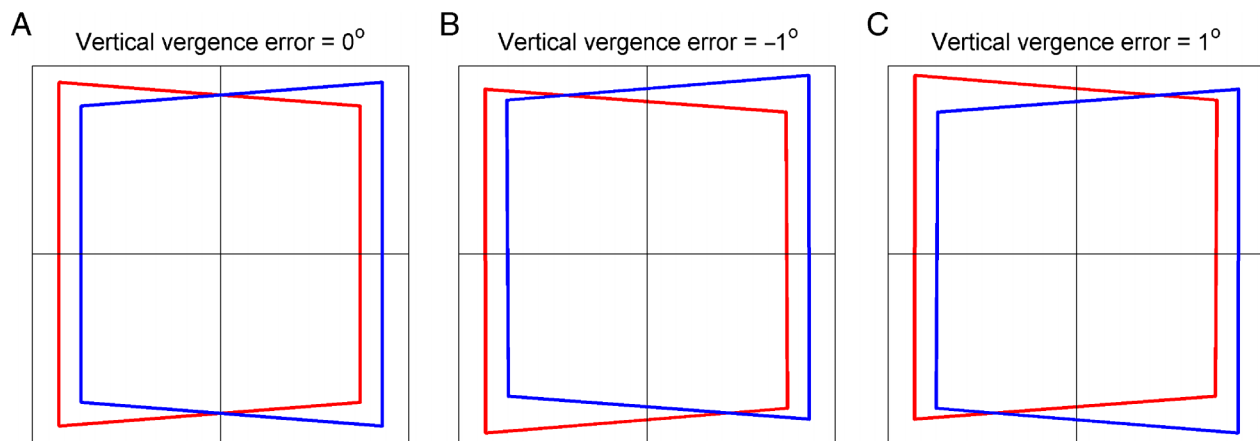


Figure 13. The effect of vertical vergence error on the locus of zero vertical disparity. Similar to Figure 5, except that, here, the azimuthal gaze angle is fixed at 0° and there is no induced effect. In Panel A, the elevation is zero for both eyes. In Panels B and C, there is a vertical vergence error of magnitude equal to 1° (B: $V_L = +0.5^\circ$, $V_R = -0.5^\circ$; C: $V_L = -0.5^\circ$, $V_R = +0.5^\circ$).

left eye is looking down 0.5° , whereas the right eye is looking up 0.5° (right hypervergence). The effect of this, to a good approximation, is to shift the square's image down 0.5° on the left retina and up 0.5° on the right retina. Figure 13C shows a vertical vergence error of the same magnitude but of the opposite sign. Now, consider what this means for the locus of zero vertical disparity, visible in Figure 13, as the places where the two images of the square intersect. When there is no vertical vergence error (Figure 13A), this locus is the vertical retinal meridian, $x = 0$. But when the eyes are misaligned vertically, the intersections move away from the vertical meridian. For right hypervergence (Figure 13B), the top intersection moves to the left of the retina, whereas the bottom intersection moves to the right. However, for left hypervergence (Figure 13C), this pattern is reversed. Now, the locus of zero vertical disparity occurs on the top right and bottom left of the retina. Thus, from tracking the locus of zero vertical disparity, we can deduce the sign of the vertical vergence error.

Figure 14 examines how this effect shows up in the response of our population of horizontal-disparity detectors. The visual scene is the exploded sphere shown in Figure 8, and the details—apart from eye elevation—are the same as in the top row of Figure 9: gaze angle, -2° ; horizontal vergence, 3.5° . However, now, this scene is viewed with a vertical vergence error of 0.2° . The top row of Figure 14 shows the vertical-disparity field experienced on the retina in the presence of vertical vergence error (0.2° right hypervergence in Panel A; 0.2° left hypervergence in Panel B), whereas the bottom row shows the effective binocular correlation field (expected value for Gaussian receptive fields, as in Figure 9D). Right hypervergence means that the image is lower on the left retina, which introduces a positive vertical disparity in our notation (Equation 10). Thus, the whole vertical-disparity field in Figure 14A is increased by 0.2° compared with the situation in Figure 9C, where the eyes were aligned. The dashed line, which was the locus of zero

vertical disparity in Figure 9C, now has a vertical disparity of $+0.2^\circ$. Zero vertical disparity now occurs along the contours marked in white, where vertical disparity would be -0.2° in the absence of vergence error (cf. Figure 9C). These contours occur in the top left and bottom right of the retina, and hence, this is where the effective binocular correlation is maximal (Figure 14C). Figures 14B and 14D show analogous results for left hypervergence. Now, the vertical-disparity field has been reduced by 0.2° everywhere, relative to its value in the absence of vergence error (Figure 9C). Zero vertical disparity and, hence, maximal correlation now occur in the top right and bottom left of the retina.

Thus, binocular correlation fields like those in Figures 14C and 14D could, in principle, be used to derive vertical vergence error and gaze angle. First, correlation will be approximately constant along the horizontal meridian (fluctuations are due to nonuniformities in stimulus depth). If this constant level of correlation is less than 1, then this indicates a vertical vergence error. The magnitude of the vergence error can be deduced from the amount of decorrelation. For the example shown in Figure 14, the correlation along the horizontal meridian is $C_{\max} = 0.96$ for sensors tuned to the horizontal disparity of the stimulus. From Equation 21, we deduce that vergence error is causing a vertical disparity of $2\sigma\sqrt{\ln(C_{\max}^{-1})} = 0.2^\circ$, where σ is the standard deviation of the Gaussian RFs used in the simulation, 0.5° . Thus, we have correctly obtained the magnitude of the vergence error. Its sign can be deduced from the location of the peaks in the population response: If they are on the top left and bottom right of the retina, the vergence is negative. Gaze angle and vergence can also be deduced. To obtain gaze angle, we locate the vertical line along which the response is approximately constant at the same value, 0.96, as it had on the horizontal meridian. The position of this line, here -2° , gives the azimuthal gaze angle. Vergence can be deduced from the rate of change of response away from this cross-shaped contour of constant

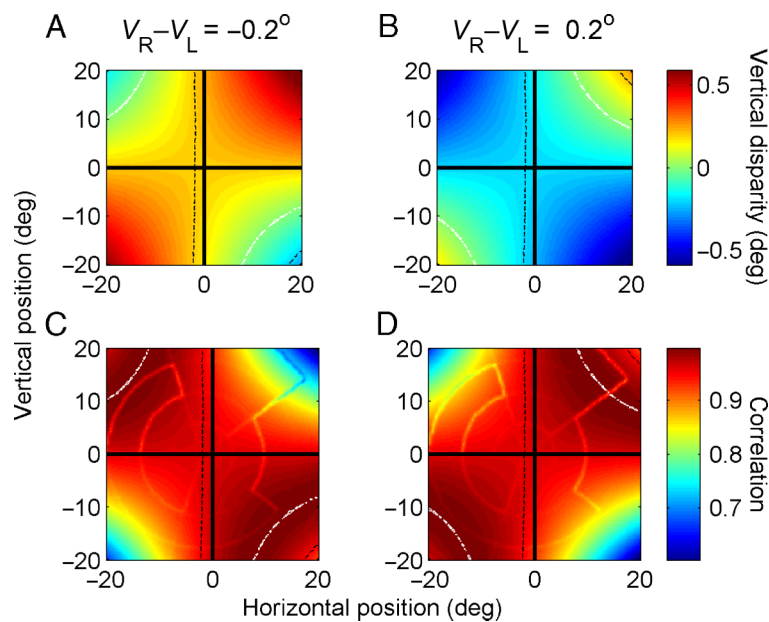


Figure 14. Vertical disparity (A and B) and correlation (C and D) fields in the presence of a vertical vergence error. The visual scene and population details are the same as in the top row of Figure 9. The two columns show results for equal and opposite vergence errors (A and C: 0.2° right hypervergence; B and D: 0.2° left hypervergence; recall that positive V means the eye is looking downward). (A and B) Vertical-disparity field of the stimulus as experienced on the retina. (C and D) Expected binocular correlation reported by sensors with Gaussian receptive fields, averaging over many random-dot patterns (Equation 18). The solid black lines show the horizontal and vertical retinal meridians; the dashed black lines show where the vertical disparity of the stimulus is equal in magnitude and opposite in sign to the vertical vergence error. The white contours show where the vertical disparity of the stimulus is zero on the retina. In both cases, the gaze angle H_c is -2° and the horizontal vergence angle D is 3.5° . The neurons' receptive fields are isotropic Gaussians with an SD of 0.5° and a horizontal position disparity equal to the horizontal disparity of the stimulus at that location in the visual field.

activation. We have not considered an example with elevation, but it is easy to see qualitatively how this would work. With elevation, the horizontal contour along which correlation is approximately constant would be shifted upward or downward from the horizontal meridian. The amount of this shift would indicate the elevation, and the rest of the calculation would proceed in an analogous way.

Of course, this way of estimating vergence error depends on information about effective correlation being available across the visual field. It could not be implemented if the visual stimulus were simply one or two points of light. However, we are not aware of any evidence that single point targets outside the fovea elicit appropriate vertical vergence movements. Schor, Maxwell, and Stevenson (1994) showed that when the eyes saccade to peripheral targets, they make the appropriate changes in vertical vergence. However, the vertical vergence of a peripheral point target can be predicted from stereo geometry; thus, it is not clear that the vertical disparity of the peripheral point targets is explicitly measured. Indeed, Schor et al. showed that saccades during monocular viewing were associated with the same vertical vergence movements, suggesting that these movements are open loop, not a response to the vertical disparity in the stimulus. Thus, existing data on vertical vergence eye movements do not establish that the sign of vertical disparity is detected in local regions outside the fovea.

