A laminar cortical model of stereopsis and three-dimensional surface perception

Stephen Grossberg and Piers D. L. Howe[1]

Department of Cognitive and Neural Systems
and
Center for Adaptive Systems
Boston University
677 Beacon Street, Boston, MA 02215, USA

Running title: Laminar cortical model of depth perception

Submitted: March, 2002
Revised: August, 2002

Technical Report CAS/CNS-2002-002

All correspondence should be addressed to
Professor Stephen Grossberg
Department of Cognitive and Neural Systems
Boston University
677 Beacon Street
Boston, MA 02215
Phone: 617-353-7858
Fax: 617-353-7755
Email:steve@bu.edu

[1]Authorship in alphabetical order. SG was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397) and the Office of Naval Research (ONR N00014-01-1-0624). PDLH was supported in part by the Air Force Office of Scientific Research (AFOSR F49620-01-1-0397), the Defense Advanced Research Projects Agency and the Office of Naval Research (ONR N00014-95-1-0409), and the Office of Naval Research (ONR N00014-95-1-0657).

Abstract
A laminar cortical model of stereopsis and later stages of 3D surface perception is developed and simulated. The model describes how initial stages of monocular and binocular oriented filtering interact with later stages of 3D boundary formation and surface filling-in in the lateral geniculate nucleus (LGN) and cortical areas V1, V2, and V4. In particular, it details how interactions between layers 4, 3B, and 2/3A in V1 and V2 contribute to stereopsis, and clarifies how binocular and monocular information combine to form 3D boundary and surface representations. Along the way, the model modifies and significantly extends the disparity energy model. Neural explanations are given for psychophysical data concerning: contrast variations of dichoptic masking and the correspondence problem, the effect of interocular contrast differences on stereoacuity, Panum's limiting case, the Venetian blind illusion, stereopsis with polarity-reversed stereograms, da Vinci stereopsis, and various lightness illusions. By relating physiology to psychophysics, the model provides new functional insights and predictions about laminar cortical architecture.

Keywords: Cortical model; Depth perception; Stereopsis; Surface perception; Cortical layers; Lightness perception; Monocular-binocular interactions

1 Introduction

This article describes a model of how the LGN and cortical areas V1, V2 and V4 utilize both monocular and binocular visual information to produce three-dimensional (3D) surface percepts. Despite some explanatory successes, many previous cortical models, for example the disparity energy model of Ohzawa, DeAngelis and Freeman (1990), considered only stereopsis which is an early stage of depth perception that occurs in cortical area V1. Stereopsis is important, but on its own is insufficient to explain the 3D surface percepts that form an integral part of our visual consciousness.

The present model, shown in Figure 1, goes beyond these previous analyses in several ways. First, it provides a refined model of stereopsis in V1 which clarifies the role of cells in cortical layers 4, 3B, and 2/3A. In particular, the model revises how the disparity energy model achieves stereopsis, in a manner that is more consistent with recent data. Second, the model shows how monocular and binocular information are combined and selected in V2 to form 3D boundary representations. Third, the model shows how these 3D boundaries give rise to visible 3D surface percepts in V4. Taken together, these model processes are used to explain and simulate a much larger set of neurophysiological, anatomical, and psychophysical data about stereopsis and 3D surface perception than has previously been possible.

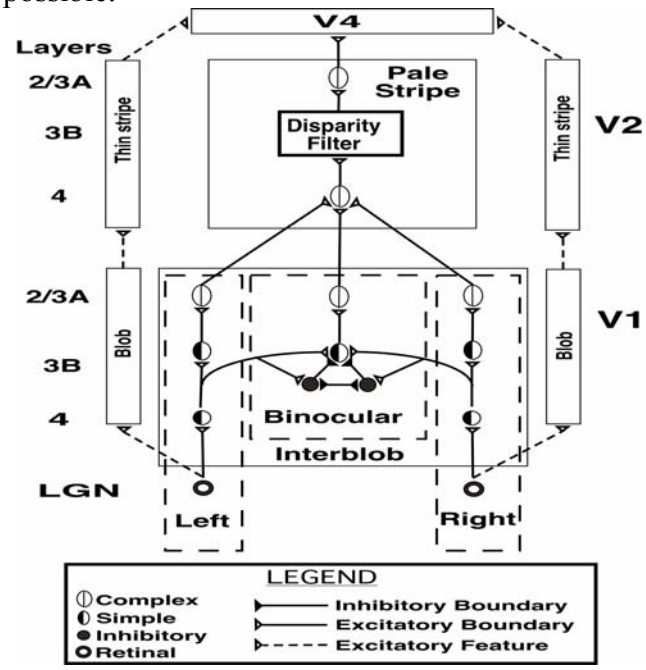

Figure 1: Model circuit diagram.

The model's explanatory range is larger still since it is consistent with, and generalizes, a recent laminar model of V1 and V2, called the LAMINART model, that does not incorporate binocular interactions (Grossberg, 1999; Grossberg, Mingolla & Ross, 1997; Grossberg & Raizada, 2000; Raizada & Grossberg, 2001; Grossberg & Williamson, 2001). The LAMINART model explained data about perceptual development, learning, grouping, and attention. Because the present model consistently generalizes the LAMINART model to 3D vision, it is called the 3D LAMINART model. The Discussion section outlines how this synthesis leads to a model that predicts how cellular and network mechanisms of 3D vision are linked to mechanisms of development, learning, grouping, and attention.

The model achieves these goals by embodying five basic psychophysical constraints in its neural circuitry:

(1) *Reconciles contrast-specific binocular fusion with contrast-invariant boundary perception.* It is well known that only edges in the left and right retinal images that have the same contrast polarity (i.e., their luminance gradients have the same sign) can be binocularly fused to form a percept of depth (Howard & Rogers, 1995). Otherwise expressed, binocular fusion obeys the *same-sign hypothesis;* see Figure 2. However, fused boundaries must also be able to form around objects whose contrast polarity with respect to the background can reverse along their perimeters (Grossberg, 1994). In other words, binocular boundaries need to be represented in a contrast-invariant way. How can the brain reconcile contrast-specific fusion with the need to form contrast-invariant object boundaries? The model proposes that both constraints are realized by interactions between cells in layers 4, 3B, and 2/3A of cortical area V1 interblobs; see Figure 1.

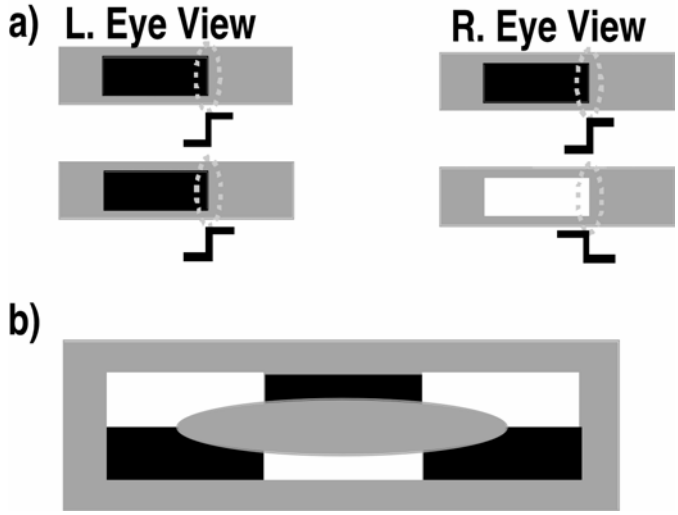

Figure 2: (a) The same-sign hypothesis: only edges that have the same contrast polarity can be stereoscopically fused to produce a percept of depth; (b) As it is traversed, the boundary of the ellipse changes its contrast polarity relative to the background, thereby illustrating the need for object boundaries to be represented in a contrast-invariant manner. See text for details.

(2) *Implements the contrast constraint on binocular fusion.* The brain needs to determine which of the many potential same-sign edges in the two retinal images should be binocularly fused, since veridical stereoscopic depth perception will occur only if the two edges belong to the same object. This is commonly referred to as the *correspondence problem* (Julesz, 1971; Howard & Rogers, 1995). An early step in solving the correspondence problem is to binocularly fuse only edges with approximately the same magnitude of contrast (McKee et al., 1994). This constraint naturally arises when the brain fuses edges that derive from the same objects in the world. The model satisfies this constraint through interactions between excitatory and inhibitory cells in layer 3B of V1 that endow the binocular cells there with an *obligate property* (Poggio, 1991), whereby they respond preferentially to left and right eye inputs of approximately equal size.

(3) *Solves the correspondence problem.* Even if all binocular matches are of the same-sign and similar contrast magnitude, there can still exist many false binocular matches between edges that did not derive from the same objects. This problem has often been approached by imposing a *unique-matching rule*, which states that any given feature in one retinal image is matched at most with one feature in the other retinal image (Marr & Poggio, 1976; Grimson, 1981; for a review see Howard & Rogers, 1995, pp. 42-43). However, this rule fails in situations like Panum's limiting case (Panum, 1858; Gillam et al., 1995; McKee et al., 1995) where a bar presented to one eye is simultaneously

matched to two separate bars presented to the other eye. The present model does not enforce unique matches. Rather, the model encourages them by using a *disparity filter* that is proposed to occur in the pale stripes of cortical area V2, possibly in layer 3B; see Figure 3. This disparity filter uses two types of inhibitory interactions: line-of-sight inhibition and inhibition across depth but within cyclopean position, to encourage unique matches, as described in more detail below.

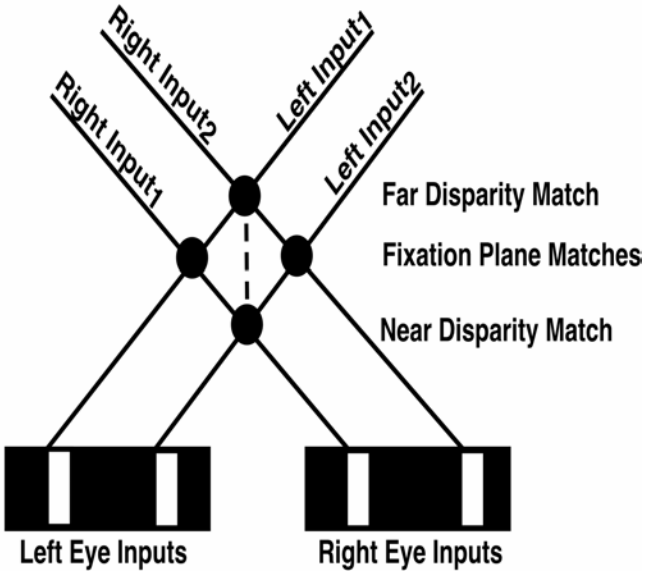

Figure 3: The V2 disparity filter. The V1 binocular boundaries network matches an edge in one retinal image with every other edge in the other retinal image whose relative disparity is not too great, that has the same contrast polarity and whose magnitude of contrast is not too different. In response to this image, the V1 boundary network creates four matches, with the two not in the fixation plane being false matches between edges that do not correspond to the same object. As described in the text, these false matches are suppressed by the disparity filter in V2, wherein each neuron is inhibited by every other neuron that shares either of its monocular inputs (i.e., shares a monocular line-of-sight represented by the solid lines) or is directly in front of or behind it (i.e., is connected to it by a dashed line). Note in particular that the solid lines that represent the monocular lines-of-sight also represent the allelotropic shifts: an edge in the left retinal image is shifted to the right for matches increasingly further away whereas an edge in the right retinal image is shifted in the opposite direction.

(4) *Combines monocular and binocular information to form depth percepts*. Although Panum's limiting case seems to be a laboratory curiosity at first, there are many naturally occurring situations where there is only one edge in one eye and two possible edges with which to match it in the other eye. For example, due to the lateral displacement of the eyes, an object's edge that is seen by one eye may be occluded in the other eye, as occurs during da Vinci stereopsis (Nakayama & Shimojo, 1990). Despite this lack of binocular information, the monocularly viewed region has a definite depth conferred to it by the binocularly viewed parts of the scene. The brain can thus utilize monocular information to build up seamless 3D percepts of the world. In fact, in experiments involving Panum's limiting case, varying the relative contrast of the bars alters the perception of depth in a manner that reveals clear monocular-binocular interactions (Smallman & McKee, 1995). Dichoptic masking, where an object presented to one eye is obscured (i.e., masked) by one presented to the other eye, illustrates a third way in which monocular and binocular information may interact (McKee et al., 1994).

Once monocular information is included, the problem immediately arises about how to combine monocular and binocular boundaries. This is a problem because monocular boundaries do not have a definite depth associated with them. How, then, can we decide to which depth they should be assigned? A proposed approach to this *monocular-binocular interface problem* was suggested in Grossberg (1994, 1997) in order to explain data about 3D figure-ground perception. Here the same hypothesis is shown to play a crucial role in explaining many other data about 3D surface perception. Namely, the model assumes that the outputs of the monocular boundary cells are added to all depth planes in the pale stripes of cortical area V2 along their respective lines-of-sight, possibly in layer 4; see Figures 1 and 4a. The disparity filter, which helps to solve the correspondence problem, also solves the monocular-binocular interface problem by automatically eliminating most of the monocular boundaries that are not at the correct depths.

(5) *Forms 3D surface percepts*. So far we have considered only how the brain constructs a 3D boundary representation of an object. There is considerable evidence that boundary representations on their own do not give rise to visible percepts which rather are a property of surface representations (Grossberg, 1994). In the present model surface representations derive from a filling-in process whereby lightness and color mark the depths at which the surfaces occur. Filling-in is needed to recover lightness and color estimates in regions where they have been suppressed by the process of discounting the illuminant (Grossberg & Todorović, 1988). Boundaries control the depths at which particular lightnesses and colors can fill-in, a process that we call 3D *surface capture*. The present article considers only the filling-in of achromatic lightness.

How does the brain ensure that lightness fills-in at only the correct depths? Grossberg (1994) proposed properties of this boundary-surface interaction that helped to explain many data about 3D figure-ground perception. Here, one of these properties proved essential to explain 3D surface percepts that arise in stereopsis research. Namely, visible surfaces arise in cortical area V4 only if they are enclosed by *connected* boundaries; see Figure 4. In particular, a rectangular connected boundary may be composed of one vertical binocular boundary, one vertical monocularly viewed boundary, and two horizontal boundaries that code no disparity information. This connected boundary can support a visible surface percept at the depth corresponding to the binocular boundary if all other constraints are satisfied. Such a boundary can contain the filling-in process. However, if the vertical binocular boundary is missing, as it would be at a different depth plane, then the total boundary is not connected, and a visible percept will not be evident at that depth because filling-in can dissipate out of the boundary gap. This example illustrates how the monocular-binocular interface problem (item (4) above), and thus the correspondence problem (item (3) above), influence visible percepts of 3D surfaces.