Subjects cannot learn to report the sign of vertical disparity

Although the oculomotor system clearly measures the sign of vertical disparities driving vergence correction, it does not share this information with the perceptual system. To verify this, we asked whether subjects could learn to discriminate the sign of vertical disparity. Whereas several studies have shown that human observers can detect the existence of vertical disparity, although with poorer acuity than horizontal disparity (Duwaer & van den Brink, 1982; Farell, 2003; McKee et al., 1990; Westheimer, 1978, 1984), no published study has examined whether they can discriminate its sign (although one unpublished study found that one of three observers could do so: Backus & Banks, 1998).

Previous psychophysical studies have shown that, to demonstrate an effect of vertical disparity on depth perception, stimuli must be large, subtending more than $\sim 10^\circ$ (Howard & Pierce, 1998; Kaneko & Howard, 1996, 1997a, 1997b; Pierce et al., 1998; Rogers & Bradshaw, 1993; Stenton et al., 1984). To give subjects the best chance of detecting signed vertical disparity, we therefore used random-dot stereograms that filled the screen, subtending $22^\circ \times 18^\circ$. The stimulus was presented for 140 ms, which is too short for voluntary vergence movements. A 2° square region around the central fixation cross was presented with zero disparity; the rest of

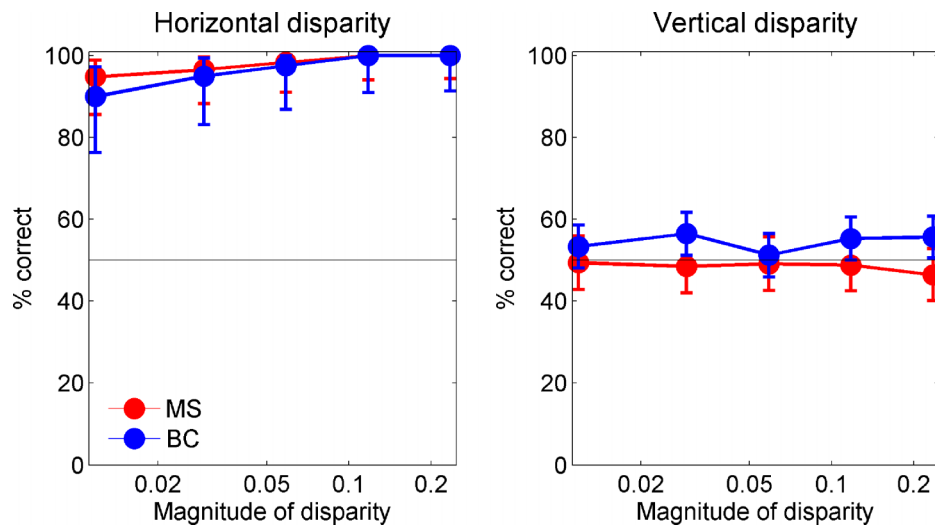


Figure 15. Results of a one-interval forced-choice task in which subjects were asked to discriminate the sign of disparity. Left: horizontal disparity; right: vertical disparity. A 2° square region around the central fixation cross was presented with zero disparity; the rest of the pattern had a uniform disparity, either horizontal or vertical. Subjects had to report the sign of this disparity. Different signs and magnitudes of disparity were interleaved randomly; vertical and horizontal disparities were applied in separate blocks. The stimulus was presented for 140 ms, which is too short to allow vergence movements. The data for vertical disparities represent a total of 3,020 trials for the two subjects. Error bars show 68% confidence intervals, assuming a simple binomial distribution.

the screen had a uniform disparity, either horizontal or vertical. Subjects had to report the sign of this disparity. Subjects pressed a mouse button after each trial; visual feedback indicated whether they had answered “correctly” or not. The feedback indicated the sign of the disparity, and subjects were directed to maximize the number of correct responses. When the disparity was applied horizontally, the naive subjects quickly realized that the correct strategy was to press the left mouse button if the central region appeared in front and the right mouse button if it appeared behind (Figure 15A). However, when the disparity was applied vertically, subjects were unable to find any strategy that enabled them to perform above chance (Figure 15B). Similar results were obtained with other stimulus configurations, for example, those in which there was a central disparate region on a large zero disparity background. We found no evidence that subjects could ever learn to report the sign of vertical disparity, even after training with feedback.

Discussion

In this article, we have examined the theoretical grounds for believing that the visual system detects and encodes vertical disparity. As an extreme example, we considered a model stereo system made up of binocular correlation sensors lying on the epipolar lines appropriate to primary position (zero vertical disparity in our coordinate system). We showed that this very simple model is, in theory, capable of supporting several phenomena that are usually taken as evidence that the visual system must contain a range of vertical-

disparity detectors, allowing for the rotation of epipolar lines as the eyes move. We have shown that all these perceptual phenomena could be experienced by a visual system that contains only sensors tuned to zero vertical disparity. In this view, the rotation of epipolar lines is taken into account by higher visual areas when decoding the activity of correlation detectors early in the visual system, as when our model deduces gaze direction and vergence. The model is consistent with the physiological evidence available to date. However, as we discuss below, the model makes psychophysical predictions that have not been borne out in our preliminary investigations.

The earliest physiological studies did not find any cells responding optimally to vertical disparities other than zero (Gonzalez et al., 1993; Maunsell & Van Essen, 1983; Poggio, 1995). Later studies, explicitly motivated by the psychophysical evidence of the perceptual effects of vertical disparity, have looked for evidence of cells tuned to a range of vertical disparities. Some of these, including the only study so far to have probed the full response matrix to combinations of vertical and horizontal disparity for each cell, did not find any convincing evidence for cells tuned to vertical disparities significantly different from zero (Cumming, 2002; Gonzalez et al., 2003), but both these used cells within 10° of the fovea. Two studies—one in the owl (Nieder & Wagner, 2001) and one in the monkey (Durand et al., 2002)—have reported cells tuned to nonzero vertical disparities, probably reflecting the fact that Durand et al. probed further out into the periphery, up to 22° . However, in these studies, vertical disparity was defined in terms of screen coordinates, not relative to epipolar lines. For near viewing distances, points placed in matching peripheral locations on

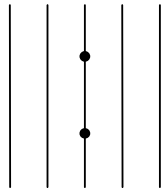


Figure 16. Ogle's minimal stimulus (Ogle, 1964, chap. 15). The stimulus, viewed with a vertically magnifying lens over one eye, consists of vertical rods. Two spheres are attached to the central fixated rod, providing the only vertical disparity cue in the stimulus. Although the induced effect was very weak in this stimulus, some subjects perceived the five rods as lying in a plane slanted away from frontoparallel.

two flat monitors do have a vertical disparity, which was not corrected for in these studies. Thus, even the model neurons used in our simulations, which are tuned to zero vertical disparity on the retina, would have been reported by Durand et al. as being tuned to vertical disparities ranging from 0° to 0.3° . It is unclear whether there are cells that are tuned to *retinal* vertical disparities significantly different from zero. Thus, for the moment at least, our model must stand or fall by psychophysical evidence.

We believe that our model is consistent with all published demonstrations of the perceptual effects of vertical disparity. Admittedly, it cannot explain slant illusions in stimuli in which the only stimulus vertical disparity is on the vertical meridian, as in Ogle's minimal stimulus (Ogle, 1964; Figure 16). At first sight, this is inconsistent with Ogle's reports that some subjects were able to obtain a weak induced effect even with this extremely impoverished vertical disparity cue. However, eye movements were not controlled in Ogle's experiments, and the viewing duration lasted many seconds. This raises the possibility that eye movements are responsible for the illusion of slant. If eye movements occur, our model too can explain the induced effect in this stimulus. By observing how binocular correlation increases and decreases as the gaze direction changes, reflecting the summation of the vertical disparity in the stimulus with the vertical disparity introduced by oblique gaze, the model stereo system would conclude that the central fixation rod was being viewed at an oblique gaze angle, and the induced effect follows.

The model provides a natural explanation of why vertical disparity is pooled across relatively large areas of the visual field (Adams et al., 1996; Kaneko & Howard, 1996; Rogers & Bradshaw, 1995; Stenton et al., 1984). Because it depends on the large-scale pattern of binocular correlation, it is insensitive to local fluctuations in vertical disparity. It naturally reproduces the results of Stenton et al. (1984), in which applying vertical magnification to restricted regions of the visual field produces a weaker version of the induced effect, with the degree of slant increasing as the percentage of mag-

nified points increases. A scrambled version of the induced effect, in which the magnitude of applied vertical disparity is the same as the induced effect (Figure 6C) but its sign is picked at random for each point, produces no slant in our model because the response of the population of correlation sensors is noisy but, on average, symmetrical about the vertical meridian.

In our discussion of the induced effect, we have assumed that its perceptual effects are mediated via an estimate of eye position. But what if the effect of vertical disparity is not mediated through gaze angle? There is a large strand of evidence suggesting that vertical disparity is used in a more ad hoc way to estimate visual scene quantities like slant or curvature, without being used to construct an explicit estimate of eye position (Banks et al., 2002, 2001; Duke & Howard, 2005; Garding et al., 1995; Kaneko & Howard, 1996; Koenderink & van Doorn, 1976; Rogers & Bradshaw, 1993). We do not regard this as critical to our argument. We framed the discussion in terms of eye position because it seemed simple and intuitive, but we believe that the essential point still holds. Our population of horizontal-disparity detectors provides a retinal map of the magnitude of vertical disparity, as specified in Equation 21. Thus, any scheme that uses vertical disparity gradients across a large area of the visual field could still be supported by this population. Furthermore, because the global pattern of vertical disparities is so regular, it is simple to infer the signs of each local vertical disparity, if the global pattern has been identified. As we have seen, the population response of horizontal-disparity detectors (Figure 9) is characterized by a cross-shaped contour along which the response is constant and equal to 1 (in the absence of vergence error). The sign of vertical disparity is always positive in the top-right and bottom-left quarters defined by this cross and negative in the top-left and bottom-right quarters. In other words, the magnitude of vertical disparity could be deduced from the population of horizontal-disparity detectors, and its sign is then constrained by stereo geometry. Thus, irrespective of exactly how vertical disparity supports the induced effect, it could be implemented by this population.

The model faces more serious challenges in accounting for the effects of vertical disparity on vergence. We have shown that the sign of vertical vergence error can be deduced from the population of horizontal-disparity detectors. If there is a vergence error, then the response along the cross-shaped contour will be less than 1. The magnitude of the vertical vergence error can be deduced from the lowered response on the contour, whereas its sign can be deduced from the quadrants in which the response is maximal (Figure 14). However, this method fails when the eyes are in primary position. We have been unable to find any published studies of vertical vergence movements in response to vertical disparity in stimuli viewed at infinity, but it certainly seems very unlikely that vertical disparity in such stimuli would fail to evoke vertical vergence movements. One possible way around this would be to consider a slightly different model

in which all disparity detectors are again restricted to a single set of epipolar lines, but these are no longer the epipolar lines associated with primary position. If the epipolar lines were those associated with a slight divergence (negative vergence angle), then this method would work for all gaze positions.

In practice, our guess is that the visual system does contain a small number of specific vertical-disparity detectors (i.e., a population tuned to a range of vertical disparities) to drive corrective vergence movements. We suggest that these could be kept entirely separate from the disparity detectors used to support perception. Vertical disparity caused by gaze direction/elevation when the eyes are fixating eccentrically has very different properties to vertical disparity caused by vergence error. In the former case, vertical disparity is always zero at the fovea and gets larger as one moves toward the periphery. To detect and use this vertical disparity, it would make sense to concentrate detectors at the periphery (Durand et al., 2002; Rogers & Bradshaw, 1993; Trotter et al., 2004). In contrast, vertical vergence error causes a uniform vertical disparity across the retina, including at the fovea, where changes in conjugate gaze angle do not produce vertical disparities. It would therefore make sense to concentrate vergence-error detectors around the fovea. Thus, we arrive at a slightly modified version of the model considered so far in this article. Perception is supported by a large population of pure horizontal-disparity detectors all across the visual field, tuned to a range of horizontal disparity but all to zero vertical disparity. As we have shown, the *perceptual* consequences of vertical disparity could all be due to its effect on these detectors, produced via an effective reduction in binocular correlation. Vertical vergence eye movements are supported by a very small population of vertical-disparity detectors at the fovea, which are of little use for perception because vertical disparity is always zero at the fovea once correct alignment has been achieved. This accords with evidence that vertical disparity is more potent at eliciting vergence movements if it is closer to the fovea (Howard et al., 2000). It also explains the different pattern of saccades to peripheral targets with horizontal versus vertical disparity. Under normal viewing conditions, the vertical disparity at each location in the visual field can be predicted from a knowledge of the eyes' position and stereo geometry. The brain takes advantage of this and programs saccades to peripheral targets with the appropriate vertical vergence, based on the vertical disparity that is expected at that location in the visual field. If this vertical vergence turns out to be incorrect, a new "expected vertical disparity map" can be learnt quite rapidly (McCandless, Schor, & Maxwell, 1996). In contrast, no such open-loop programming exists for saccades to horizontally disparate peripheral targets. Here, the horizontal vergence (prior to a saccade) is appropriate to the individual target and does not have to be learnt (Collewijn, Erkelens, & Steinman, 1997; Rashbass & Westheimer, 1961). This strongly suggests that the oculomotor system has access to a detailed local map of horizontal disparity, measured instantaneously across the whole visual field. In contrast, for vertical disparity, the oculomotor sys-

tem has access only to a remembered map, built up gradually from measurements made at the fovea. While doubtless an oversimplification, this version of our model explains all existing psychophysical and physiological data in a very economical way.

We noted above that our model stands or falls by psychophysics. Here, it makes a number of clear predictions. If the model is correct in postulating that the perceptual effects of vertical disparity are mediated by a reduction in binocular correlation, it should be possible to mimic these perceptual effects by manipulating binocular correlation. For instance, one should be able to produce a percept of a slanted plane in the absence of any disparity—either vertical or horizontal—simply by altering binocular correlation to mimic the induced effect. We have tried to do so, without success. However, the comparison is complicated by the fact that the mapping between vertical disparity and binocular correlation depends on the spatial scale of the disparity sensors. Equation 21 shows that, for a sensor whose RFs have standard deviation σ , a vertical disparity of Δy is roughly equivalent to reducing the binocular correlation by a factor of $\exp(0.25\Delta y^2/\sigma^2)$. Thus, it is not possible to reproduce the effects of vertical disparity with binocular correlation in a broadband image because the reproduction will not agree across scales. Even if the image is filtered, it is impossible to stimulate just one spatial frequency/orientation channel; hence, one would expect the illusion to be less compelling than in the real induced effect. Therefore, failing to mimic the induced effect in this way still leaves open the possibility that vertical disparity is equivalent to decorrelation within a single channel.

A more compelling approach is to produce an induced effect with a uniform 80% correlated stimulus and then try to null the illusion by bringing the binocular correlation back up to 100% in a ramp across the visual field, producing a correlation gradient that mimics vertical disparities in the opposite direction to those produced by the vertical magnification. Although this nulling might not be perfect for any channel, one would expect it to disrupt the induced effect. We have attempted this, but so far, we have been unable to demonstrate any nulling effect of a binocular correlation gradient on the induced effect. This suggests that the effects of vertical disparity are *not* simply mediated by binocular correlation. Because we have shown here that the other perceptual effects of vertical disparity can be explained purely by pure horizontal-disparity detectors, this null result, if confirmed, would be the first conclusive perceptual evidence that the stereo system does contain vertical-disparity detectors. It therefore warrants further investigation.

It is also possible that, even if the visual system contains some dedicated vertical-disparity detectors (and reads them out as such), the mechanism proposed here may also contribute to perception. It seems clear from the physiology that sensors tuned to nonzero vertical disparities, if they exist at all, are a small minority of disparity-tuned neurons, while we have shown that most pure horizontal-disparity detectors also contain valuable information about vertical

disparity. Thus, it would seem sensible for the visual system to take this information into account when making judgements about vertical disparity. It may be possible to design psychophysical stimuli to test this.

Even if the model system considered here, containing purely horizontal-disparity detectors, proves not to be an accurate model of the visual system, the exercise has nevertheless been instructive. Understanding all that can be achieved with purely horizontal-disparity detectors is essential for understanding what the brain achieves by having vertical-disparity detectors (and keeping track of their vertical disparity when decoding). It also raises some stimulating questions about the brain's encoding strategy. A common assumption in neuroscience is that the brain's representation of the world is efficient, matched to the statistical properties of the world it encounters. In the case of disparity, this should mean that the brain devotes vastly more resources to encoding horizontal, rather than vertical, disparity. Several physiological studies (Durand et al., 2002; Gonzalez et al., 2003, 1993; Maunsell & Van Essen, 1983; Nieder & Wagner, 2001; Trotter et al., 2004; but see Cumming, 2002) and even some psychophysical studies (Farell, 1998) have suggested that this is not the

case. Previous workers have argued that the brain needs to devote neuronal resources to encoding vertical disparity to achieve a local map of vertical disparity across the retina, which can then be used to extract quantities such as eye position, slant, and so forth. In other words, this is a case where disproportionate neuronal resources are devoted to statistically rare events, such as large vertical disparity, because they are particularly informative when they do occur. However, this article shows that this is not a valid explanation. Resources could be devoted exclusively to horizontal disparity, and a map of vertical disparity would “come free.” A full understanding of the role of vertical disparity in perception will have to explain why the brain does not adopt this seemingly highly efficient strategy. A possible reason is that this strategy depends on binocular correlation being close to 100% in natural stimuli (Appendix, Equation 21). If this assumption is too often violated—due to effects such as occlusion at scene boundaries, significant changes in depth over a receptive field, or luminance differences between the eyes—it may be necessary to include a population of explicit vertical-disparity detectors, despite the computational cost.

Appendix A

Coordinate systems

Head-centered space coordinates

In discussing stereo geometry, it is important to have suitable coordinate systems. We use the same coordinate system developed in Read and Cumming (2004). To describe an object's position relative to the head, we use a head-centered Cartesian coordinate system (X, Y, Z), whose origin is at the midpoint between the two eyes' nodal points (Figure A1). Z is the depth axis (Z increases with distance from the observer), Y is the vertical axis (Y increases as the object moves upward), and X is the horizontal axis (X increases as the object moves leftward).

Eye position coordinates

For describing eye position, we use the Helmholtz coordinate system, again as in Read and Cumming (2004), except that we do not consider torsional eye movements in this article. The eyes are assumed to rotate about their nodal points (Figure A1(B)); thus, the nodal points remain at the same place in the head as the eyes move. Azimuthal eye position H is the angle by which the eye's optic axis is rotated about an axis passing through the nodal point and parallel to the Y -axis (Figure A1(B)). Positive values of H indicate that the eye is turned to the left. When the eyes are converged, H will be different for the two eyes. We use subscripts to denote the value for individual eyes: H_L, H_R . In expressions that could apply to either eye, we shall write H without any subscript; it should then be understood that H should be replaced with H_L to obtain an expression valid for the left eye and with H_R for the right eye.