The present model refines aspects of the FACADE model of 3D vision and figure-ground perception (Grossberg, 1994, 1997). The FACADE model included a (non-laminar) model of stereopsis and 3D planar surface perception (Grossberg & McLoughlin, 1997; McLoughlin & Grossberg, 1998) that modified and generalized the disparity energy model of stereopsis (Ohzawa et al., 1990). This generalization incorporated rectification prior to binocular combination, absent from the original disparity energy model, which has recently received independent experimental support (Read et al., 2002; Cumming, 2002). It also proposed that positional shifts between left and right eye cortical inputs code disparities, rather than phase shifts, which has also received experimental support (Tsao & Livingstone, In Press). The FACADE model also incorporated a disparity filter to help solve the correspondence problem (Howard & Rogers, 1995) as well as mechanisms for filling-in 3D surface percepts from 3D boundary representations. In particular, the FACADE model explained the fact that stereoscopic fusion is *generally* impossible when the left and right eye stimuli differ too much in contrast (Smallman & McKee, 1995). However, in the form developed by Grossberg and McLoughlin, the FACADE model could not explain why stereoscopic fusion is *always* possible in

the special case where each eye sees only a single bar, regardless of the contrast difference of the two bars (McKee et al., 1994; Smallman & McKee, 1995).

The 3D LAMINART model overcomes this limitation using identified cells in laminar circuits, and resimulates all the data previously simulated by McLoughlin & Grossberg (1998), in particular the data on contrast variations of the correspondence problem and dichoptic masking. In addition, the new model can simulate still more psychophysical data than its non-laminar predecessors, including: the Venetian blind illusion, four different examples of da Vinci stereopsis (Nakayama & Shimojo, 1990; Gillam et al., 1999), stereopsis with opposite-contrast stimuli, the effect of interocular contrast differences on stereoacuities and various lightness illusions. In so doing, it demonstrates more of the roles that boundary and surface representations play in depth perception. The 3D LAMINART model also makes neurophysiological predictions, including that there exist: (1) In V1 cells that obey the ratio constraint on binocular fusion. The model proposes that some binocular simple cells in layer 3B obey an obligate property whereby they can be activated only if they receive approximately equal inputs from both left and right eye monocular simple cells in layer 4. The constraints that determine cell firing depend upon the ratios of left and right monocular cell activity. This property explains the ratio constraint on stereoscopic fusion that is illustrated in Figure 10 below. The obligate property is predicted to be caused by a balance between excitatory inputs from layer 4 monocular simple cells and inhibitory inputs from layer 3B inhibitory interneurons. The interneurons are themselves activated by layer 4 monocular simple cells and mutually inhibit each other, in addition to inhibiting the binocular simple cells. (2) In V2 cells that solve the correspondence problem using a disparity filter. (3) In V4 a filling-in mechanism that completes visible 3D surface representations within connected boundaries. These results were briefly reported in Howe and Grossberg (2001).

2 Model description

The model consists of four component networks which process: V1 binocular boundaries, V1 monocular boundaries, V2 boundaries, and V4 surfaces. For a mathematical description, the reader is referred to Appendix A. A description of the neurophysiological and anatomical evidence that supports all the model processing stages is found in Section 4.1. In order to reduce the computational load, the model currently considers only horizontal and vertical contours and five depth planes. Even so, the model includes approximately 185,000-333,000 cells depending on the simulation. Although model cells and cells in vivo will be clearly distinguished in the text, model cells will be referred to by physiological labels because their properties so closely match those found *in vivo.*

2.1 V1 binocular boundaries

The network that processes the V1 binocular boundaries is located in the V1 interblob region and includes the binocular cells in layers 3B and 2/3A. It carries out stereoscopic fusion of vertical contours, but not of horizontal contours, which it assumes cannot be stereoscopically fused. This network implements the *same-sign hypothesis*; see Section 1.1. As shown in Figure 1, inputs to the left and right eyes activate monocular simple cells in layer 4 of the V1 interblob regions. Left and right eye monocular simple cells conjointly activate binocular simple cells in layer 3B whose depth sensitivity is determined by the relative retinal disparity of the layer 4 monocular cells that project to them. The model implements the same-sign hypothesis by assuming that only layer 4 simple cells with the same contrast polarity project to a single layer 3B simple cell. These layer 3B simple cells are therefore selective for binocular disparity and a prescribed contrast polarity. Binocularly fused vertical contours that occupy corresponding points on the two retinas are seen as a single boundary in the fixation plane, whereas vertical contours that are displaced relative to each other are seen as a single boundary either in front of or behind the fixation plane, depending on their displacement, as detailed in Appendix A, equation (A10).

There are also inhibitory cells in layer 3B. As is described in Appendix B, these cells ensure that the binocular simple cells act like the "obligate cells" of Poggio (1991): The activity of such a binocular simple cell is suppressed by these inhibitory interneurons if the magnitudes of the left and right eye inputs differ too much; see Section 1.2. In particular, these obligate cells respond to binocular, but not to monocular, stimulation. These obligate cells help to solve the correspondence problem by ensuring that only similar stimuli in the left and right eye retinal images are stereoscopically fused.

The next processing stage implements contrast-invariant boundary detection; see Section 1.1. Layer 3B simple cells that are sensitive to the same position and disparity, but *opposite* contrast polarities, pool their signals at layer 2/3A complex cells. These complex cells therefore respond to both contrast polarities and so can generate three-dimensional object boundaries even if the object's contrast polarity, with respect to the background, reverses as the boundary is transversed. In summary, the two layers 3B and 2/3A, acting together, can realize the same-sign hypothesis and also begin to compute object boundaries in front of textured backgrounds.

These proposed interactions between layers 4, 3B and 2/3A are consistent with neurophysiological data, as detailed in Section 4.1, and instantiate key operations of the disparity energy model (Ohzawa et al., 1990), which itself is strongly supported by physiological evidence; for a review, see Ohzawa (1998). As discussed in Section 1, the need for preprocessing before the site of binocular combination, such as that carried out by layer 4 of our model, has recently been demonstrated by Read et al. (2002) and Cumming (2002), who showed that preprocessing was required to explain subtleties in physiological data not captured by the original disparity energy model.

2.2 V1 monocular boundaries

The network that processes the V1 monocular boundaries comprises the monocular cells in layers 4, 3B and 2/3A of the V1 interblob region. It is similar to the binocular boundaries network, but represents both horizontal and vertical boundaries whereas the binocular boundaries network represents only vertical boundaries. Binocular boundary cells preferentially represent a particular depth plane, but this is not true of monocular boundary cells. How, then, do monocular and binocular boundaries interact? A proposed solution of this *monocular-binocular boundary interface problem* assumes that the outputs of the monocular boundary cells are added to *all* depth planes in cortical area V2 along their respective lines-of-sight; see Section 1.4 and Figure 3. Appendix A, equations (A12)-(A13), describe this process quantitatively.

As noted in Section 1.4, the V2 disparity filter helps to solve the monocular-binocular interface problem, as well as the correspondence problem, by eliminating most of the monocular representations that are not at the correct depth. This previously unexpected property of the disparity filter is crucial to understanding the monocular-binocular interactions described in this paper. It can best be understood by studying the model simulations in Section 3.

2.3 V2 boundaries

The disparity filter network that processes V2 boundaries is located in the V2 pale stripes; see Section 1.3. The V1 binocular boundaries network attempts to match every vertical edge in one retinal image with every other nearby vertical edge in the other retinal image that has the same contrast polarity and approximately the same magnitude of contrast. Figure 3 shows the resultant matches if each eye sees two bars. V1 makes four matches. Only the two in the fixation plane are correct matches. The other two are false matches between retinal images that do not correspond to the same object. Such false matches are known to occur in V1 but less readily in V2 (Bakin, Nakayama, & Gilbert, 2000; Cumming & Parker, 2000). As they typically do not give a veridical depth perception, these false matches must be suppressed.

Figure 3 illustrates how the disparity filter works. To encourage unique-matching, the model assumes that each neuron inhibits all other neurons that share either of its monocular inputs; that is, shares one of its monocular lines-of-sight. This is represented by the solid lines between neurons in Figure 3. This rule on its own could ensure that only two of the four initial matches in Figure 3 survive, but it could not guarantee that it is the false matches that are suppressed. A second form of inhibition ensures this. This inhibition acts across depth and within cyclopean position. It is represented by the dashed line between each neuron with every other neuron that is directly in front of or behind it. These two types of inhibition work together to ensure that the two matches in the fixation plane typically win, thereby solving the correspondence problem. It should be stressed that the disparity filter operates only on vertically oriented cells, as the model assumes that horizontal boundaries cannot be fused and therefore cannot give rise to false matches. It will be shown in Section 3.1.4 how this filter is also able to explain how, in some situations, double matching can occur, as in Panum's limiting case, an example of stereopsis that many previous models (e.g., Marr & Poggio, 1976; Grimson, 1981) could not explain.

2.4 Surfaces

Boundaries help give rise to 3D surface percepts in the manner summarized in Section 1.5. Although our main goal is to explain percepts of surface depth, percepts of surface lightness are also simulated to show that our development of cortical depth perception mechanisms are consistent with simulations in related modeling studies of surface brightness and lightness (e.g., Grossberg & Kelly, 1999; Kelly & Grossberg, 2000). Such a unified set of simulations supports the key FACADE prediction that the same process fills-in surface lightness, color, and depth (Grossberg, 1994). Previous simulations of lightness often focused on computing the *relative* lightnesses of surface regions (but see Grossberg, Mingolla, & Williamson, 1995). Once relative lightness is estimated, then absolute lightness can be computed in many cases by assuming that the lightest surface of the group is white and calculating the absolute lightnesses of all other surfaces relative to that one (Wallach, 1976).

Grossberg and Todorović (1988) computed the relative lightness of two surfaces by first discounting the effects of a spatially non-uniform illumination; see Section 1.5. Discounting the illuminant can be achieved by neurons that obey cell membrane equations and that interact through on-center, off-surround circularly symmetric receptive fields. The present model utilizes such model neurons, which are analogous to those found in the LGN, as summarized in Section 4.1 and defined in Appendix A equations (A1)-(A3). These model neurons are excited by spots of light applied to the center of their receptive fields but are inhibited by those applied outside this central region. The excitatory and inhibitory components of the receptive fields are balanced so that cell responses are attenuated to spatially uniform or slowly varying stimulation. The cells therefore respond preferentially to luminance borders. At a later processing stage, these border signals propagate throughout those surface regions that are completely enclosed by boundaries to complete the lightness representation. Propagation occurs via a filling-in process that is akin to a diffusion process, as defined in Appendix A, equations (A17)-(A23). Propagating signals can dissipate across space unless the region is surrounded by a connected boundary; see Figure 4. As in Grossberg (1994), the present model proposes that the final stage of filling-in occurs in V4, where visible surface percepts are predicted to occur. Section 3 summarizes how such a filling-in process, when confined by the 3D boundaries of the present model, can explain da Vinci stereopsis, as well as many aspects of lightness perception, thereby linking the model's explanations of surface depth and lightness.

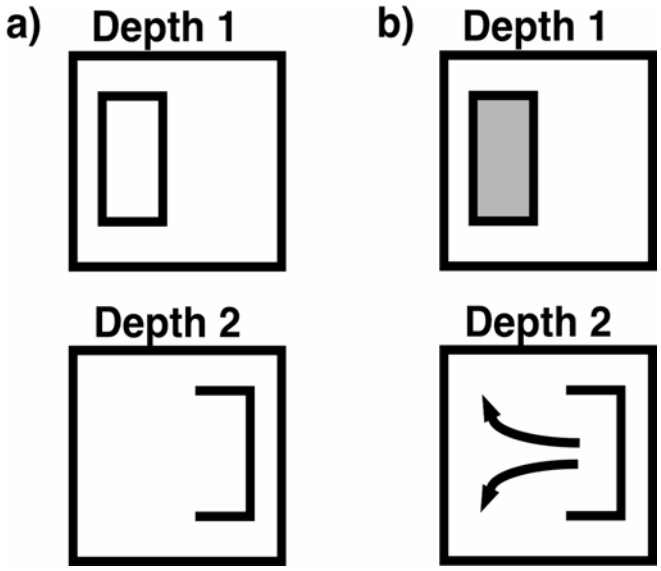

Figure 4: (a) Open and connected boundaries; (b) Filling-in of surface lightness is contained or not depending on the connectedness of the boundary. Note that the monocular boundaries (i.e., two horizontal boundaries and the right vertical boundary) have been added to all depth planes whereas the binocular boundary (i.e., the left vertical boundary) is present only in the near depth plane, thereby creating a connected boundary, and thus containment of filling-in, only in the near depth plane.

3 Model simulations

This section summarizes simulations that predict how monocular and binocular information interact in the visual cortex. We will consider, in turn, contrast variations of dichoptic masking, stereoacuity, Panum's limiting case, contrast variations of the correspondence problem, the Venetian blind illusion, stereopsis with opposite contrast stimuli, da Vinci stereopsis, and the Craik-O'Brian-Cornsweet lightness illusion. The main aim of these simulations is to illustrate how the model's four component networks interact with each other to explain the percepts reported by human subjects. These explanations constitute testable predictions for linking psychophysical percepts to their cortical mechanisms. Like the model diagram shown in Figure 1, the simulation figures should be read from the bottom up, with the bottom two rows representing the input and the V1 boundary representations, the next two rows representing the V2 boundary representations and the top row representing the V4 surface representations. Furthermore for each of the top four rows, depth increases from left to right, with the middle plot representing the fixation plane, the two leftmost plots representing the two near depth planes and the two right plots representing the two far depth planes.