The vergence angle is the difference in the two eye's azimuthal gaze directions:

$$D = H_R - H_L. \quad (1)$$

Many subsequent expressions will involve half the vergence angle, which we write $D_{1/2}$. We shall also define the azimuthal gaze angle H_c to be the average of the azimuthal position of each eye:

$$H_c = (H_R + H_L)/2. \quad (2)$$

In most parts of this article, we assume that there is no elevation; hence, the fixation point lies in the XZ plane. However, when considering vertical vergence errors, we shall need the Helmholtz elevation angles, V_L, V_R , describing the angle by which

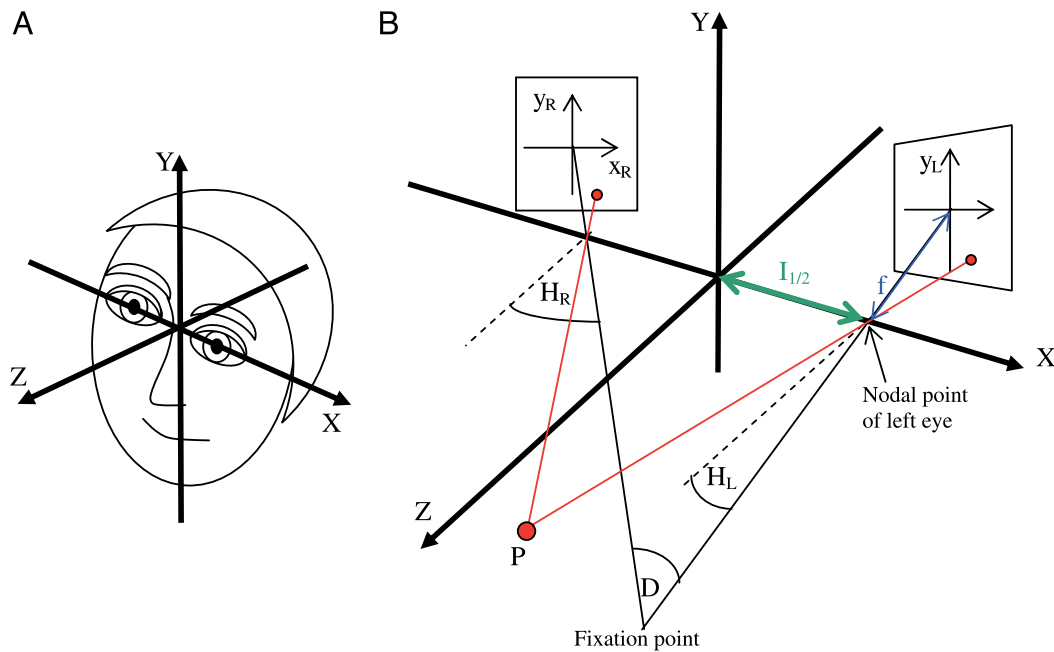


Figure A1. (A) Head-centered coordinate system. (B) Describing eye position and position on the retina. $I_{1/2}$ is half the interocular distance; f is the focal length of the eye. The points $\pm I_{1/2}$ on the X -axis are the nodal points of the two eyes.

each eye’s axis is rotated about the X -axis. Positive values of V indicate that the eye is looking down. In the Helmholtz coordinates we use, this elevation is applied *after* the azimuthal rotation. For the eyes to be correctly fixated, their Helmholtz elevations must be the same: $V_L = V_R$. If the Helmholtz elevations are different, then the gaze rays of the eyes do not intersect (even at infinity), and there is a vertical vergence error. In our previous article (Read & Cumming, 2004), we did not allow for this possibility and only considered the case $V = V_L = V_R$.

Projection onto the retinae

For calculating the position of images on the retina, it is convenient to represent the position of each eye by rotation matrices:

$$R_H = \begin{bmatrix} \cos H & 0 & \sin H \\ 0 & 1 & 0 \\ -\sin H & 0 & \cos H \end{bmatrix}; \quad R_V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos V & -\sin V \\ 0 & \sin V & \cos V \end{bmatrix}; \quad R = R_V R_H. \tag{3}$$

R_H represents the eye’s azimuthal rotation about the Y -axis, and R_V represents its elevation about the X -axis. Their product R represents the final position of the eye (the order is important; as mentioned above, the elevation in our coordinate system is applied after azimuthal rotation). Obviously, to obtain matrices for each eye, H, V in these expressions must be replaced with H_L, V_L or H_R, V_R as appropriate. As an example of how these matrices are used, consider finding the direction of the optic axis. In primary position, the eye’s optic axis is parallel to the Z -axis and may be represented by the vector $\mathbf{Z} = (0,0,1)$. With azimuth H and elevation V , the optic axis is parallel to the vector $R\mathbf{Z}$.

As described in Figure 3, the retinae are represented by planes. Position on the retina is represented by a Cartesian coordinate system (x, y) . When the eye is in primary position ($H = 0$), the x - and y -axes are parallel to the X - and Y -axes, respectively. An object at $\mathbf{P} = (X, Y, Z)$, such as the red point in Figure A1(B), projects to the point (x_L, y_L) on the left retina and to the point (x_R, y_R) on the right. The image coordinates (x, y) may be expressed very simply in terms of the eye’s rotation matrix.

$$\begin{aligned} x/f &= -(\mathbf{P} - \mathbf{N}) \cdot \mathbf{RX} / (\mathbf{P} - \mathbf{N}) \cdot \mathbf{RZ} \\ y/f &= -(\mathbf{P} - \mathbf{N}) \cdot \mathbf{RY} / (\mathbf{P} - \mathbf{N}) \cdot \mathbf{RZ}. \end{aligned} \tag{4}$$

\mathbf{X} , \mathbf{Y} , and \mathbf{Z} are unit vectors along each of the axes. \mathbf{N} is a vector representing the nodal point of the eye. The rotation matrix \mathbf{R} was given in Equation 3. When evaluating this expression for a particular eye, the appropriate values of \mathbf{N} and \mathbf{R} must be used. For the left eye, $\mathbf{N}_L = (I_{1/2}, 0, 0)$, and for the right, $\mathbf{N}_R = (-I_{1/2}, 0, 0)$, compare Figure A1(B). f is the focal length of the eye, and $I_{1/2}$ is half the interocular distance. To obtain \mathbf{R} for the left eye, replace H, V with H_L, V_L in Equation 3.

Image position for zero elevation

Although Equation 4 is the most compact way of writing the retinal image coordinates for a general eye position, most parts of this article considers the case of zero elevation: $V_L = V_R = 0$. In this case, $R = R_H$, and the retinal coordinates of the images of an object at (X, Y, Z) are:

$$x_L = f \frac{Z \sin H_L - (X - I_{1/2}) \cos H_L}{[(X - I_{1/2}) \sin H_L + Z \cos H_L]}; \quad x_R = f \frac{Z \sin H_R - (X + I_{1/2}) \cos H_R}{[(X + I_{1/2}) \sin H_R + Z \cos H_R]}; \tag{5}$$

$$y_L = -\frac{fY}{[(X - I_{1/2}) \sin H_L + Z \cos H_L]}; \quad y_R = -\frac{fY}{[(X + I_{1/2}) \sin H_R + Z \cos H_R]}. \tag{6}$$

In the induced-effect stimulus, we artificially adjust the vertical position of the images in the two eyes. We apply the distortion symmetrically, expanding the right eye’s image by a magnification factor \sqrt{M} , and shrinking the left eye’s image by $1/\sqrt{M}$. For induced-effect stimuli, therefore, Y in Equation 6 should be replaced with Y/\sqrt{M} for the left eye and $Y\sqrt{M}$ for the right eye.

Angular coordinates

The head-centered coordinates (X, Y, Z) and retinal coordinates (x, y) are in units of distance. As we shall see below, these are convenient mathematically. However, it is more usual in visual science to present results in degrees. Figure 3 showed how retinal coordinates could be expressed as angles:

$$\hat{x} = \arctan (x/f), \quad \hat{y} = \arctan (y/f), \tag{7}$$

where f is the focal length of the eye. Similarly, the direction to an object in space can be expressed as

$$\hat{X} = \arctan (X/Z) \text{ and } \hat{Y} = \arctan (Y/Z). \tag{8}$$

These specify the object’s direction in degrees from the vertical and horizontal axes, respectively.

We use these definitions to convert the image coordinates given in Equation 6 into angles. If we also allow for an induced effect with magnification factor M , we obtain

$$\hat{y}_L = -\arctan \frac{\tan \hat{Y}}{\sqrt{M} \left(\left(\tan \hat{X} - \frac{I}{2Z} \right) \sin H_L + \cos H_L \right)}; \quad \hat{y}_R = -\arctan \frac{\sqrt{M} \tan \hat{Y}}{\left(\left(\tan \hat{X} + \frac{I}{2Z} \right) \sin H_R + \cos H_R \right)}, \tag{9}$$

where the angle \hat{y} says how many degrees the image falls above the retina’s horizontal meridian. We define the angular vertical disparity as the difference between these two angles:

$$\Delta\hat{y} = \hat{y}_R - \hat{y}_L. \tag{10}$$

The angular vertical cyclopean location is their mean:

$$\hat{y}_c = \frac{1}{2}(\hat{y}_R + \hat{y}_L). \quad (11)$$

The angular horizontal disparity and cyclopean position are defined similarly. In the resulting figures, quantities like stimulus disparity and so forth are plotted as a function of cyclopean angular position on the retina, (\hat{x}_c, \hat{y}_c) .

Reconstructing the visual scene

In [Figure 12](#), we show the visual scene reconstructed from the retinal stimulus, using the estimates of eye position available from retinal information, and the cyclopean position x_c and horizontal disparity Δx of each point. We do this by inverting [Equation 5](#), expressing X and Z in terms of x_L and x_R . We obtain

$$X = I_{1/2} \frac{f^2 \sin 2H_c - 2fx_c \cos 2H_c - x_L x_R \sin 2H_c}{f^2 \sin D - f\Delta x \cos D + x_R x_L \sin D}; \quad Z = I \frac{[x_R \sin H_R + f \cos H_R][x_L \sin H_L + f \cos H_L]}{f^2 \sin D - f\Delta x \cos D + x_R x_L \sin D}$$

or equivalently (because $x_c = (x_L + x_R)/2$, $\Delta x = x_R - x_L$, $H_c = (H_L + H_R)/2$, $D = H_R - H_L$):

$$X = I_{1/2} \frac{(f^2 - x_c^2 + \Delta x^2/4) \sin 2H_c - 2fx_c \cos 2H_c}{(f^2 + x_c^2 - \Delta x^2/4) \sin D - f\Delta x \cos D} \quad (12)$$

$$Z = I_{1/2} \frac{(f^2 + x_c^2 - \Delta x^2/4) \cos D + (f^2 - x_c^2 + \Delta x^2/4) \cos 2H_c + 2fx_c \sin 2H_c + f\Delta x \sin D}{(f^2 + x_c^2 - \Delta x^2/4) \sin D - f\Delta x \cos D}.$$

If we use the correct values for gaze angle H_c and vergence D , we of course reconstruct the actual position in space of the object whose images fell at x_L , x_R in the two retinae ([Figure 12A](#)). If we use the estimates of H_c , D derived from fitting the neuronal responses (cf. [Equation 23](#)), we can reconstruct the position as it would be perceived by the visual system ([Figure 12D](#)).

Predicting the vertical disparity, given cyclopean location, horizontal disparity, and eye position

Given the horizontal-disparity field of the stimulus and the position of the eyes, it is possible to derive the vertical-disparity field that must necessarily exist under natural viewing conditions (i.e., assuming no manipulations such as vertical magnification, as in the induced effect). To keep this derivation simple, we assume that the eyes are fixating in the XZ plane; that is, Helmholtz elevation is zero for both eyes, and we work in positional, rather than angular, coordinates. We define positional vertical disparity to be

$$\Delta y = y_R - y_L \quad (13)$$

and positional vertical cyclopean location to be

$$y_c = (y_R + y_L)/2. \quad (14)$$

These are of course entirely analogous to the corresponding definitions for angular disparity and cyclopean location ([Equations 10](#) and [11](#)), but note that there is in general no straightforward relationship between positional and angular disparity: In the periphery, a given positional disparity corresponds to a smaller angular disparity than it would do near the fovea. However, when positional disparity is zero, then angular disparity is also zero, a fact we shall exploit below.

Substituting for y_L and y_R from Equation 6 and then eliminating the object's vertical position Y , Equations 13 and 14 yield the following relationship between vertical cyclopean position and disparity in positional coordinates:

$$\Delta y = -2y_c \frac{(X \sin D_{1/2} \cos H_c + I_{1/2} \cos D_{1/2} \sin H_c - Z \sin D_{1/2} \sin H_c)}{(X \cos D_{1/2} \sin H_c + I_{1/2} \sin D_{1/2} \cos H_c + Z \cos D_{1/2} \cos H_c)}. \quad (15)$$

Note that no such simple relationship exists between the equivalent quantities in angular coordinates, $\Delta \hat{y}$ (Equation 10) and \hat{y}_c (Equation 11), because of the tangents/arctangents in Equation 9. This is why we used positional coordinates for this simulation.

Rearranging Equation 5 to express X and Z as a function of x_L and x_R on the planar retina, we obtain

$$X = I_{1/2} \frac{\left[\left(1 - \frac{x_L x_R}{f^2} \right) \sin (H_R + H_L) - \frac{(x_L + x_R)}{f} \cos (H_R + H_L) \right]}{\left[\left(1 + \frac{x_L x_R}{f^2} \right) \sin (H_R - H_L) + \frac{(x_L - x_R)}{f} \cos (H_R - H_L) \right]}$$

$$Z = I_{1/2} \frac{\left[\cos H_L \cos H_R + \frac{x_R}{f} \sin H_R \cos H_L + \frac{x_L}{f} \sin H_L \cos H_R + \frac{x_L x_R}{f^2} \sin H_R \sin H_L \right]}{\left[\left(1 + \frac{x_L x_R}{f^2} \right) \sin (H_R - H_L) + \frac{(x_L - x_R)}{f} \cos (H_R - H_L) \right]}.$$

We can replace x_L , x_R with the cyclopean location and disparity: $x_L = x_c - \Delta x/2$, $x_R = x_c + \Delta x/2$. Then, substituting these expressions into Equation 15 and simplifying, we obtain

$$\Delta y(x_c, y_c) = -2y_c \frac{\left[\cos H_L - \cos H_R + \left(\frac{2x_c - \Delta x(x_c, y_c)}{2f} \right) \sin H_L - \left(\frac{2x_c + \Delta x(x_c, y_c)}{2f} \right) \sin H_R \right]}{\left[\cos H_L + \cos H_R + \left(\frac{2x_c - \Delta x(x_c, y_c)}{2f} \right) \sin H_L + \left(\frac{2x_c + \Delta x(x_c, y_c)}{2f} \right) \sin H_R \right]}. \quad (16)$$

This is the vertical disparity that must be experienced at the cyclopean position (x_c, y_c) , given that the horizontal disparity at that position is $\Delta x(x_c, y_c)$, the Helmholtz elevation is zero, and the Helmholtz azimuths are H_L, H_R .

Retinal locus of zero vertical disparity (for zero elevation)

When the Helmholtz elevation is zero for both eyes, objects in the XZ plane project to the horizontal meridian on the retina, irrespective of the eyes' azimuthal gaze directions or the position of the object within the XZ plane. Thus, vertical disparity is zero along the horizontal retinal meridian, $\hat{y} = 0$ (Equation 9). However, there is also a vertical line on the retina along which vertical disparity is also zero, producing a cross-shaped contour of constant vertical disparity. It is the location of this cross that enables us to derive eye position. In the Results section, we state that the location of the vertical arm of the cross tells us the azimuthal gaze direction, H_c . This is a slight approximation, and we here do a more rigorous analysis.

Equation 16 gives the positional vertical disparity, which, as noted, has no simple relationship to the angular disparity. However, the locus of zero disparity will be the same for both types of disparity; thus, we can exploit the relatively simple expression we were able to derive in positional coordinates to deduce the conditions under which angular vertical disparity is zero. From Equation 16, we find that vertical disparity is zero when either $\hat{y}_c = 0$ (the horizontal arm of the cross), or

$$x_c = \left(1 - \frac{\Delta x(x_c, y_c)}{2f \tan D_{1/2}} \right) f \tan H_c.$$

If the horizontal disparity is zero, this becomes simply $\hat{x}_c = H_c$. In other words, the gaze angle can be simply read off from the horizontal position of the zero vertical-disparity cross. However, this is only true when the horizontal disparity is small compared with the vergence angle. Stereopsis only operates up to horizontal disparities of $<1^\circ$ or so; hence, under many

relevant situations, the approximation is valid, and it gives an intuitive idea of how eye position may be recovered. However, our fitting routines did *not* use this approximation. The predicted vertical-disparity field was calculated exactly, using Equation 16, and optimization was performed on the whole field, not just the locus of zero disparity.

Deriving eye position

In Figure 9, we show how eye position may be estimated from the response of a population of binocular correlation detectors all tuned to zero vertical disparity. We restrict ourselves to the case of zero elevation. Conceptually, this is very simple. We assume that the brain has been able to solve the stereo correspondence problem to arrive at an accurate map of horizontal disparity at each point in the image. Thus, the brain knows the horizontal disparity of every object in the visual scene. If it also makes a guess about the current gaze angle and vergence, it can deduce the position of each object in space and, hence, its predicted vertical disparity, according to Equation 16. But once the horizontal and vertical position and disparity of each object are known, then the response of the population of correlation detectors can be predicted and compared with the actual response. Our fitting routine adjusts the values of gaze angle and vergence until the predicted response best matches the actual response. The sections that follow lay out the math involved. As we have seen, this is greatly simplified if we use position on the planar retina (Figure 3) rather than angle. For this reason, the following sections will use position coordinates (x, y) rather than the more intuitive angular coordinates (\hat{x}, \hat{y}) used in the figures.

Measuring binocular correlation

In Figure 2, we postulated a “correlation sensor” that measured the effective binocular correlation between particular regions of the retina. What does this mean in practice? Let us take a concrete example. Suppose that the visual stimulus is binary noise, made up of infinitesimal pixels colored either black or white, and that it has both binocular disparity and imperfect binocular correlation. Suppose that at a particular cyclopean position (x_c, y_c) —say $(1, 2)$ —the correlation is $C_{\text{stim}} = 0.8$ and the 2D disparity is $\Delta x_{\text{stim}} = 0.4$, $\Delta y_{\text{stim}} = 0.02$. The disparity means that the pixel at $(x_c - \Delta x_{\text{stim}}/2, y_c - \Delta y_{\text{stim}}/2) = (0.8, 1.99)$ in the left eye corresponds to the pixel at $(x_c + \Delta x_{\text{stim}}/2, y_c + \Delta y_{\text{stim}}/2) = (1.2, 2.01)$ in the right eye. If the stereogram were perfectly correlated, then these pixels would therefore be the same, either both white or both black; thus, their product would always be 1 (taking white to be +1 and black to be -1). In fact, the correlation is only 80% at that point in the image; hence, the expected value of their product is only 80% (i.e., there is a 90% chance that both pixels are black or both are white, but a 10% chance that they have opposite polarities).

We shall model the correlation sensors very simply as binocular neurons whose receptive fields are isotropic Gaussians on the planar retina. The RFs in the two eyes are identical apart from their position. The mean of the RF positions in the two planar retinæ defines their preferred cyclopean stimulus location, $(x_{\text{pref}}, y_{\text{pref}})$, and their horizontal position disparity defines their preferred stimulus disparity, Δx_{pref} . The RFs always have the same vertical location y ; thus, their preferred vertical disparity is zero.