3.1 Dichoptic masking

3.1.1 The basic paradigm

In the basic paradigm considered by McKee et al. (1994), the contrast threshold for the detection of a low contrast bar presented to one eye was found to increase radically when a high contrast bar was presented to the other eye. Furthermore it wasn't necessary for the two bars to be at retinal correspondence. The model explanation of this percept is as follows. The high contrast bar is presented to the left eye and the low contrast bar to the right, as shown by the middle two plots in the bottom row of Figure 5. The outer two plots of the bottom row show the simulated monocular boundary representations. Since their contrasts differ greatly, these two bars cannot be

stereoscopically fused in V1 due to the inhibitory circuit in layer 3B, as explained in Section 2.1. This accounts for the absence of V1 binocular boundaries representations in the second row. As the monocular boundaries do not yet have a depth associated with them, they are added to all depth planes in V2 along their respective monocular lines-of-sight, as shown in the third row of this figure. In this row, each of the 5 plots represent a different depth, with those on the left representing depth planes nearer than the fixation plane and those on the right the converse. As we move across this row the allelotropic shifts (cf., Figure 3) cause the left monocular boundaries to be added to locations further to the right in successive depth planes, while the right monocular boundaries are added to locations further to the left. The left and right monocular boundaries coincide in the near disparity plane represented by the second plot of this row. The vertical boundaries in this disparity plane are consequently stronger than those in the other four depth planes, which they then suppress via the line-of-sight inhibition of the V2 disparity filter (cf., Figure 3) to give the final V2 boundary representations shown in the fourth row. Notice, in particular, that all horizontal boundaries have survived since the disparity filter only inhibits vertical boundaries. In contrast to the horizontal boundaries, only the vertical boundaries in the near disparity plane, represented by the second plot of this row, have survived. As explained in Section 2.4, lightness signals, originating at the location of the boundaries, propagate throughout this disparity plane. Because the near disparity plane contains a connected boundary that completely encloses a bar-shaped region, these boundaries can contain the filling-in of the lightness signals to cause the bar-shaped surface percept shown in middle plot of the top row. The other filling-in signals dissipate and do not give rise to a conscious surface percept; see Figure 4. Because the bars in the left and right eye inputs are perceived to occupy the same position in 3D space, the high contrast bar masks the low contrast bar. In summary, this simulation shows how the left and right inputs can be fused to form a single percept in V4 even though their contrasts are so different that they can not be fused by the binocular cells in V1.

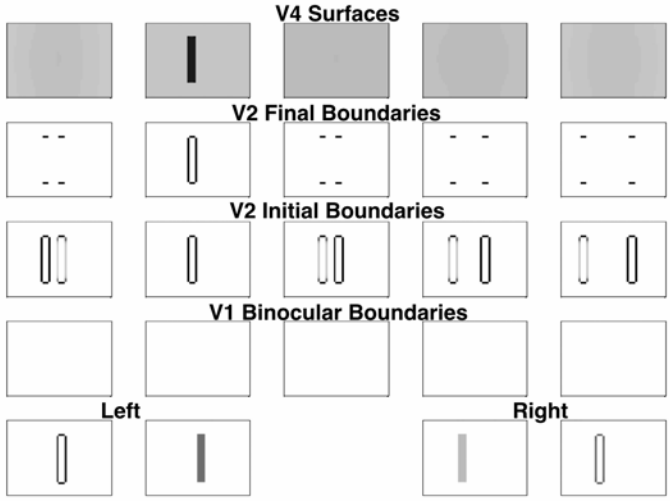

Figure 5: Model simulation showing that retinal correspondence is not needed for dichoptic masking (McKee et al., 1994). The first row represents the inputs and the V1 monocular boundaries, the second row the V1 binocular boundaries, the third and fourth rows the V2 boundaries and the fifth row the V4 surface percepts. In the top four rows depth increases from left to right for successive plots, with the middle plot representing the fixation plane. All other simulation plots use the same format. See text for details.

The fact that the V2 disparity filter can fuse bars whose contrasts are too different to be fused in V1 has ramifications for stereoacuity. In particular Schor & Heckmann (1989) noted that increasing the contrast of the image equally in both eyes increases stereoacuity, but increasing the contrast of the image in just one eye decreases stereoacuity. The model explanation is simply that in the first case fusion could occur in V1 but in the second only in V2. Since V1 cells in general have smaller receptive fields than V2 cells that correspond to the same region of visual space, the model is therefore able to explain why stereoacuity is greater in the first case than in the second.

3.1.2 Release from dichoptic masking

McKee et al. (1994) continued their study of dichoptic masking by demonstrating that, for the particular case where the two bars of Figure 5 were in retinal correspondence, the addition of a second high contrast bar to the right input releases the low contrast bar from masking. The model explanation is summarized in Figure 6a. As before, the stimuli are shown in the middle two plots of the first row and the monocular boundary representations in the outer two plots. Now the high contrast bar of the left input is able to binocularly fuse with the high contrast bar of the right input to form the vertical V1 binocular "far" boundary representations shown in the fourth plot of the second row. As before, the monocular boundaries are added to all disparity planes in V2 along their respective monocular line-of-sights (cf., Figure 3) as shown in the third row. In addition, the V1 binocular boundaries are also added to V2, coinciding with the right bar representation of the fourth plot of this row. The vertical boundaries of this bar representation are therefore stronger then the vertical boundaries in the other depth planes, in particular those of the fixation plane, which they consequently suppress via the line-of-sight inhibition of the V2 disparity filter. This then is the reason why the left bar of the right input is not perceived to lie in the fixation plane even though it is seen only monocularly. The vertical boundaries of the left bar representation of the fourth plot are not suppressed because they do not share any lines-of-sight with the vertical boundaries of the right bar of this plot. The final V2 boundary representations are shown in the fourth row.

Only those vertical boundaries in the fourth plot have survived. Consequently, only the fourth plot contains regions that are completely enclosed by boundaries, and so give rise to surface percepts in V4, as is shown in the fourth plot of the top row. The high contrast bar of the left input is no longer perceived to occupy the same 3D position as the low contrast bar of the right input. The low contrast bar is therefore no longer masked. This simulation explains the observation of McKee et al. (1994) that the addition of a high contrast bar to the right input causes the low contrast bar of the right input to be released from dichoptic masking.

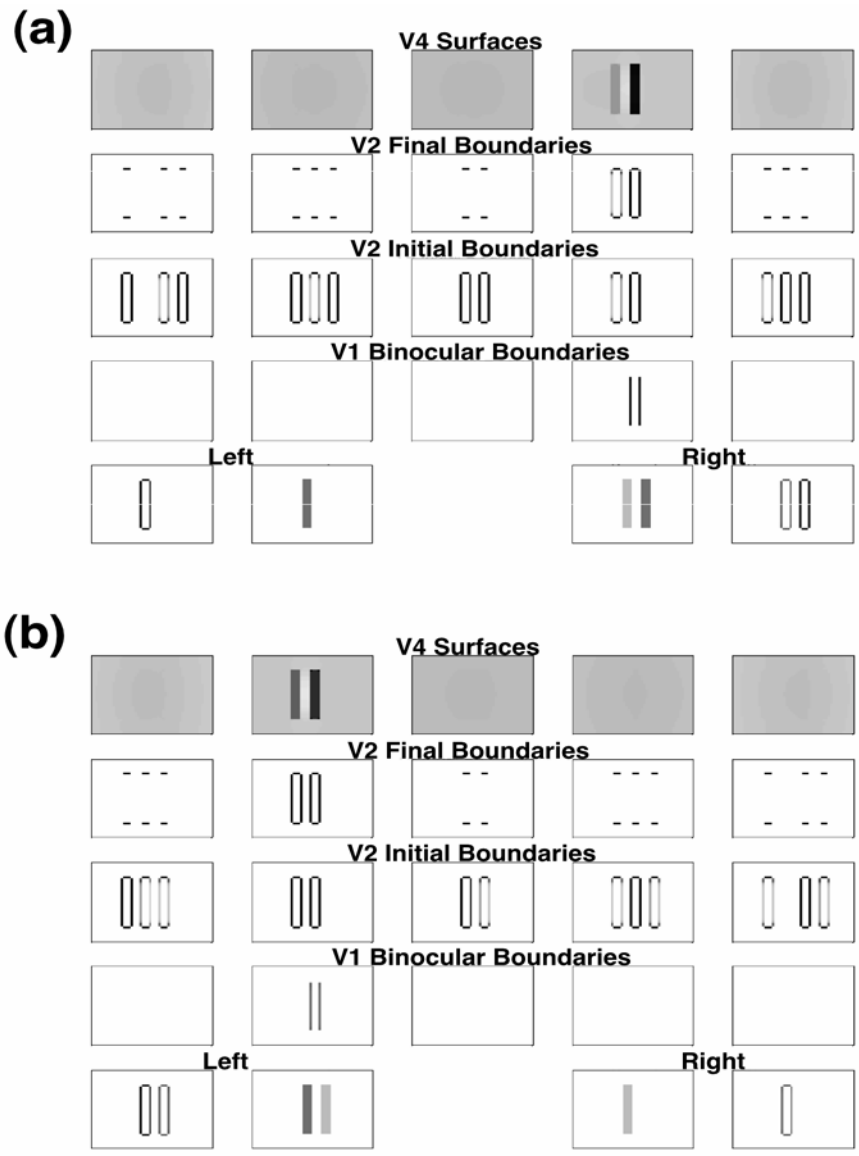

Figure 6: (a) Simulation of the release from dichoptic masking reported by McKee et al. (1995). (b) Simulation of another way a release from dichoptic masking may be achieved. See text for details.

In Figure 6b, instead of placing a high contrast bar next to the low contrast of the right input bar, as in Figure 6a, a low contrast bar is placed next to the high contrast bar of the left input. The resultant simulation is very similar to that shown in Figure 6a. In particular, the two low contrast bars binocularly fuse and the resultant allelotropic shifts mean that once again the high contrast bar of the left input is not perceived to cover the low contrast bar of the right input resulting in a release from dichoptic masking.

Figures 6a and 6b together show that the release from dichoptic masking can be achieved by adding either a high or low contrast bar to the original stimulus of Figure 5. Since McKee et al. (1994) considered only the stimulus configuration of Figure 6a, Figure 6b represents a novel prediction.

3.1.3 Return to dichoptic masking

McKee et al. (1994) also observed that the release from dichoptic masking observed with the stimulus configuration of Figure 6a did not occur when the additional bar in the right input had a low contrast. Figure 7a describes the model simulation of this property. The key difference between Figure 7a and Figure 6a is that the right bar of the right input no longer fuses with the bar of the left input because, as explained in Section 2.1, their contrasts differ too greatly. This accounts for the lack of V1 boundary representations in the fourth plot of the second row. The situation is now very similar to the basic dichoptic masking paradigm, depicted in Figure 5, thereby explaining the return to dichoptic masking.

The monocular boundaries are added to all disparity planes in V2, shown in the third row of this figure, in exactly the same manner as in the previous section. Because the left bar of the right input is in retinal correspondence with the single bar of the left input, their boundary representations overlap in the zero disparity plane, thereby forming the leftmost bar representation in the middle plot of this row. The vertical boundaries of this bar representation are consequently stronger and so suppress all the other vertical boundaries that share either of their lines-of-sight. They do not, however, suppress the vertical boundaries corresponding to the right bar representation of this plot because these boundaries do not share any of their lines-of-sight. The final V2 boundary representations are shown in the fourth row.

Only the vertical boundaries in the fixation plane, represented by the middle plot, have survived. These two sets of boundaries completely enclose two bar-shaped regions. As before, these boundaries confine the lightness signals, which originate at the locations of the boundaries, to generate the two bar-shaped surfaces shown in the middle plot of the top row. The high contrast bar of the left input coincides with, and therefore again masks, the left bar of the right input. In summary, this simulation explains the observation of McKee et al. (1994) that dichoptic masking returns when the contrast of the right bar of the right input is reduced.

A key difference between this simulation and that shown in Figure 6a is that in this simulation the monocularly viewed bars are perceived to lie in the fixation plane where as in the previous simulation the line-of-sight inhibition of the V2 disparity filter prevented the monocularly viewed bar of that simulation from being perceived to lie in the fixation plane. Taken together these two simulations show how the line-of-sight inhibition of the V2 disparity filter can interact with the spatial layout of the stimuli to determine the perceived depth arrangement.

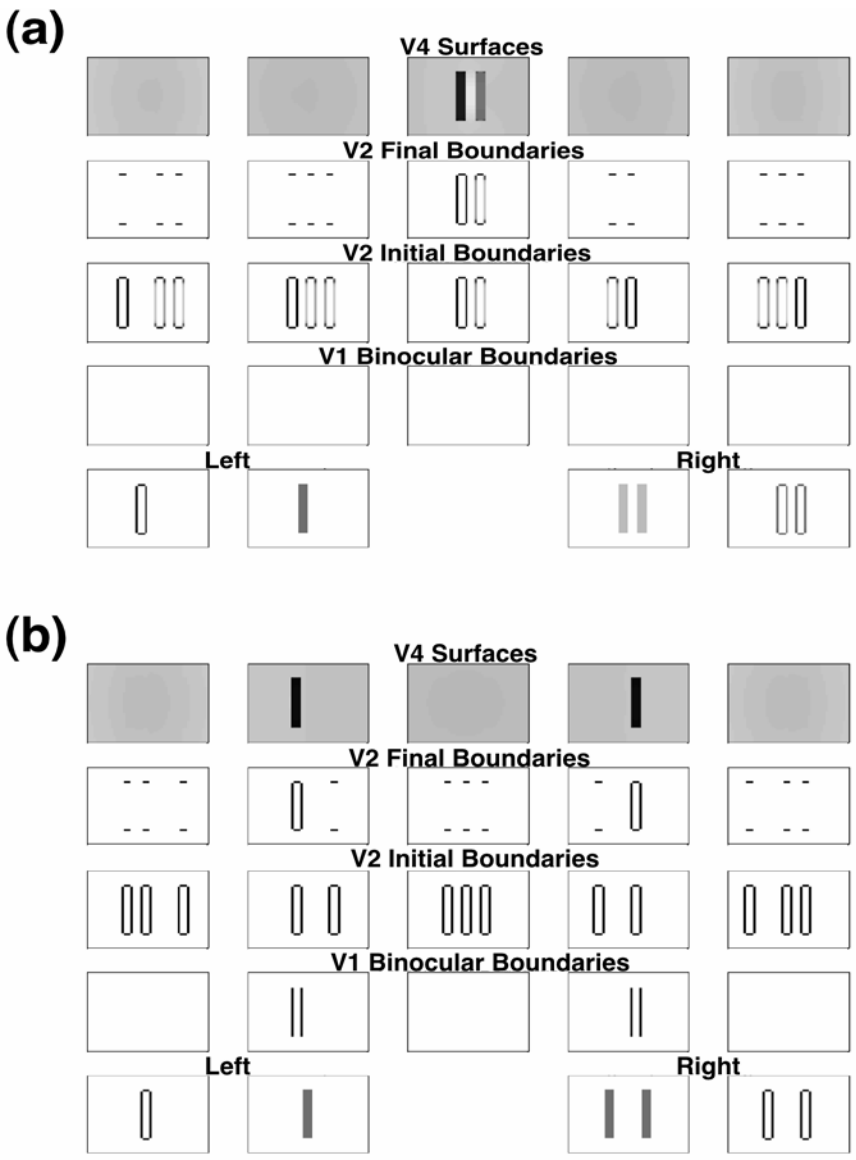

Figure 7: (a) Simulation of the return to dichoptic masking reported by McKee et al. (1995). (b) Simulation of dichoptic masking in Panum's limiting case reported by McKee et al. (1995). See text for details.