To obtain units whose output reflects the binocular correlation of the stimulus, we begin with energy-model subunits (Ohzawa et al., 1990), whose response is the square of the summed output from left and right receptive fields, $(L + R)^2$. This full-squared output can be thought of as the combined outputs of a push-pull pair of simple cells, each of which computes a half-squared output. We used tuned-excitatory units, for which the receptive-field profiles are identical in the two eyes, differing only in their horizontal position. Thus, the inputs from the two eyes are

$$L = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy I_L(x, y) \rho \left(x - x_{\text{pref}} + \frac{\Delta x_{\text{pref}}}{2}, y - y_{\text{pref}} \right)$$

$$R = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dx dy I_R(x, y) \rho \left(x - x_{\text{pref}} - \frac{\Delta x_{\text{pref}}}{2}, y - y_{\text{pref}} \right).$$

$I_L(x, y)$ and $I_R(x, y)$ are the images on the two retinæ. These are expressed relative to the mean luminance, so that $I(x, y)$ is positive for bright features and negative for dark. $\rho(x, y)$ is a receptive-field profile centered on zero. For an individual unit, this profile is displaced on the retina depending on the unit’s preferred horizontal disparity and cyclopean position. $(x_{\text{pref}}, y_{\text{pref}})$ is the unit’s preferred cyclopean location on the retina. Δx_{pref} is its preferred horizontal disparity; the centers of the left and right receptive fields feeding into the unit are offset horizontally from one another by Δx_{pref} , giving the unit its disparity tuning. In our simulations, we consider only units tuned to the stimulus horizontal disparity, so that $\Delta x_{\text{pref}} = \Delta x_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}})$.

The unit’s response can be divided into a linear sum of monocular terms, $M = L^2 + R^2$, and a binocular term $B = 2LR$. When the stimulus is 100% correlated and the unit is viewing corresponding regions of the image in its two receptive fields, then $L = R$, and thus, these two terms become equal: $M = B$. In general, for images with arbitrary disparity and correlation, we

can calculate the expected values, $\langle B \rangle$ and $\langle M \rangle$, where the average is taken over many different random-dot patterns with the same disparity and correlation fields:

$$\begin{aligned} \langle M \rangle &= 2 \int dx_c dy_c \rho^2(x_c - x_{\text{pref}}, y_c - y_{\text{pref}}) \\ \langle B \rangle &= 2 \int dx_c dy_c C_{\text{stim}}(x_c, y_c) \\ &\quad \rho\left(x_c - x_{\text{pref}} - \frac{(\Delta x_{\text{stim}}(x_c, y_c) - \Delta x_{\text{pref}})}{2}, y_c - y_{\text{pref}} - \frac{\Delta y_{\text{stim}}(x_c, y_c)}{2}\right) \\ &\quad \rho\left(x_c - x_{\text{pref}} + \frac{(\Delta x_{\text{stim}}(x_c, y_c) - \Delta x_{\text{pref}})}{2}, y_c - y_{\text{pref}} + \frac{\Delta y_{\text{stim}}(x_c, y_c)}{2}\right). \end{aligned} \tag{17}$$

The integration variables x_c, y_c represent position on a cyclopean retina. $\Delta x_{\text{stim}}(x_c, y_c)$ and $\Delta y_{\text{stim}}(x_c, y_c)$ are the horizontal- and vertical-disparity fields of the stimulus, and $C_{\text{stim}}(x_c, y_c)$ its binocular correlation. Note that all three are allowed to vary as a function of position on the cyclopean retina; that is, these expressions are not restricted to frontoparallel stimuli. Similar expressions were derived in Prince, Pointon, Cumming, and Parker (2002, p. 206) and Read and Cumming (2003, p. 2814). Although we have generalized to allow for varying vertical- and horizontal-disparity fields and for varying binocular correlation, the details of the derivation are sufficiently similar that it does not seem worth reproducing them.

Figure 9D shows the ratio of these quantities, $C = \langle B \rangle / \langle M \rangle$, for Gaussian receptive fields:

$$\begin{aligned} C = \frac{\langle B \rangle}{\langle M \rangle} &= \frac{1}{\pi \sigma^2} \int_{-\infty}^{+\infty} dx_c \int_{-\infty}^{+\infty} dy_c C_{\text{stim}}(x_c, y_c) \exp\left(-\frac{1}{2\sigma^2} \left[\left(y_c - y_{\text{pref}} + \frac{\Delta y_{\text{stim}}(x_c, y_c)}{2} \right)^2 + \left(y_c - y_{\text{pref}} - \frac{\Delta y_{\text{stim}}(x_c, y_c)}{2} \right)^2 \right. \right. \\ &\quad \left. \left. + \left(x_c - x_{\text{pref}} + \frac{\Delta x_{\text{stim}}(x_c, y_c)}{2} - \frac{\Delta x_{\text{pref}}}{2} \right)^2 + \left(x_c - x_{\text{pref}} - \frac{\Delta x_{\text{stim}}(x_c, y_c)}{2} + \frac{\Delta x_{\text{pref}}}{2} \right)^2 \right] \right). \end{aligned} \tag{18}$$

If, in addition to the receptive fields being Gaussian, the stimulus disparity and correlation remain approximately constant over the unit’s receptive field, then the quantity C has a particularly simple form:

$$C = C_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}}) \exp\left\{-\frac{1}{4\sigma^2} \left[(\Delta x_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}}) - \Delta x_{\text{pref}})^2 + \Delta y_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}})^2 \right]\right\}. \tag{19}$$

Thus, for sensors that are perfectly tuned to the disparity of the stimulus, this is simply the local binocular correlation of the stimulus at the receptive field, $C = C_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}})$. However, notice that any mismatch between the sensor’s preferred disparity and that of the stimulus causes a reduction in response. The response falls off as a Gaussian function of the disparity mismatch, with SD equal to $\sqrt{2}$ that of the Gaussian RF. A population of these correlation detectors, which included all possible horizontal and vertical disparities, would encode both the local 2D disparity and the local correlation of the stimulus. Roughly speaking—ignoring the complexities of the correspondence problem—at each position on the cyclopean retina (x_c, y_c), the local correlation $C_{\text{stim}}(x_c, y_c)$ would be given by the response of the maximally responding sensor tuned to that cyclopean position (i.e., with $x_{\text{pref}} = x_c, y_{\text{pref}} = y_c$), and the local disparity would be given by the disparity tuning of that maximally responding sensor (i.e., $\Delta x(x_c, y_c) = \Delta x_{\text{pref}}, \Delta y(x_c, y_c) = \Delta y_{\text{pref}}$). The model stereo system considered here falls short of this in that the population contains only horizontal disparity sensors. Thus, the horizontal disparity of the stimulus can still be deduced from the response of the maximally responding sensor, but the vertical disparity and binocular correlation are confounded. A maximal response of

$$C_{\text{max}} = C_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}}) \exp\left\{-\frac{1}{4\sigma^2} \left[\Delta y_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}}) \right]^2\right\} \tag{20}$$

(obtained when the horizontal-disparity tuning matches the stimulus) could mean that the stimulus has zero vertical disparity and binocular correlation of $C_{\text{stim}} = C_{\text{max}}$ or that it has 100% binocular correlation and a vertical disparity of magnitude

$$|\Delta y_{\text{stim}}| = 2\sigma\sqrt{-\ln C_{\text{max}}}. \quad (21)$$

Fitting for eye position, given the response of a neuronal population to a single noisy image

Equations 17, 18, 19, 20, and 21 are for the *average* response, averaged over all binary noise stimuli. For any individual noise stimulus, the value of the energy-model components B and M may be quite different. This leads to considerable noise in the field B/M for any individual stimulus. This noise only affects regions of the image where there is significant vertical disparity. Along the locus of zero vertical disparity, and because we are considering only sensors tuned to the horizontal disparity of the stimulus, the receptive fields in each eye are viewing corresponding regions of the visual scene. This means that although the output from each eye, L and R , fluctuates depending on the particular pattern of black and white dots, L is always equal to R , because each eye always sees the same dot pattern as the other eye. Thus, $B/M = 2LR/(L^2 + R^2)$ is always equal to 1. Thus, the locus of zero vertical disparity and, hence, the gaze angle can be reliably deduced even from the response of a single sensor to a single image. However, estimates of vergence are much more seriously affected. The estimate of vergence depends on measuring how rapidly the effective binocular correlation falls off from its peak value of 1 along the locus of zero vertical disparity. Away from this locus, vertical disparity in the stimulus means that the receptive fields are not, in general, seeing exactly corresponding regions of the image. This means that L and R are not quite equal, even if the sensor's horizontal disparity is matched to that of the stimulus. Not only is the mean value $\langle B \rangle / \langle M \rangle$ less than 1, but the actual value B/M for any individual image is very noisy. This makes the estimates of vergence returned by fitting very unreliable.