3.1.4 Dichoptic masking in Panum's limiting case

As described in the next section, the present model solves the correspondence problem by using a disparity filter that encourages unique matching, via line-of-sight inhibition, but does not enforce it. One advantage of this is that the model can simulate Panum's limiting case, where a bar in one eye is simultaneously fused with two bars in the other eye (Panum 1858, Gillam et al., 1995, McKee et al., 1995; but see Frisby (2001) and Wang et al. (2001) for variations where double matching does not seem to occur). Figure 7b shows the model simulation where a bar in one eye masks equally two bars presented to the other eye as reported by McKee et al. (1995).

The left eye sees a single bar and the right eye sees two bars as shown by the middle two plots of the first row. The resultant monocular boundaries are shown by the outer two plots of this row. Area V1 fuses the bar of the left input with both bars of the right input, to form binocular boundaries

in both a near and a far disparity plane, represented by the second and fourth plots of the second row. The monocular boundaries are added to all disparity planes in V2 along their respective lines-of-sight, as shown by the third row. The left monocular boundaries form the left bar representation in the first two plots, the middle bar representation in the third plot, and the right bar representation in the fourth and fifth plots of the third row. Similarly, the right monocular boundaries form the two right bar representations in the first two plot, the outer two bar representations in the third plot, and the leftmost two bar representations in the fourth and fifth plots of this row. The V1 binocular boundaries are also added to V2, coinciding with the vertical boundaries of the left bar representation in the second plot and the right bar representation in the fourth plot. Those boundaries in V2 that receive binocular input are stronger than and consequently suppress via the recurrent inhibition of the V2 disparity filter those V2 vertical boundaries that receive only monocular input and that share one of their lines-of-sight. The surviving V2 boundary representations are shown in the fourth row. Those regions in V2 that are enclosed by a connected boundary give rise to surface percepts in V4. The model correctly predicts that the bar of the left input is matched with both bars of the right input, and so masks them both equally (McKee et al*.,* 1995).

3.2 Contrast variations of the correspondence problem

The previous simulations have demonstrated the crucial role the disparity filter plays in explaining properties of dichoptic masking. However, as was discussed in Section 1, the disparity filter also helps to solve the correspondence problem by eliminating matches between edges that belong to different objects. The unique-matching rule that various other models have imposed cannot hold in general, since Panum's limiting case shows that, in certain circumstances, a feature in one eye can be matched to two features in the other eye. This is why the disparity filter of the present model encourages unique matching but does not enforce it. The model will now be shown to simulate all the data from the Smallman and McKee (1995) extensive study of the correspondence problem even though it does not enforce the uniqueness constraint. In so doing, it clarifies the crucial role that monocular-binocular interactions play in these percepts.

3.2.1 Control experiment

Smallman and McKee (1995) initiated their study by performing a control experiment in which each eye was presented with two bars, all four bars having the same high contrast. Subjects reported seeing two identical bars, both in the far disparity plane. Figure 8a shows the corresponding model simulation.

Since the left input is displaced leftwards relative to the right input, the vertical edges of the two bars fuse in the far disparity plane in V1, as is shown by the fourth plot of the second row. In addition to this, there is a false match in the near disparity plane of V1, shown in the second plot of this row, which is caused by the inappropriate fusion of the right bar of the left input with the left bar of the right input. As usual, the monocular boundaries are added to all depth planes in the V2 disparity filter along their respective monocular lines-of-sight, as shown in the third row of this figure. In addition, the binocular bar representations are also added to V2, coinciding with the middle bar representation in the second plot and both bar representations of the fourth plot. Those vertical boundaries that receive binocular input, being stronger, quickly inhibit via the V2 disparity filter all other vertical boundaries that share their lines-of-sight and only receive monocular input. The two sets of vertical boundaries in the fourth plot, both of which receive binocular input, cooperate via the disparity filter to inhibit the vertical boundaries of the middle bar representation of the second plot, which also receive binocular input. This happens because the middle bar boundaries receive binocular inputs that share monocular inputs with their inhibitors. The final V2 boundary representations are shown in the fourth row. The model correctly predicts that subjects see both bars in the far disparity plane. In summary, this simulation shows how the line-of-sight inhibition of the

V2 disparity filter ensures that the false match that is present in V1 (second plot of the second row) is eliminated. The V2 disparity filter is therefore the reason why the model can solve the correspondence problem.

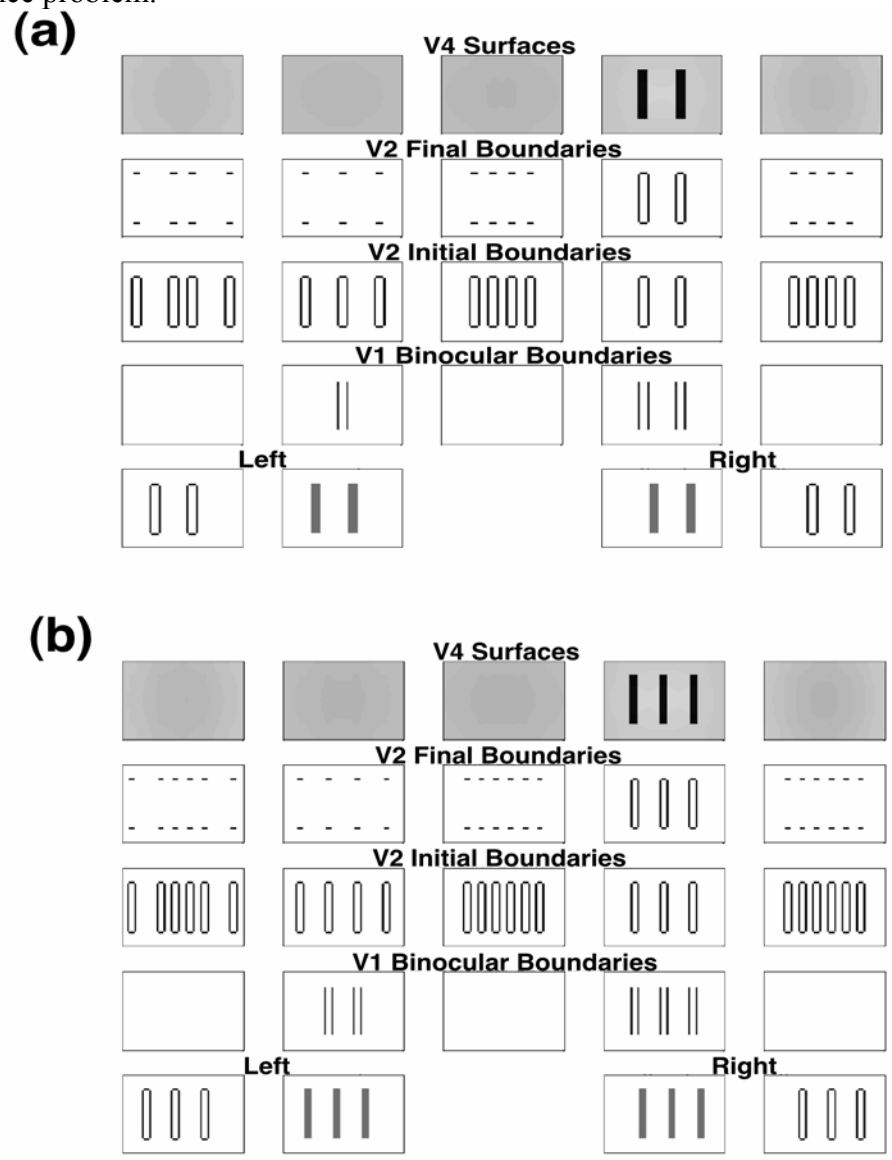

Figure 8: (a) Simulation of the control experiment Smallman and McKee (1995) used for subsequent studies of the correspondence problem. (b) Simulation of a more complicated version of the correspondence problem. See text for details.

Figure 8b shows a more complicated version of the correspondence problem. Once again the false matches are shown in the second plot of the second row and the correct matches in the fourth plot. Since there are more correct matches than false matches, the latter are again suppressed by the former via the line-of-sight inhibition of the V2 disparity filter. This simulation shows that the model can be applied to more general versions of the correspondence problem than that shown in Figure 8a. In Section 3.3 the model is applied to a particularly complex version of the correspondence problem known as the Venetian blind illusion. These simulations of the correspondence problem, the Venetian blind illusion (Figures 11-12) and da Vinci stereopsis (Figures 14-15), among others, clarify how the

model will generalize to natural images by showing how it deals with a variety of potentially confusing matches within the fusion range.

3.2.2 Contrast variations

After performing their control experiment, Smallman and McKee (1995) then proceeded to study contrast variations. They first considered the case where the left bar of the left input had a much lower contrast than the other three bars. They found that observers perceived this bar to lie in the zero disparity plane while also perceiving two high contrast bars, the left lying in a near disparity plane and the right lying in a far disparity plane, as depicted by plots in the top row of Figure 9a. The model simulation explains this percept as follows.

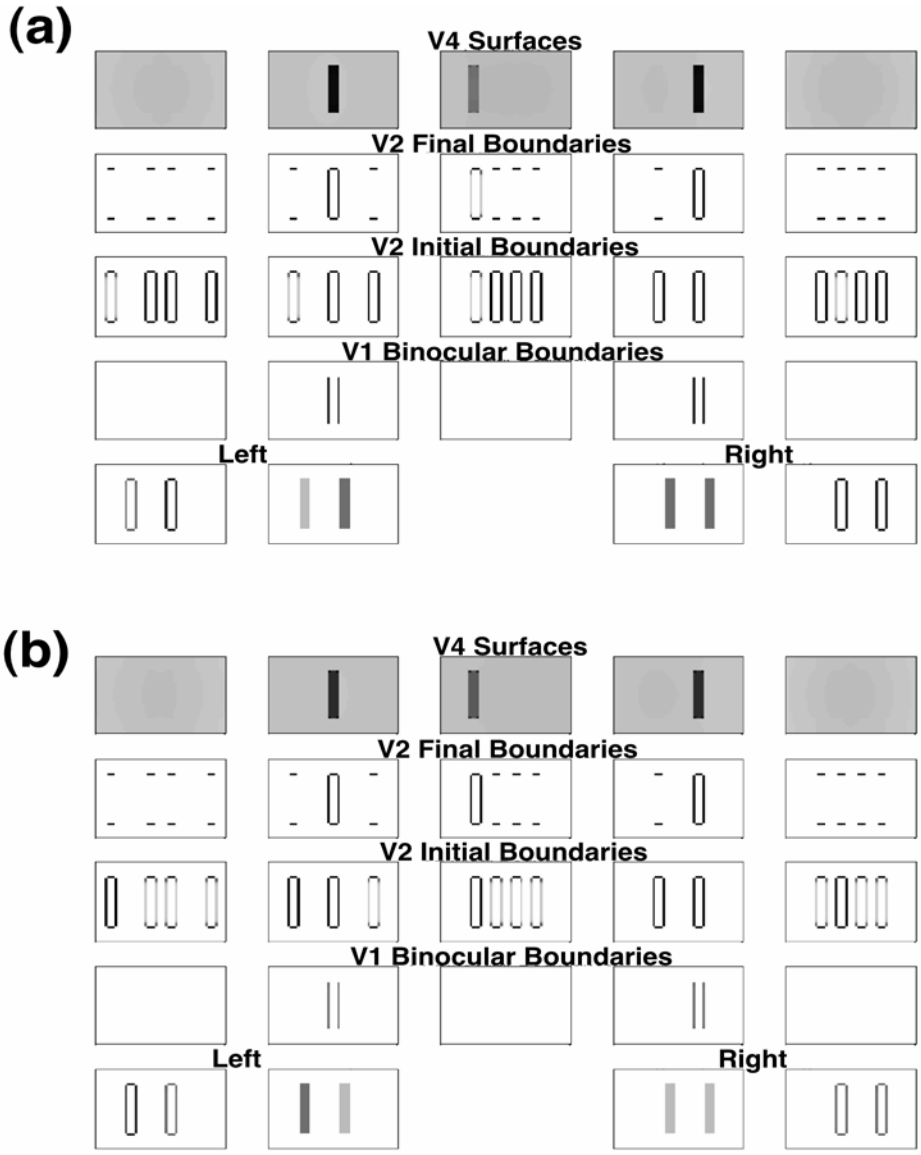

Figure 9: (a) Simulation of a contrast variant of the correspondence problem studied by Smallman and McKee (1995). (b) Simulation of another contrast variant of the correspondence problem studied by Smallman and McKee (1995). See text for details.

The model asserts that the left bar of the left input cannot be matched with either of the bars of the right input because its contrast differs too greatly from theirs. Instead, the right bar of the left input matches with both bars of the right input, forming near and far disparity vertical boundary representations in V1, as shown by the second and fourth plots of the second row. As detailed in Section 3.2.1, the monocular boundary representations are added to all depth planes in the V2 disparity filter along their respective lines-of-sight, as is shown by the plots in the third row of this figure. The V1 binocular boundary representations are also added to the V2 disparity filter, coinciding with the middle bar representation of the second plot and the right bar representation of the fourth plot. These two sets of boundaries, being stronger, then suppress, via the recurrent line-of-sight inhibition of the V2 disparity filter, those vertical boundary representations that share their lines-of-sight. They cannot, however, suppress the vertical boundaries of the leftmost bar representation of the middle plot because these boundaries do not share any of their lines-of-sight. The final V2 boundary representations are shown in the plots of the fourth row. After the surviving connected boundaries fill-in, three surface representations form where humans see them.

Smallman and McKee (1995) then studied the inverse situation where the left bar of the left input had a much higher contrast than the other three bars, which all had the same contrast, as depicted by the middle two plots of the first row of Figure 9b. They found that this situation produced very similar results to the last situation with the left bar of the left input being perceived to lie in the zero disparity plane and the two bars of the right input being perceived to lie in the near and far disparity planes as before.

According to the model, the left bar of the left input once again cannot fuse with either of the bars of the right input, this time because its contrast is too high. The situation is therefore virtually identical to that depicted by Figure 9a, thereby explaining the similar percept reported by the subjects.

3.2.3 The ratio rule

The only difference between the control experiment and these last two experiments was that, in the latter, the left bar of the left input had a significantly different contrast from the other three bars. A key question is how great this contrast difference must be to cause the percept to change from that obtained in the control experiment to that obtained in the last two experiments. Smallman and McKee (1995) determined that this critical difference was best described in terms of a ratio constraint on the magnitudes of the contrast of the inputs to the two eyes, where the exact value of the critical ratio varied between subjects. Figure 10 shows the maximum contrast difference between the two eyes that still allows the model to perform stereoscopic fusion, plotted on logarithmic axes. "○" designates a data point obtained when the odd bar had a lower contrast than the other three bars (i.e., the situation depicted by Figure 9a) and "+" the converse situation (Figure 9b). The line of best fit for the two sets of data combined is a straight line of slope equal to 1, indicating that the model's behavior is consistent with the ratio rule reported by Smallman and McKee (1995).

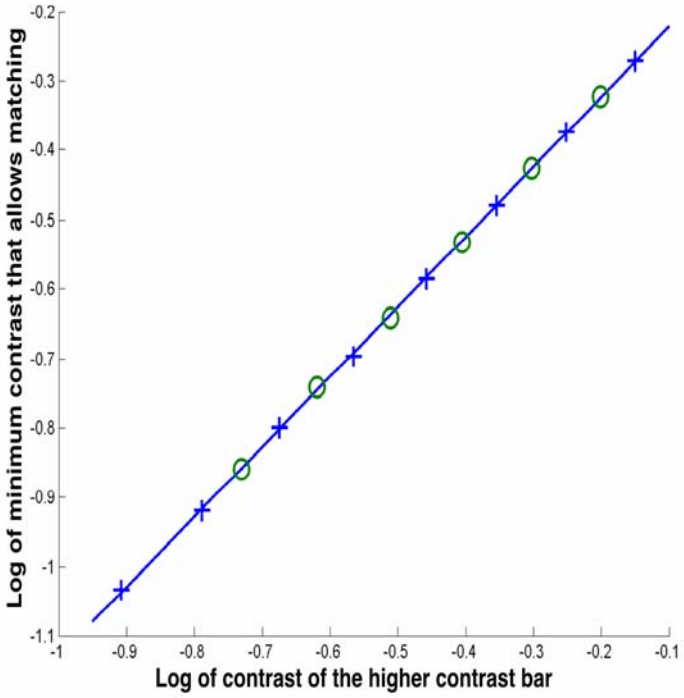

Figure 10: Simulation of the ratio constraint on stereoscopic fusion. On logarithmic axes the contrast of the higher contrast bar is plotted against the minimum contrast of the low contrast bar that can still be fused with it. Crosses (+) denote the situation where one bar has a higher contrast than the other three bars (e.g., Figure 9a) and circles (○) the converse (e.g., Figure 9b). The line of best fit for the total data set has a slope of 1. The model thus obeys the ratio constraint on stereoscopic fusion.

3.2.4 Exception to the ratio rule

Smallman and McKee (1995) also discovered one notable exception to the ratio rule: When each eye sees only a single bar, these bars match regardless of their contrast difference. This situation was simulated as an example of dichoptic masking without retinal correspondence, and the reader is referred back to Section 3.1.1 (Figure 5) for an explanation.

3.3 The Venetian blind effect

We continue our discussion of the correspondence problem by considering the Venetian blind effect. A Venetian blind stereogram is shown in Figure 7.21 of Howard and Rogers (1995) and consists of two gratings, a low frequency one that is presented to the left eye, and a high frequency one presented to the right. When fused, the frequency of the gratings are such that every second bar of the left grating is in retinal correspondence with every third bar of the right grating. The stimulus is reproduced in the middle two plots of the first row of Figure 11.

According to Howard and Rogers, this stereogram produces a percept of short ramps, each containing three bars, sloping up from left to right interspaced with steep returns. The total percept is that of a Venetian blind. The model is able to correctly predict this percept, as shown by the top row of Figure 11. Numbering from left to right, this row shows that the first bar of the percept is in the zero disparity plane, the second in the near disparity plane, then there is a step return to the third bar which is located in the far disparity plane after which the pattern repeats.

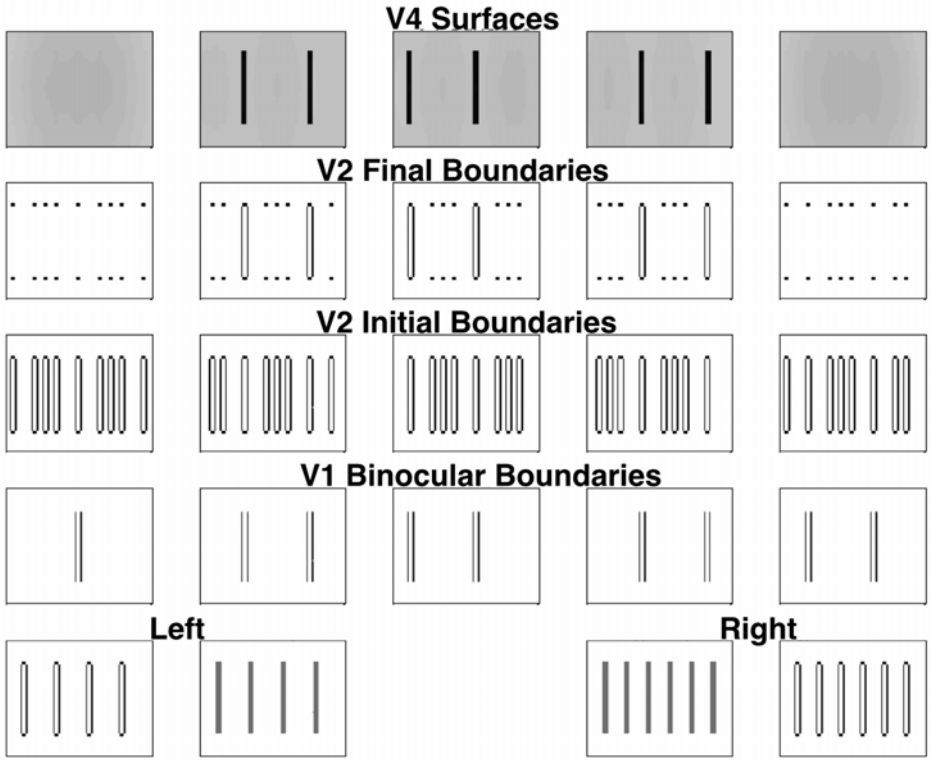

Figure 11: Simulation of the Venetian blind effect (Howard & Rogers, 1995).

Although the model is able to correctly simulate the percept, this simulation is too complicated to explain simply. Instead, we will divide the stimulus into two components and consider these separately.

First, we note that every second bar of the left input is in retinal correspondence with every third bar of the right input. We extract these bars to form the stimulus shown in the middle two plots of the first row of Figure 12a. Since the bars in the left and right inputs of this figure are in retinal correspondence, the model correctly predicts that they will appear in the zero disparity plane, as shown by the middle plot of the fifth row.

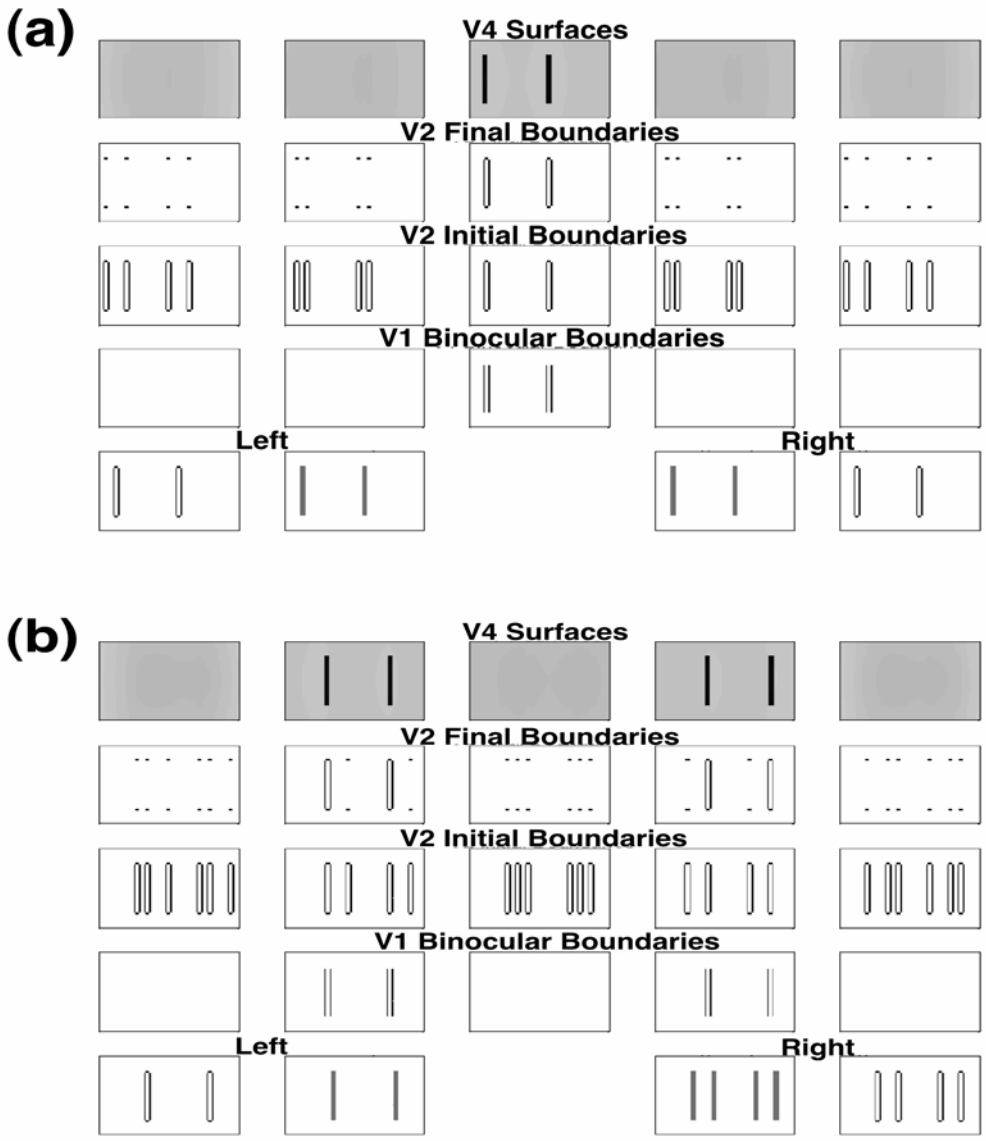

Figure 12: (a) Simulation of one component of the Venetian blind effect. (b) Simulation of the other component.

We now consider the remaining bars, which are shown in Figure 12b. The right eye sees exactly twice the number of bars as the left eye. This is therefore an example of Panum's limiting case that was considered Section 3.1.4. As before, the model predicts that each bar of the left input is fused with two bars of the right input to generated the percept shown in the top row of this figure.

Adding together the percepts shown in top rows of Figure 12a and 12b, we achieve the percept shown in the top row of Figure 11, thereby explaining the Venetian blind effect. The insight that the model provides is that the Venetian blind effect is just a complex version of the correspondence problem and Panum's limiting case, when it is properly understood by combining early stereo matching, later selection by a disparity filter, and surface filling-in of those regions that are completely enclosed by boundaries.

3.4 Stereopsis with opposite contrast stimuli

Polarity-reversed stereograms are those stereograms where corresponding elements in the two stereo half images have opposite luminances. In other words, for every white element in one stereo half image there is a corresponding black element in the other stereo half image. The model simulation of the depth percept induced by a polarity-reversed stereogram is shown in Figure 13a.