This problem can be overcome by looking at the response of many sensors, with a variety of receptive-field orientations and phases. This corresponds to looking at the activity of several complex cells. Because the different receptive fields see different aspects of the dot pattern, the total response of this population to any one random-dot pattern is close to its expected total response averaged over all random-dot patterns. This expected total response can be found from Equation 17, summing the expressions for $\langle B \rangle$ and $\langle M \rangle$ over all the receptive fields used in the population. In practice, these expressions are too slow to evaluate for use in a fitting algorithm, because they involve an integration over the entire cyclopean retina. However, excellent results are obtained if we make the approximation that the stimulus disparity remains constant across the receptive field (the stimulus correlation is assumed to be constant at 1). We use receptive fields that are 2D Gabor functions with an isotropic Gaussian envelope. Thus, for the n th unit in the population:

$$\rho(x, y) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \cos(2\pi f(x \cos \theta_n + y \sin \theta_n) + \phi_n). \quad (22)$$

θ_n is the preferred orientation of the n th neuron, and ϕ_n is its overall phase (note that the phase of the Gabor is the same in both eyes; thus the phase disparity is always zero). Under the assumption of constant stimulus disparity, it can be shown that the expected monocular and binocular components of this energy unit's response are:

$$\langle M_n \rangle = I_n + J_n; \quad \langle B_n \rangle = \exp\left(-\frac{\Delta y_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}})^2}{4\sigma^2}\right) [J_n + I_n \cos(2\pi f \Delta y_{\text{stim}}(x_{\text{pref}}, y_{\text{pref}}) \sin \theta_n)], \quad (23)$$

where

$$I_n = \int_{-\infty}^{+\infty} dx_c \int_{-\infty}^{+\infty} dy_c \exp\left(-\frac{(x_c - x_{\text{pref}})^2 + (y_c - y_{\text{pref}})^2}{\sigma^2}\right)$$

$$J_n = \int_{-\infty}^{+\infty} dx_c \int_{-\infty}^{+\infty} dy_c \exp\left(-\frac{(x_c - x_{\text{pref}})^2 + (y_c - y_{\text{pref}})^2}{\sigma^2}\right) \cos(4\pi f((x_c - x_{\text{pref}}) \cos \theta_n + (y_c - y_{\text{pref}}) \sin \theta_n) + 2\phi_n).$$

The double integrals I_n and J_n only have to be calculated once for each neuron in the population; the expected value of $\langle B_n \rangle$ for different eye positions can then be calculated very quickly from Equation 23 (recall that different eye positions imply different vertical-disparity fields Δy_{stim} , according to Equation 16).

Simulations in Figure 10

For clarity, we here summarize exactly how the simulations shown in Figure 10 were produced. A single visual scene was generated, consisting of 10,000 black and white dots placed at random over the surface of an exploded sphere (Figure 8). The images of each dot were projected onto the two planar retinae to obtain the positions (x_{Lj}, y_{Lj}) and (x_{Rj}, y_{Rj}) at which the j th dot struck each retina. For each sensor, the output from each eye's receptive field was calculated by summing the values of the receptive field at each white dot position and subtracting the values of the receptive field at each black dot position. Thus, for the n th sensor:

$$L_n = \sum_{j=1}^{10,000} c_j \rho_n \left(x_{Lj} - x_{\text{pref}} + \frac{\Delta x_{\text{pref}}}{2}, y_{Lj} - y_{\text{pref}} \right); \quad R_n = \sum_{j=1}^{10,000} c_j \rho_n \left(x_{Rj} - x_{\text{pref}} - \frac{\Delta x_{\text{pref}}}{2}, y_{Rj} - y_{\text{pref}} \right), \quad (24)$$

where c_j is +1 for white dots and -1 for black dots. The monocular and binocular components for each sensor were then computed as $B_n = 2L_n R_n$, $M_n = L_n^2 + R_n^2$. At each cyclopean location shown in Figure 10, we calculated B_n and M_n for 30 simple cells, with Gabor receptive fields (Equation 22). The 30 units were made up of three different orientations ($\theta = 0^\circ, 60^\circ, 120^\circ$) and 10 different receptive-field phases ($\phi = 0^\circ, 36^\circ, \dots, 288^\circ, 324^\circ$). In each case, the spatial-frequency full-width half-power bandwidth was 1.5 octaves, the preferred spatial frequency was 0.3 cpd, and the envelope was an isotropic Gaussian with an SD of 1° . For every binocular unit, the receptive-field profiles were identical in the two eyes. The receptive-field positions differed only horizontally. Each unit was given a horizontal position disparity equal to the stimulus horizontal disparity at the center of its cyclopean receptive field.

Figure 10A shows the ratio B_1/M_1 for one sensor in the population, with orientation $\theta = 0^\circ$ and phase $\phi = 0^\circ$. This is very noisy, reflecting the wide variation depending on the particular pattern of black and white dots experienced by sensors in different parts of the retina. Figure 10B shows what happens if we first sum the binocular and monocular components over all sensors in the population, before taking the ratio, that is, $(\sum_n B_n)/(\sum_n M_n)$. This surface is much smoother. For comparison, Figure 10C shows the expected values, $(\sum_n \langle B_n \rangle)/(\sum_n \langle M_n \rangle)$, which we would expect to get if we averaged the binocular and monocular components obtained from many different random-dot stimuli. Because we have summed over 30 units with different receptive-field properties, the value obtained from just one random-dot pattern (Figure 10B) is very similar to the value expected from averaging over all possible random-dot patterns (Figure 10C).

To recover an estimate of eye position, we assumed that the summed response ratio shown in Figure 10A and the horizontal stimulus field shown in Figure 9B are both computed in the brain. For a particular value of gaze angle H_c and vergence D , the predicted vertical-disparity field, Δy_{pred} , can be obtained from Equation 16. Because the properties of each sensor (σ, θ, ϕ) are known, approximate expressions for the expected components for each sensor, $\langle B_n \rangle$ and $\langle M_n \rangle$, can be calculated from Equation 23. Recall that this ignores variation in stimulus disparity across a receptive field. The predicted correlation

$$\frac{\sum_n \langle B_n \rangle}{\sum_n \langle M_n \rangle} = \exp \left(-\frac{\Delta y_{\text{pred}}(x_{\text{pref}}, y_{\text{pref}})^2}{4\sigma^2} \right) \frac{\sum_n [J_n + I_n \cos(2\pi f \Delta y_{\text{pred}}(x_{\text{pref}}, y_{\text{pref}}) \sin \theta_n)]}{\sum_n [J_n + I_n]} \quad (25)$$

is then compared with the value actually obtained for this dot pattern at this cyclopean location. The fitting algorithm finds the values of gaze angle and vergence that produce the closest match between the predicted and actual results. Figure 10D shows the best match found in this case.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, National Eye Institute, and by a Royal Society University Research Fellowship awarded to J.C.A.R. Thanks to Chris Hillman and Mark Szarowicz for being excellent psychophysical subjects.

Commercial relationships: none.
Corresponding author: Jenny Read.

Email: j.c.a.read@ncl.ac.uk.

Address: Henry Wellcome Building for Neuroecology,
Newcastle University, Newcastle upon Tyne, NE2 4HH, UK.

Footnotes

¹Note that “sign of vertical disparity” is sometimes used (e.g., Westheimer, 1978), in the context of the induced effect, to

mean which eye's image is magnified: "negative vertical disparity" means that the left eye's image is larger than the right and so forth. This is not the sense in which we use the term.

²Provided the absolute horizontal disparity of the viewed objects remains small and the vergence angle is large, as in this example where the fixation point is in the plane of the square and the vergence is 10°. See [Retinal locus of zero vertical disparity \(for zero elevation\)](#) section.

References

- Adams, W., Frisby, J. P., Buckley, D., Garding, J., Hippisley-Cox, S. D., & Porrill, J. (1996). Pooling of vertical disparities by the human visual system. *Perception*, *25*, 165–176. [[PubMed](#)]
- Allison, R. S., Howard, I. P., & Fang, X. (2000). Depth selectivity of vertical fusional mechanisms. *Vision Research*, *40*, 2985–2998. [[PubMed](#)]
- Backus, B. T., & Banks, M. S. (1998). Vertical disparity: Absolute or relative? *Investigative Ophthalmology & Visual Science*, *39*, S616.
- Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*, 1143–1170. [[PubMed](#)]
- Banks, M. S., & Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vision Research*, *38*, 187–194. [[PubMed](#)]
- Banks, M. S., Backus, B. T., & Banks, R. S. (2002). Is vertical disparity used to determine azimuth? *Vision Research*, *42*, 801–807. [[PubMed](#)]
- Banks, M. S., Hooge, I. T., & Backus, B. T. (2001). Perceiving slant about a horizontal axis from stereopsis. *Journal of Vision*, *1*(2), 55–79, <http://journalofvision.org/1/2/1/>, doi:10.1167/1.2.1. [[PubMed](#)] [[Article](#)]
- Barlow, H. (1961). Possible principles underlying the transformation of sensory messages. In W. Rosenblith (Ed.), *Sensory communication* (pp. 217–234). Cambridge, MA: MIT Press.
- Berends, E. M., & Erkelens, C. J. (2001). Strength of depth effects induced by three types of vertical disparity. *Vision Research*, *41*, 37–45. [[PubMed](#)]
- Berends, E. M., van Ee, R., & Erkelens, C. J. (2002). Vertical disparity can alter perceived direction. *Perception*, *31*, 1323–1333. [[PubMed](#)]
- Brenner, E., Smeets, J. B., & Landy, M. S. (2001). How vertical disparities assist judgements of distance. *Vision Research*, *41*, 3455–3465. [[PubMed](#)]
- Busettoni, C., Fitzgibbon, E. J., & Miles, F. A. (2001). Short-latency disparity vergence in humans. *Journal of Neurophysiology*, *85*, 1129–1152. [[PubMed](#)] [[Article](#)]
- Clement, R. A. (1992). Gaze angle explanations of the induced effect. *Perception*, *21*, 355–357. [[PubMed](#)]
- Collewijn, H., Erkelens, C. J., & Steinman, R. M. (1997). Trajectories of the human binocular fixation point during conjugate and non-conjugate gaze-shifts. *Vision Research*, *37*, 1049–1069. [[PubMed](#)]
- Cumming, B. G. (2002). An unexpected specialization for horizontal disparity in primate primary visual cortex. *Nature*, *418*, 633–636. [[PubMed](#)]
- Duke, P. A., & Howard, I. P. (2005). Vertical-disparity gradients are processed independently in different depth planes. *Vision Research*, *45*, 2025–2035. [[PubMed](#)]
- Durand, J. B., Zhu, S., Celebrini, S., & Trotter, Y. (2002). Neurons in parafoveal areas V1 and V2 encode vertical and horizontal disparities. *Journal of Neurophysiology*, *88*, 2874–2879. [[PubMed](#)] [[Article](#)]
- Duwaer, A. L., & van den Brink, G. (1982). Detection of vertical disparities. *Vision Research*, *22*, 467–478. [[PubMed](#)]
- Farell, B. (1998). Two-dimensional matches from one-dimensional stimulus components in human stereopsis. *Nature*, *395*, 689–693. [[PubMed](#)]
- Farell, B. (2003). Detecting disparity in two-dimensional patterns. *Vision Research*, *43*, 1009–1026. [[PubMed](#)]
- Friedman, R. B., Kaye, M. G., & Richards, W. (1978). Effect of vertical disparity upon stereoscopic depth. *Vision Research*, *18*, 351–352. [[PubMed](#)]
- Frisby, J. P., Buckley, D., Grant, H., Garding, J., Horsman, J. M., Hippisley-Cox, S. D., et al. (1999). An orientation anisotropy in the effects of scaling vertical disparities. *Vision Research*, *39*, 481–492. [[PubMed](#)]
- Garding, J., Porrill, J., Mayhew, J. E., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, *35*, 703–722. [[PubMed](#)]
- Gillam, B., Chambers, D., & Lawergren, B. (1988). The role of vertical disparity in the scaling of stereoscopic depth perception: An empirical and theoretical study. *Perception & Psychophysics*, *44*, 473–483. [[PubMed](#)]
- Gillam, B., & Lawergren, B. (1983). The induced effect, vertical disparity, and stereoscopic theory. *Perception & Psychophysics*, *34*, 121–130. [[PubMed](#)]
- Gonzalez, F., Justo, M. S., Bermudez, M. A., & Perez, R. (2003). Sensitivity to horizontal and vertical disparity and orientation preference in areas V1 and V2 of the monkey. *Neuroreport*, *14*, 829–832. [[PubMed](#)]
- Gonzalez, F., Relova, J. L., Perez, R., Acuna, C., & Alonso, J. M. (1993). Cell responses to vertical and