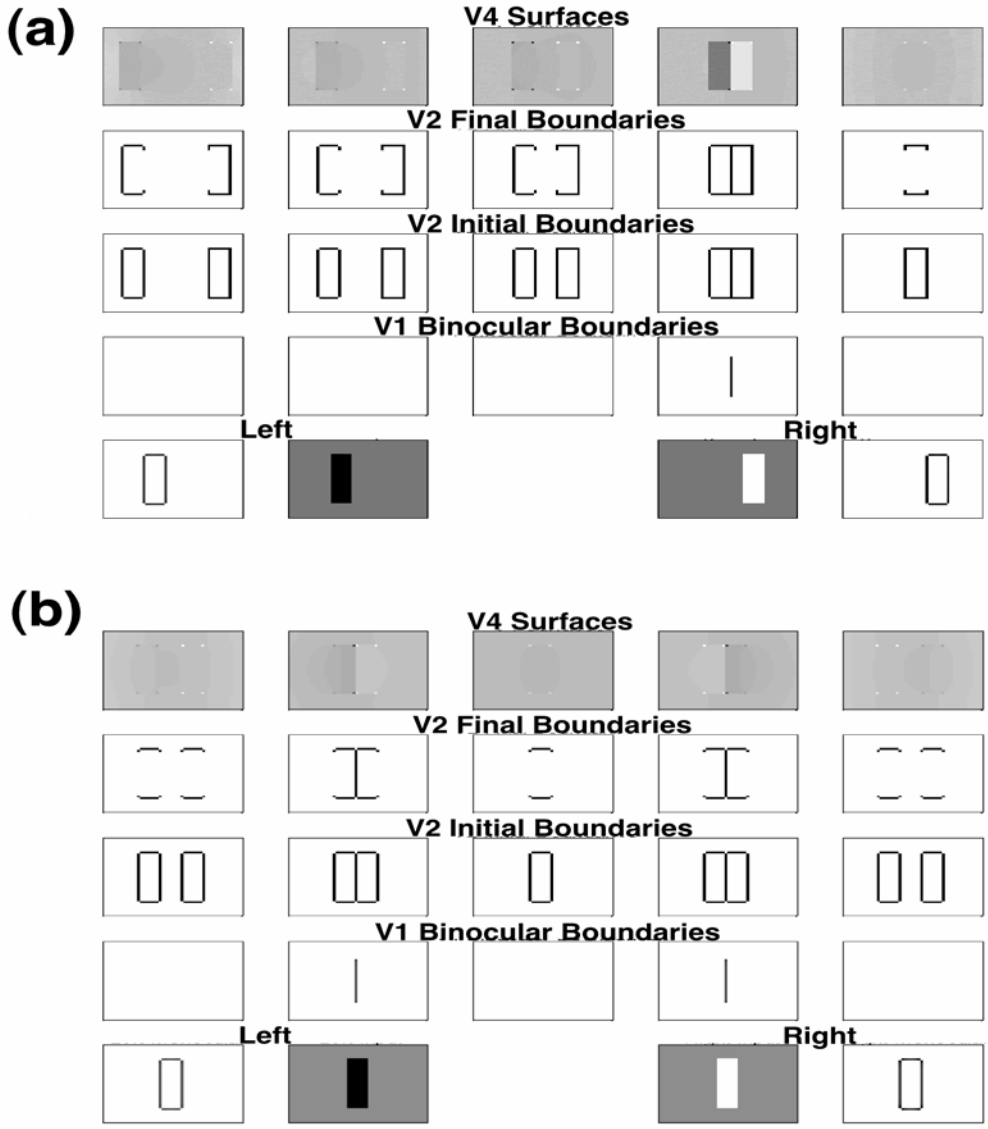

Figure 13: Simulation of stereopsis with a polarity-reversed stereogram. See text for details.

Here the left eye sees a black bar and the right eye a white bar both on the same gray background. Since they have the same contrast polarity, the left edge of the white bar fuses with the right edge of the black bar to form a vertical boundary in a far disparity plane of V1, shown by the fourth plot of the second row. Even though they have the same contrast-polarity, the right edge of the white bar cannot fuse with the left edge of the black bar because they are too disparate. As always, the monocular boundaries are added to V2 along their respective lines-of-sight, shown by the plots in the third row. The vertical binocular V1 boundary is also added to V2 and coincides with the middle vertical boundary in the far disparity plane represented by the fourth plot. This boundary, being

stronger, then suppress, via the line-of-sight inhibition of the V2 disparity filter, the middle two vertical boundaries in all other disparity planes of this row, resulting in the final V2 boundary representations shown in plots of the fourth row. Only the fourth plot of this row contains regions completely enclosed by boundaries, which is why surfaces are perceived only in this disparity plane, as shown by the fifth row.

This simulation suggests that the stereoscopic depth perception induced by polarity-reversed stereograms is mediated by the fusion of those edges in the two stereo half images that have the same contrast polarity. In particular, the model predicts that the degree of the induced depth should be completely determined by the disparity of these fused edges. An alternative prediction would be that, since most of the display is perceived only monocularly, and since monocular objects tend to be perceived to lie in the fixation plane (Krol & van de Grind 1983), the depth perceived should be biased towards the fixation plane. Howe and Watanabe (2002) ran a series of psychophysical experiments to investigate which of these two predictions were true, and found that the prediction of the model was a better description of the degree of depth experienced by subjects.

Suppose, however, that the subject's vergence were to change so that the stimulus became that shown in Figure 13b. Now, because they have the same contrast polarity, the right edge of the white bar fuses with the left edge of the black bar to form a boundary representation in the far disparity plane of V1; see the fourth plot of the second row. Also, the left edge of the white bar fuses with the right edge of the black bar to form a boundary representation in the near disparity plane of V1; see the second plot of the second row. Unlike Figure 13a, there are now two boundary representations in V1. The monocular boundaries are added to V2 along their respective lines-of-sight. The binocular boundaries are also added to V2, overlapping with the middle vertical boundaries in the second and fourth plots of the third row. These two boundaries, being stronger, suppress all other vertical boundaries via the recurrent inhibition of the V2 disparity filter. However, because they are equally strong, they can not suppress each other. The final boundary representations are shown in fourth row. No regions are completely enclosed by boundaries and so the model predicts that there will be no stable depth percepts.

This prediction is correct in as far as it goes, in that subject do not achieve any stable surface percepts, but in practice unstable surface percepts may form if subjects experience binocular rivalry. Describing binocular rivalry is beyond the scope of our simulations. However, it has been qualitatively modeled in Grossberg (1987) in a manner that is consistent with the present model simulations.

The key point here is that whether or not an anticorrelated stereogram induces a stable depth percept depends on the vergence of the subject. One vergence position enables the visual system to match the left and right inputs only in a single way. Other vergence positions lead to two binocular boundaries in V1, and consequently no stable depth percepts in V4, as demonstrated by Figure 13b. Subjects may also be able to use attention to choose between the two possible ways of matching the left and right inputs. Section 4.4 shows how the model may be extended to incorporate attentional effects.

Regardless of whether subjects use vergence or attention to make sure their visual system can only fuse the left and right inputs in one way, as more elements are included in the left and right inputs, the harder it is to ensure unambiguous fusion. The model suggests that this is the reason why complex anticorrelated stereograms (i.e., those anticorrelated stereograms that contain many separate elements) induce little or no depth perception whereas simple anticorrelated stereograms do (Julesz, 1971; Howard & Rogers, 1995).

3.5 Da Vinci stereopsis

Da Vinci stereopsis describes those situations where a monocular object has a definite depth conferred to it by its relationship to a binocularly-viewed object. Such situations are often caused by

each eye viewing the world from a slightly different position, leading to partial occlusions where part of a scene is visible to only one eye. The model clarifies how the percept of depth caused by such stimuli can be explained in terms of monocular-binocular interactions.

3.5.1 Stimuli of Nakayama and Shimojo (1990)

In this set of experiments, a thick bar was presented to both eyes and a thin bar only to the right eye, as shown in the first row of plots of Figure 14a. Subjects reported perceiving the thin bar behind the thick bar, at a depth that was consistent with the right edge of the thin bar of the right input being fused with the right edge of thick bar of the left input.

The model explanation is as follows. The vertical boundaries of the thick bar are registered binocularly in the near disparity plane in V1, as shown by the second plot of the second row, and the right edge of the thin bar is matched with the right edge of the thick bar to be registered binocularly in the far disparity plane in V1, as shown by the fourth plot. The left edge of the thin bar is registered only monocularly because it cannot be matched with either of the edges of the left input. As usual, the monocular boundaries are added to all depth planes in the V2 disparity filter along their respective monocular lines-of-sight as shown by the plots in the third row of this figure. The vertical binocular boundaries are also added to the disparity filter, overlapping with the vertical boundaries of the thick bar representation in the second plot and with the rightmost vertical boundary in the fourth plot. These vertical boundaries, being stronger, eliminate all other vertical boundaries that share their lines-of-sight via the disparity filter's line-of-sight inhibition. However, they do not eliminate the vertical boundaries originating from the left edge of the thin bar because these do not share any of their lines-of-sight. The final V2 boundary representations are shown in the fourth row. As usual, V4 fills in surfaces in those regions that are completely enclosed by a connected boundary. This produces a percept of a thick bar in a near disparity plane, represented by the second plot of the top row, and a thin bar in a far disparity plane, represented by the fourth plot. The very small squares seen in the top row are artifacts of the implementation of the diffusion process with a relatively small number of pixels and have no physiological significance. In particular, they disappear when the simulations are carried out at a sufficiently high resolution, at a high computational cost. The model therefore correctly predicts that the thin bar will appear behind the thick bar at a depth that is consistent with the right edge of the thin bar being stereoscopically fused with the right edge of the thick bar, as has been reported experimentally (Nakayama & Shimojo, 1990).

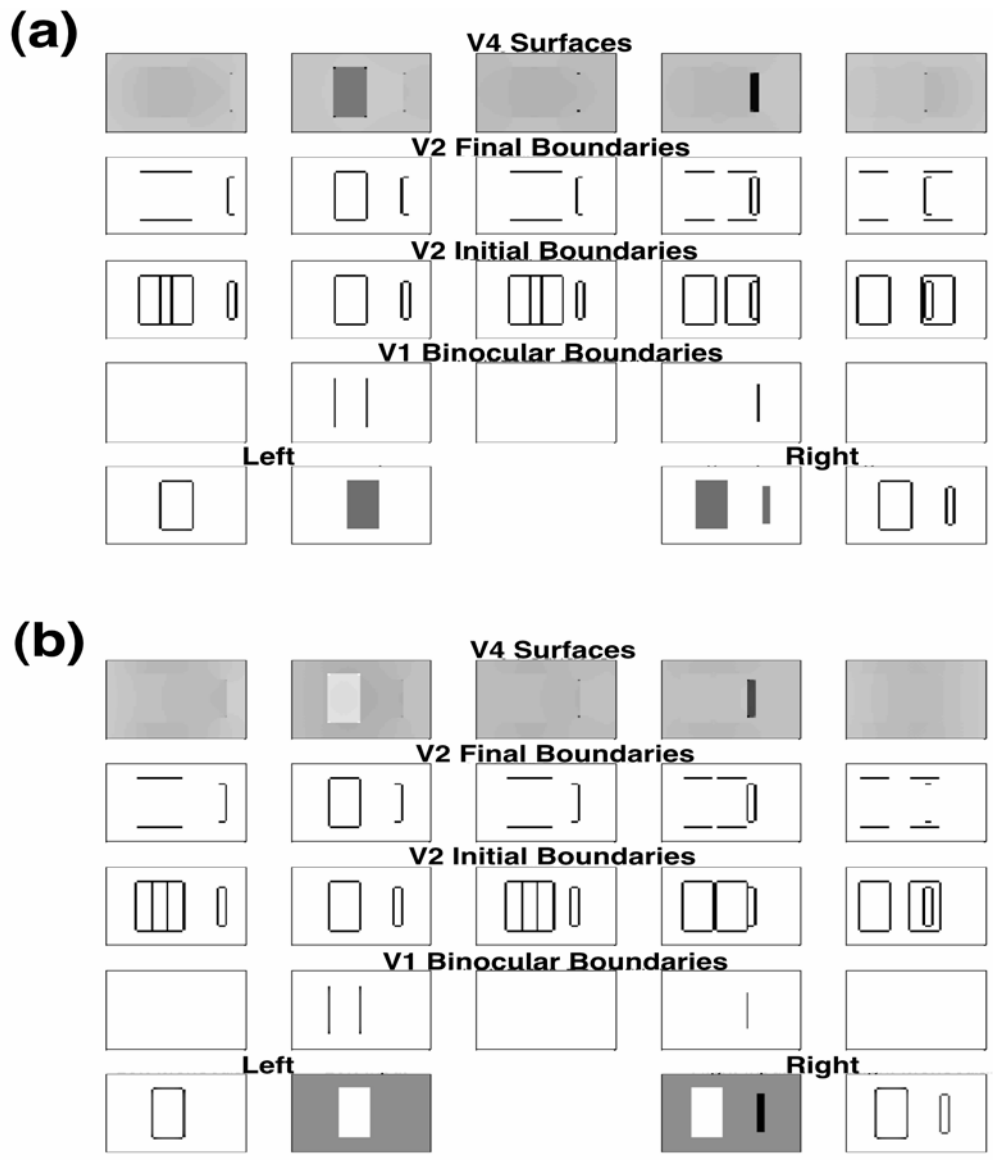

Figure 14: (a) Simulation of the depth percept invoked by the conventional da Vinci stereopsis stimuli of Nakayama and Shimojo (1990). (b) Simulation of the depth percept invoked by the polarity-reversed da Vinci stereopsis stimuli. See text for details.

Nakayama and Shimojo (1990) also showed that the percept remained the same even when the monocular and binocular bars had opposite luminance polarities: specifically, the binocular bar being white and the monocular bar being black. It might be thought that the model could not simulate this observation, since the model can only binocularly fuse edges that have the same contrast polarity. The model's successful simulation is shown in Figure 14b.

The only difference between this simulation and the previous simulation is that in the previous simulation it was the right edge of the thin bar of the right input that fused with the right edge of the thick bar of the left input whereas in this simulation it is the left edge of the thin bar that fuses with the right edge of the thick bar of the right input since these two edges now have the same contrast polarity. This simulation then proceeds in the same manner as the previous one, thereby explaining the similar percept. Contrary to the claims of Nakayama and Shimojo (1990), this simulation shows

that the reported percept can be generated without independent knowledge of occlusion relationships; see section 4.2.

3.5.2 Stimulus of Gillam, Blackburn and Nakayama (1999)

In Figure 15a the right eye sees two thin bars and the left eye a single thick bar. Subjects report seeing two thin bars, the left in the near disparity plane and other in the far disparity plane. Gillam et al. suggested that, because the right eye input contains a gap not present in the left eye input, this display demonstrates that stereopsis can be induced by monocular gaps. It should be stressed that, although they are superficially very similar, this display is quite different to that of Nakayama and Shimojo (1990) (Figure 14) which demonstrated an entirely different point: that depth perception could be determined by the separation of a monocular bar from a binocular bar.