- horizontal retinal disparities in the monkey visual cortex. *Neuroscience Letters*, *160*, 167–170. [[PubMed](#)]
- Helmholtz, H. v. (1925). *Treatise on physiological optics*. Rochester, NY: Optical Society of America.
- Hering, E. (1942). *Spatial sense and movements of the eye*. Baltimore: American Academy of Optometry.
- Hillebrand, F. (1893). Die Stabilität der Raumwerte auf der Netzhaut. *Zeitschriften der Psychologischen und Physiologischen Sinnesorganen*, *5*, 1–60.
- Howard, I. P., Allison, R. S., & Zacher, J. E. (1997). The dynamics of vertical vergence. *Experimental Brain Research*, *116*, 153–159. [[PubMed](#)]
- Howard, I. P., Fang, X., Allison, R. S., & Zacher, J. E. (2000). Effects of stimulus size and eccentricity on horizontal and vertical vergence. *Experimental Brain Research*, *130*, 124–132. [[PubMed](#)]
- Howard, I. P., & Pierce, B. J. (1998). Types of shear disparity and the perception of surface inclination. *Perception*, *27*, 129–145. [[PubMed](#)]
- Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford: Oxford University Press.
- Ito, H. (2005). Illusory depth perception of oblique lines produced by overlaid vertical disparity. *Vision Research*, *45*, 931–942. [[PubMed](#)]
- Kaneko, H., & Howard, I. P. (1996). Relative size disparities and the perception of surface slant. *Vision Research*, *36*, 1919–1930. [[PubMed](#)]
- Kaneko, H., & Howard, I. P. (1997a). Spatial limitation of vertical-size disparity processing. *Vision Research*, *37*, 2871–2878. [[PubMed](#)]
- Kaneko, H., & Howard, I. P. (1997b). Spatial properties of shear disparity processing. *Vision Research*, *37*, 315–323. [[PubMed](#)]
- Koenderink, J. J., & van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, *21*, 29–35. [[PubMed](#)]
- Longuet-Higgins, H. C. (1982). The role of the vertical dimension in stereoscopic vision. *Perception*, *11*, 377–386. [[PubMed](#)]
- Maunsell, J. H., & Van Essen, D. C. (1983). Functional properties of neurons in middle temporal visual area of the macaque monkey: II. Binocular interactions and sensitivity to binocular disparity. *Journal of Neurophysiology*, *49*, 1148–1167. [[PubMed](#)]
- Mayhew, J. E. (1982). The interpretation of stereo-disparity information: The computation of surface orientation and depth. *Perception*, *11*, 387–403. [[PubMed](#)]
- Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, *297*, 376–378. [[PubMed](#)]
- McCandless, J. W., Schor, C. M., & Maxwell, J. S. (1996). A cross-coupling model of vertical vergence adaptation. *IEEE Transactions on Biomedical Engineering*, *43*, 24–34. [[PubMed](#)]
- McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, *30*, 1763–1779. [[PubMed](#)]
- Nieder, A., & Wagner, H. (2001). Encoding of both vertical and horizontal disparity in random-dot stereograms by Wulst neurons of awake barn owls. *Visual Neuroscience*, *18*, 541–547. [[PubMed](#)]
- Ogle, K. (1964). *Researches in binocular vision*. New York: Hafner.
- Ogle, K. N. (1938). Induced size effect I: A new phenomenon in binocular vision associated with the relative size of the images in the two eyes. *Archives of Ophthalmology*, *20*, 604.
- Ogle, K. N. (1952). Space perception and vertical disparity. *Journal of the Optical Society of America*, *42*, 145–146. [[PubMed](#)]
- Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America*, *43*, 907–913. [[PubMed](#)]
- Ogle, K. N. (1954). Stereopsis and vertical disparity. *A.M.A. Archives of Ophthalmology*, *53*, 495–504. [[PubMed](#)]
- Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science*, *249*, 1037–1041. [[PubMed](#)]
- Petrov, A. P. (1980). A geometrical explanation of the induced size effect. *Vision Research*, *20*, 409–413. [[PubMed](#)]
- Pettet, M. W. (1997). Spatial interactions modulate stereoscopic processing of horizontal and vertical disparities. *Perception*, *26*, 693–706. [[PubMed](#)]
- Pierce, B. J., & Howard, I. P. (1997). Types of size disparity and the perception of surface slant. *Perception*, *26*, 1503–1517. [[PubMed](#)]
- Pierce, B. J., Howard, I. P., & Feresin, C. (1998). Depth interactions between inclined and slanted surfaces in vertical and horizontal orientations. *Perception*, *27*, 87–103. [[PubMed](#)]
- Poggio, G. E. (1995). Mechanisms of stereopsis in monkey visual cortex. *Cerebral Cortex*, *5*, 193–204. [[PubMed](#)]
- Porrill, J., Frisby, J. P., Adams, W. J., & Buckley, D. (1999). Robust and optimal use of information in stereo vision. *Nature*, *397*, 63–66. [[PubMed](#)]
- Prince, S. J., Pointon, A. D., Cumming, B. G., & Parker, A. J. (2002). Quantitative analysis of the responses of

- V1 neurons to horizontal disparity in dynamic random-dot stereograms. *Journal of Neurophysiology*, 87, 191–208. [[PubMed](#)] [[Article](#)]
- Rashbass, C., & Westheimer, G. (1961). Independence of conjugate and disjunctive eye movements. *Journal of Physiology*, 159, 361–364. [[PubMed](#)] [[Article](#)]
- Read, J. C., & Cumming, B. G. (2003). Testing quantitative models of binocular disparity selectivity in primary visual cortex. *Journal of Neurophysiology*, 90, 2795–2817. [[PubMed](#)] [[Article](#)]
- Read, J. C. A., & Cumming, B. G. (2004). Understanding the cortical specialization for horizontal disparity. *Neural Computation*, 16, 1983–2020. [[PubMed](#)] [[Article](#)]
- Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, 361, 253–255. [[PubMed](#)]
- Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, 24, 155–179. [[PubMed](#)]
- Schor, C. M., Maxwell, J. S., & Stevenson, S. B. (1994). Isovergence surfaces: The conjugacy of vertical eye movements in tertiary positions of gaze. *Ophthalmic & Physiological Optics*, 14, 279–286. [[PubMed](#)]
- Schreiber, K., Crawford, J. D., Fetter, M., & Tweed, D. (2001). The motor side of depth vision. *Nature*, 410, 819–822. [[PubMed](#)]
- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24, 1193–1216. [[PubMed](#)]
- Stenton, S. P., Frisby, J. P., & Mayhew, J. E. (1984). Vertical disparity pooling and the induced effect. *Nature*, 309, 622–623. [[PubMed](#)]
- Trotter, Y., Celebrini, S., & Durand, J. B. (2004). Evidence for implication of primate area V1 in neural 3-D spatial localization processing. *Journal of Physiology (Paris)*, 98, 125–134. [[PubMed](#)]
- Wei, M., DeAngelis, G. C., & Angelaki, D. E. (2003). Do visual cues contribute to the neural estimate of viewing distance used by the oculomotor system? *Journal of Neuroscience*, 23, 8340–8350. [[PubMed](#)] [[Article](#)]
- Westheimer, G. (1978). Vertical disparity detection: Is there an induced size effect? *Investigative Ophthalmology & Visual Science*, 17, 545–551. [[PubMed](#)]
- Westheimer, G. (1984). Sensitivity for vertical retinal image differences. *Nature*, 307, 632–634. [[PubMed](#)]
- Westheimer, G., & Pettet, M. W. (1992). Detection and processing of vertical disparity by the human observer. *Proceedings of the Royal Society of London Series B: Biological Sciences*, 250, 243–247. [[PubMed](#)] [[Article](#)]
- Williams, T. D. (1970). Vertical disparity in depth perception. *American Journal of Optometry and Archives of American Academy of Optometry*, 47, 339–344. [[PubMed](#)]
- Yang, D. S., FitzGibbon, E. J., & Miles, F. A. (2003). Short-latency disparity-vergence eye movements in humans: Sensitivity to simulated orthogonal tropias. *Vision Research*, 43, 431–443. [[PubMed](#)]