The model explanation is as follows. The model suggests that the left edge of the thick bar fuses with the left edge of the left thin bar to appear in a near disparity plane in V1, represented by the second plot of the second row, while the right edge of the thick bar fuses with the right edge of the right thin bar to appear in a far disparity plane in V1, represented by the fourth plot of this row, since in both cases these edges have the same contrast polarity. The two other vertical edges of the thin bars of the right input are registered only monocularly because they cannot be matched to either of the edges of the left input. As usual, the V1 monocular boundary representations are added to all depth planes in the V2 disparity filter along their respective lines-of-sight. This is why two thin bar representations and one thick bar representation are seen in all disparity planes of the third row, with the slight complication that in all cases the thick bar representation overlaps with at least one of the two thin bar representations. The V1 binocular boundary representations are also added to the V2 disparity filter, overlapping with the leftmost vertical boundary in the second plot and the rightmost vertical boundary in the fourth plot. These vertical boundaries, being stronger, inhibit, via the recurrent line-of-sight inhibition of the disparity filter, all the other vertical boundaries that share any of their lines-of-sight. This means that they do not inhibit those vertical boundary representations originating from the two monocularly viewed edges of the right input because these vertical boundaries do not share any of their lines-of-sight. The final V2 boundary representations are shown in the fourth row. V4 fills-in surfaces in those regions that are completely enclosed by boundaries, resulting in the percept of a thin near bar and a thin far bar, as reported by human subjects (Gillam et al., 1999).

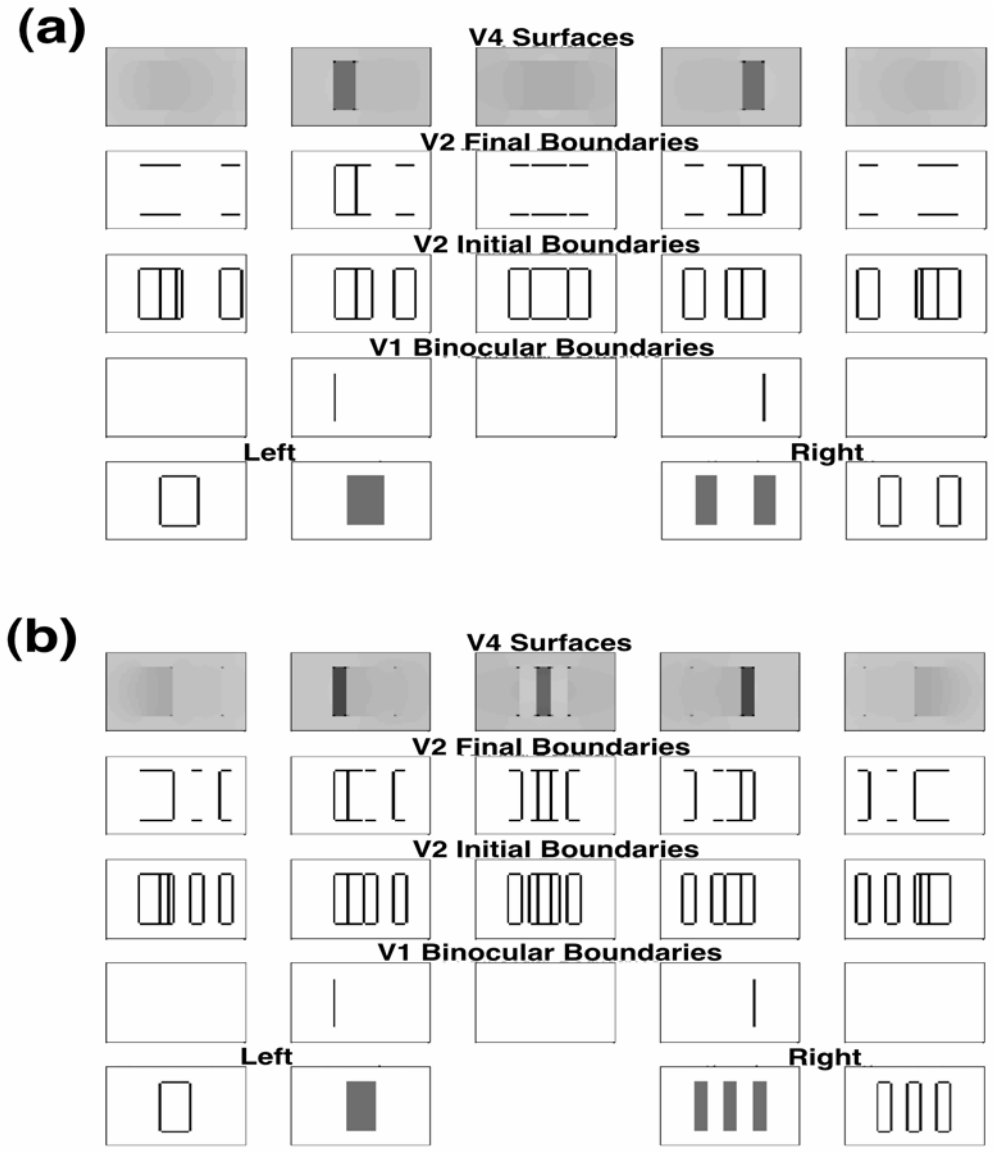

Figure 15: (a) Simulation of depth percept invoked by the da Vinci stereopsis stimuli of Gillam, Blackburn and Nakayama (1999). (b) Simulation of a variant of the original Gillam et al. stimuli.

In the previous display, at least one edge of each region could be binocularly fused. In contrast, in Figure 15b the middle bar of the right eye stimulus is perceived entirely monocularly.

The model simulation is as follows. The left eye sees a single bar while the right eye sees three separate bars. The left edge of the bar of the left input again fuses with the left edge of the leftmost bar of the right input to form a binocular boundary in the second plot of the second row. Similarly, the right edge of the bar of the left input again fuses with the right edge of the rightmost bar of the right input to form a binocular boundary in the fourth plot of the second row. Again the monocular boundaries are added to V2 along their respective lines-of-sight, as shown by the third row. The binocular V1 boundaries are also added to V2. The binocular boundary in the second plot of the second row overlaps with the first vertical boundary in the second plot of the third row. Similarly, the binocular boundary of the fourth plot of the second row overlaps with the last vertical boundary in the fourth plot of the third row. The surviving V2 boundaries are shown in the fourth row.

Only those boundaries that completely enclose a region can contain the lightness signals that originate at the location of the boundaries, and so only these regions give rise to surface percepts in V4. The model therefore correctly predicts that three surfaces will be seen, each at a different depth as reported experimentally (Gillam et al., 1999).

3.6 Lightness illusions

The filling-in mechanism utilized by the model V4 simulations is equivalent to that used by Grossberg and Todorović (1988) to explain several lightness illusions. It is therefore claimed that the present model can explain the same large set of lightness illusions. (See Grossberg and Kelly (1999), Grossberg and Pessoa (1998), Kelly and Grossberg (2000) and Pessoa, Mingolla, and Neumann (1995) for other articles that explain additional lightness and brightness data using this filling-in mechanism.)

The Craik-O'Brian-Cornsweet effect (COCE) is simulated to illustrate this claim. In Grossberg and Todorović (1988), the COCE was simulated using only a monocular input. The simulation herein uses inputs to both eyes and shows that the binocular model can also simulate this percept. The stimuli are shown in the middle two plots of the bottom row of Figure 16. Both eyes see the same stimulus, which consists of two abutting regions of the same uniform lightness separated by a lightness cusp. Subjects report perceiving both regions as having uniform lightness, with the left region appearing darker than the right.

The model explains the COCE as follows. The input is binocularly fused to form three vertical binocular boundaries in the nearest disparity plane of V1, represented by the leftmost plot of the second row. As always, both the V1 binocular and monocular boundaries are added to the V2 disparity filter, with the monocular boundaries being added to all depth planes along their respective lines-of-sight, as shown by the plots in the third row. The vertical boundaries in the nearest disparity plane are stronger because they receive both monocular and binocular input. They therefore inhibit the vertical boundaries in the other disparity planes via the recurrent line-of-sight inhibition of the disparity filter. The final V2 boundaries are shown by the plots in the fourth row. The boundaries in the nearest disparity plane confine the V4 diffusion of the lightness signals that originate at the edges of the regions. Those lightness signals originating from the left side of the cusp are darker than those originating from the right side. This lightness difference is propagated, by the V4 filling-in mechanism, throughout the respective regions, causing the left region to appear uniformly darker than the right, as shown by the leftmost plot of the top row.

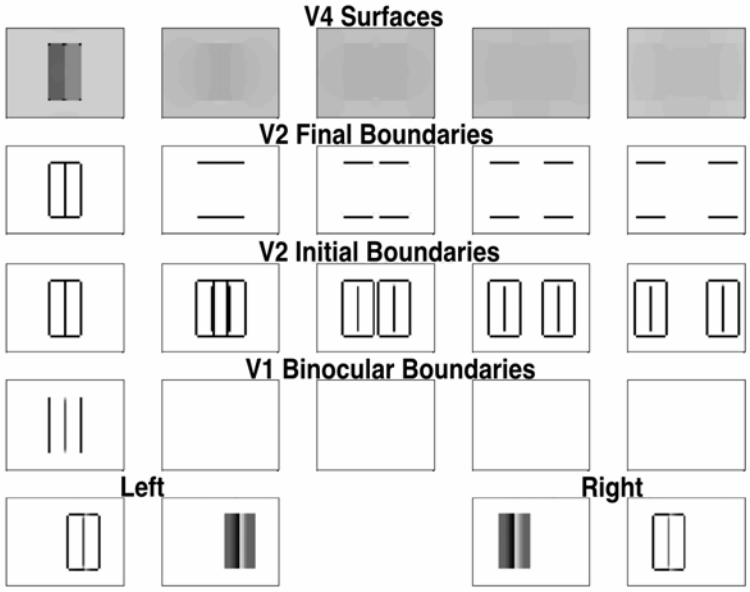

Figure 16: Simulation of the Craik-O'Brian-Cornsweet lightness illusion. See text for more details.

4 Discussion

4.1 Supporting physiological and anatomical data

This section shows that all the relevant physiological and anatomical data of which we are aware support the model. The model does not, however, consider cortical areas V3, V3A and MT, even though there is evidence that these areas play a role in depth perception (e.g. Backus et al., 2001). These areas were not needed to simulate the model's targeted data. The function of area V3A appears to be particularly controversial with studies suggesting that it is variously concerned with relative disparity (Backus et al., 2001), saccades (Nakamura & Colby, 2000a, 2000b) and prehensile hand movements (Nakamura et al, 2001). As a further complication, there is some evidence that the function of macaque V3A differs from that performed by human V3A (Tootell et al., 1997).

When the model diagram in Figure 1 is compared to the list of data below, it can be seen that the model makes predictions concerning brain physiology and anatomy beyond what is known. One prediction is that there is an inhibitory circuit in V1 which causes the binocular cells in layers 3B and 2/3A not to respond if the inputs to the left and right eyes differ too greatly in contrast. Another is that there is a disparity filter in V2 that employs line-of-sight inhibition. A third prediction is that there is a surface filling-in mechanism that leads to visible percepts and is located in V4 (among other places; see Grossberg, 1994). This section should be read in conjunction with Figure 1, which interprets each model stage anatomically.

4.1.1 V1 binocular boundaries

Consistent with the model, the LGN contains circularly symmetric on-center, off-surround receptive fields (Kandel, Schwartz, & Jessell, 2000, pp. 529). LGN lesion studies have shown that the parvocellular, but not the magnocellular, pathway is critical for fine stereopsis (Schiller, Logothetis, & Charles, 1990a, 1990b). Just as V1 layer 4 is the major recipient of this parvocellular input in vivo (Callaway, 1998), it is also the input layer of model V1. Also, in accord with the model, layer 4 is known to output to layer 3B, but not to layer 2/3A, of V1 (Callaway, 1998), a large proportion of it is

monocular (Hubel & Wiesel, 1968; Poggio, 1972), and many of its cells are simple (Hubel & Wiesel, 1968; Schiller, Finlay, & Volman, 1976).

As discussed in Section 2.1, the model assumes that polarity-specific binocular matching occurs in layer 3B. This is consistent with observations that a significant proportion of layer 3B comprises simple cells (Dow, 1974), that layer 3 contains a significant number of binocular cells (Hubel & Wiesel, 1968; Poggio, 1972), and that projections to it can be independent of ocular dominance (Katz, Gilbert, & Wiesel, 1989).

The model suggests that binocular layer 2/3A cells pool responses from layer 3B cells of both contrast polarities so that they can represent the boundaries of objects whose contrast polarity, with respect to the background, changes as the boundary is transversed. In keeping with this suggestion, it is known that layer 3B projects throughout layer 2/3A (Callaway, 1998), and that layers 2 and 3 each contain significant numbers of binocular and complex cells (Poggio, 1972).

The model further suggests that there is a group of cells in layer 2/3A and 3B that respond only to binocular, and not to monocular, stimulation. Such "obligate cells" are known to exist in macaque V1 (Poggio & Fischer, 1977; Smith, Chino, Ni, & Cheng, 1997), with about 40% of tuned excitatory neurons being obligatory (Poggio & Talbot, 1981), including almost all "tuned zero" neurons (Poggio, 1991). Obligate cells do not appear to be as prevalent in cat (Anzai et al., 1995).

The model predicts that all these interactions occur in the V1 interblob regions, which is in keeping with observations that V1 interblobs are highly selective for orientation but relatively unselective for color (Merigan & Maunsell, 1993).

4.1.2 V1 monocular boundaries

The model suggests that the V1 monocular boundaries are formed by a process that is a simplification of that which forms the V1 binocular boundaries. Consequently, much of the above data applies equally to the monocular boundaries network. Additional support for this network comes from observations that layer 3 (Hubel & Wiesel, 1968; Poggio, 1972) and layer 2 (Poggio, 1972) of V1 each comprise a large proportion of monocular cells.

4.1.3 V2 boundaries

The model assumes that the V2 boundaries are located in the V2 pale stripes. This is consistent with observations that the V2 pale stripes receive the major projection from the V1 interblob regions, while receiving no significant projection from the V1 blob regions, and are highly orientationally selective (Roe & Ts'o, 1997), while also containing a complete map of visual space (Roe & Ts'o, 1995).

The model is further consistent with data that V2 is mainly binocular (Hubel & Livingstone, 1987; Roe & Ts'o, 1997), is mainly disparity-sensitive (Poggio & Fischer, 1977; von der Heydt, Zhou, & Friedman, 2000), contains many complex cells (Hubel & Livingstone, 1987), receives input into layer 4 (Rockland & Virga, 1990) and outputs to V4 (Xiao, Zych, & Felleman, 1999), which itself is highly selective for disparity (Merigan & Maunsell, 1993). In addition, the V2 pale stripes are disparity-selective (Peterhans, 1997).

According to the model, an important function of V2 is to suppress false matches by utilizing a disparity filter. This is consistent with observations that cells readily exhibit false matches in V1 (Cumming & Parker, 2000), but not in V2 (Bakin et al., 2000).

4.1.4 Surfaces

Surfaces are built up through interactions between the V1 blobs, the V2 thin stripes, and V4, consistent with the fact all these regions are linked by major projections (Livingstone & Hubel, 1984; Xiao et al., 1999), that the V2 thin stripes are the least orientationally-selective area of V2 (Peterhans, 1997) and contain a complete map of visual space (Roe & Ts'o, 1995).

4.2 Comparison with other theories and models

One of the most popular explanations of monocular-binocular interactions is the ecological optics hypothesis of Nakayama and Shimojo (1990). This hypothesis suggests that visual systems attempt to interpret unpaired image points in terms of occlusion. For example, in Figure 14, both eyes see a thick bar but only the right eye a thin bar. According to the ecological optics hypothesis, the visual system interprets these stimuli by assuming that the thin bar is located behind the thick bar at the exact distance that would cause the thick bar to hide it from the left, but not from the right, eye.

While this hypothesis is consistent with the percepts evoked by the stimuli in Figures 14 and 15, it cannot explain the percept evoked by the stimuli of Figure 13, because this stimulus cannot be explained in terms of occlusion. If we wish to understand the response of the visual system to all possible stimuli, not just the ones that can be interpreted in terms of occlusion, then it is necessary to offer a mechanistic account that can deal with a broader data set in a unified way, as the present model does.

One of the most successful mechanistic models of stereopsis is the disparity energy model (Ohzawa et al., 1990). However, this model does not solve the correspondence problem in that it may match vertical contours in the two retinal images that correspond to different objects. Fleet et al. (1996) have proposed how the disparity energy model could be extended to avoid this problem. In their paper they note that, at least for certain stimuli, binocular neurons that are tuned to different spatial frequencies will respond to different false matches. Consequently they argue that false matches can be eliminated simply by pooling the responses of several binocular neurons, each tuned to a different spatial frequency. Although they demonstrated the proficiency of their model when it was presented with white noise stimuli, it is not clear how their model could be extended to other stimuli, in particular those situations were contrast affects the perceived solution of the correspondence problem (Section 3.2) or where monocular information contributes to depth perception (Section 3.5).

Another way to solve the correspondence problem is to utilize a disparity filter that implements the unique-matching rule, which states that any given feature in one retinal image is matched at most with one feature in the other retinal image (Marr & Poggio, 1976; Grimson, 1981; for a review see Howard & Rogers, 1995, pp. 42-43). As discussed in the Introduction, this rule fails in Panum's limiting case (Panum, 1858; Gillam et al., 1995; McKee et al., 1995).

This failure caused Grossberg and McLoughlin (1997) and McLoughlin and Grossberg (1998) to design a disparity filter that encouraged unique-matching without enforcing it. Their model forms the foundation for our own and can simulate much of the same data, including most of the dichoptic masking and the correspondence problem data. Their model also makes an incorrect psychophysical prediction: that if each eye sees a single bar, then the ratio constraint on stereoscopic fusion (Smallman & McKee, 1995) ensures that fusion will occur only if the magnitudes of the contrasts of the two bars do not differ too greatly. This is inconsistent with experimental findings which indicate that the ratio constraint does not apply to this special case (McKee et al., 1994; Smallman & McKee, 1995).

The present model refines the Grossberg and McLoughlin model to correct this short-coming. In particular, for the purposes of the disparity filter, the Grossberg and McLoughlin model assigned all unfused boundaries to the fixation plane, whereas the present model adds unfused boundaries to all fixations planes and then lets the V2 disparity filter eliminate boundary representations as necessary. As explained in Section 3.1.1 this procedure allows in the special case where each eye sees only a single bar the two bars to be binocularly fused regardless of their contrast difference. The present model has simulated all the data considered by McLoughlin and Grossberg (1998), specifically the data on contrast variations of dichoptic masking and the correspondence problem, and has also simulated additional data including the Venetian blind illusion, four different examples of da Vinci stereopsis (Nakayama & Shimojo, 1990; Gillam et al., 1999), stereopsis with opposite-contrast

stimuli, the effect of interocular contrast differences on stereoacuity and the Craik-O'Brian-Cornsweet lightness illusion. Furthermore, unlike its predecessor, it has been mapped onto known cortical cells and laminar circuits within cortical areas V1, V2 and V4.

The model presented in this article and the Grossberg and McLoughlin model both instantiated key aspects of FACADE theory. Another model, which is also a simplified version of FACADE theory, was used to explain a series of experiments on the McCollough effect, an orientation-sensitive, long-lasting, chromatic after-effect (Grossberg, Hwang & Mingolla, 2002). Unlike the present model, the binocular cells in the McCollough effect model did not exhibit the "obligate" property in that they responded to monocular inputs, albeit less strongly than to binocular inputs. Although, the McCollough effect model did not make use of obligate cells, inserting such cells into the model would not disrupt its simulations. These obligate cells would merely be unnecessary. Similarly, the addition of non-obligate binocular cells into V1 of the present model, while unnecessary, would not reduce its explanatory power. In particular, such an addition would merely increase the number of false matches that occur in V1. These false matches would be eliminated by the V2 disparity filter, in the manner outlined in Section 2.3, and so would not contribute to the final percept. Taken together, these two models help explain the differing roles of obligate and non-obligate binocular cells in the broader context of FACADE theory and help to functionally explain why both obligate and non-obligate cells have been found experimentally to exist (Poggio, 1991).

4.3 Model robustness and complexity

The model is robust in the sense that the absolute values of the model parameters can be varied over large ranges without disrupting its explanations of data; only their values relative to each other are important. Furthermore, there is considerable scope when choosing individual parameter values, since no single parameter proves to be critical in any simulation.

The model is minimally complex in the sense that each of its four interacting networks, V1 binocular boundaries, V1 monocular boundaries, V2 boundaries, and V4 surfaces, are essential. The V1 binocular boundaries network is needed to explain stereopsis and the contrast ratio constraint observed in stereoscopic fusion (Smallman & McKee, 1994). The V1 monocular boundaries network plays a role in explaining da Vinci stereopsis (Nakayama & Shimojo, 1990; Gillam et al., 1999), dichoptic masking (McKee et al., 1994), contrast variations of the correspondence problem (Smallman & McKee, 1994) as well as in some examples of stereopsis with opposite contrast stimuli (e.g., Howe & Watanabe, 2002). The V2 boundaries network is needed to solve both the correspondence problem and the monocular-binocular interface problem by utilizing its disparity filter. The correspondence problem arises because V1 sometimes incorrectly fuses contours that belong to different objects. The monocular-binocular interface problem is caused because the V1 monocular boundaries, not having a definite depth association, are initially added to all depth planes. Finally, the surface network includes cells in V4, the V2 thin stripes and the V1 blobs. It is necessary because it is surface percepts, not boundary percepts, that subjects report in the experimental studies considered by this paper and also because, as illustrated by all of the simulations, not all boundaries give rise to a percept of depth.

If anything, the model in Figure 1 is too simple to explain all data about depth perception. Fortunately, the analysis in this article has opened a clear path to generalize the model, as illustrated below.

4.4 Generalizing to Natural images, 3D Boundary Completion and 3D Attention

One of the long-term goals of this modeling work is to extend the present model so that it can be applied to natural images. The simulations already done show that the model can resolve a wide range of potentially confusing false matches. There remain, however, two impediments that the model first needs to overcome.

First, the present model can represent only 3D planes that are flat and perpendicular to the observer. To analyze natural images, the model needs to be extended to represent slanted and curved surfaces in 3D. A parallel line of research has begun to demonstrate how it can be consistently generalized to explain such data (Swaminathan & Grossberg, 2001).

Second, the present model shows how boundaries can be formed using bottom-up inputs from the outside world. It does not, however, indicate how horizontal interactions can be used to complete these boundaries where pixels are missing either due to internal brain imperfections, such as the blind spot in the retina, or due to incomplete contours in external inputs, whether due to noise, occluding surfaces, spatially discrete texture elements, illusory contour stimuli, or even missing pixels in impressionist paintings. Nor does it clarify how these circuits can develop, be modified by learning, or modulated by top-down attention. This omission can be overcome as follows.

A parallel line of modeling has developed quantitative explanations and simulations of how processes of perceptual development, learning, grouping, and attention may be achieved by laminar cortical circuits (Grossberg, 1999; Grossberg, Mingolla, & Ross, 1997; Grossberg & Raizada, 2000; Grossberg & Williamson, 2001; Raizada & Grossberg, 2001). This LAMINART model did not, however, investigate how these boundaries may be completed in three-dimensions. We now show that the LAMINART model of boundary completion is consistent with the model of three-dimensional boundary formation that is summarized in Figure 1. In particular, we demonstrate below how the LAMINART model be can generalized to cope with the positional displacement, or allelotropia, that is characteristic of binocular fusion (cf. Figure 3).

Figure 17 summarizes some of the key LAMINART interactions that govern perceptual grouping, and attention, without regard to its three-dimensional representation. Figure 18 shows how the LAMINART perceptual grouping and attention circuit naturally generalize to a 3D LAMINART model that is consistent with the system interactions in Figure 1. This extended model clarifies how 3D boundaries can be completed and how attention can be selectively paid to objects in 3D. The following new features in Figure 18 show how these properties obtain. First, layer 4 no longer directly activates layer 2/3, as in Figure 17. Instead, layer 4 simple cells first activate layer 3B simple cells, which in turn activate layer 2/3A complex cells, as shown in Figure 1. The layer 2/3A cells can then interact via horizontal interactions, like those summarized in Figures 17c and 17e, to complete boundaries. Second, binocular cells in layer 2/3A can represent different disparities, and thus different relative depths from an observer. Interactions between layer 2/3A cells that represent the same relative depth from the observer can be used to complete boundaries between object contours that lie at the same depth.

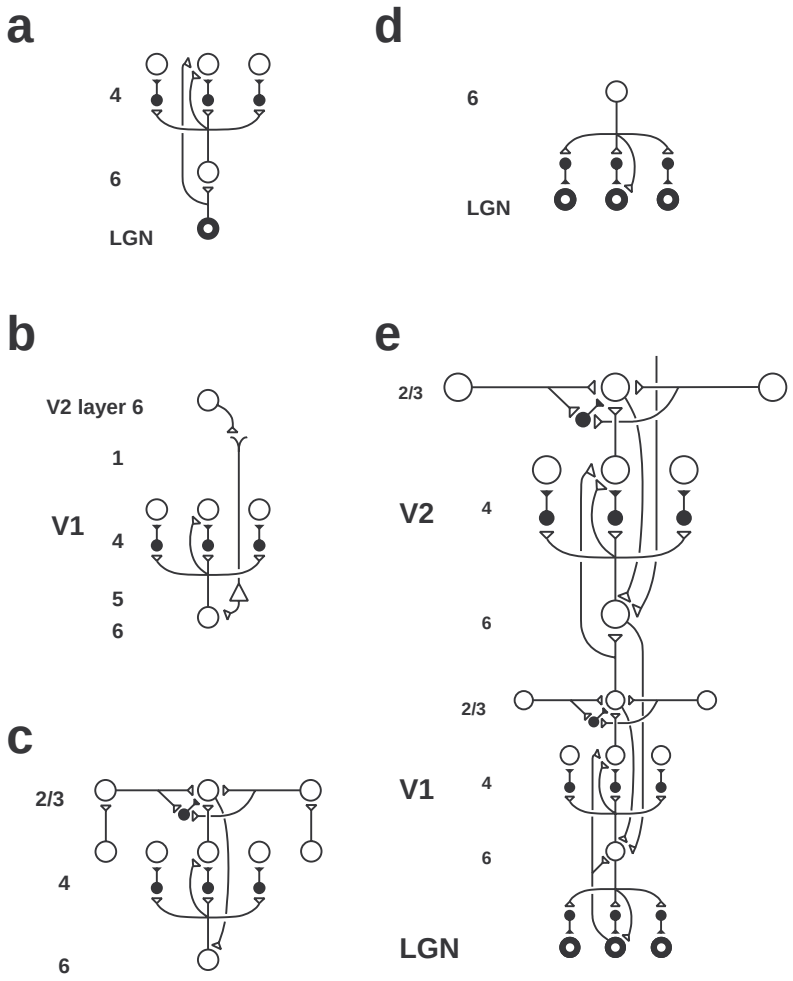

Figure 17: The LAMINART model: (a) The LGN directly activates V1 layers 4 and 6. Layer 6, in turn, sends a pattern of on-center, off-surround inputs to layer 4. These layer 6 inputs can strongly inhibit layer 4 through the off-surround, but the excitatory and inhibitory inputs in the on-center are approximately balanced so that layer 6 can modulate the excitability of layer 4 cells, but not fully drive them to fire vigorously. The direct connections from LGN to layer 4 carry out this driving function. (b) This layer 6-to-4 circuit can be used by top-down signals from V2 layer 6 to attentionally modulate the excitability of V1 layer 4 cells. (c) Boundary completion can occur when layer 4 cells activate layer 2/3 cells, which communicate with their layer 2/3 neighbors via long-range horizontal excitatory connections and shorter-range inhibitory interneurons. The balance between these excitatory and inhibitory interactions allows boundaries to form inwardly between properly oriented image contrasts, as in the case of many illusory contours, but not outwardly from individual contrasts. The strongest boundary groupings in layer 2/3 can support themselves best through the positive feedback loop between layers 2/3-to-6-to-4-to-2/3, even as their strong inhibitory signals in the layer 6-to-4 off-surround can inhibit weaker groupings. (d) A top-down on-center, off-surround from V1 layer 6 to the LGN acts like the top-down signals from V2 layer 6 to V1 layer 4. (e) The LAMINART system architecture. Note that the horizontal interactions within V2 layer 2/3 can have a broader spatial extent than those in V1 layer 2/3. The longer-range V2 interactions carry out the type of perceptual groupings that are familiar in illusory contours, texture grouping, completion of occluded objects, and bridging the blind spot. [Reprinted with permission from Grossberg and Raizada (2000).]

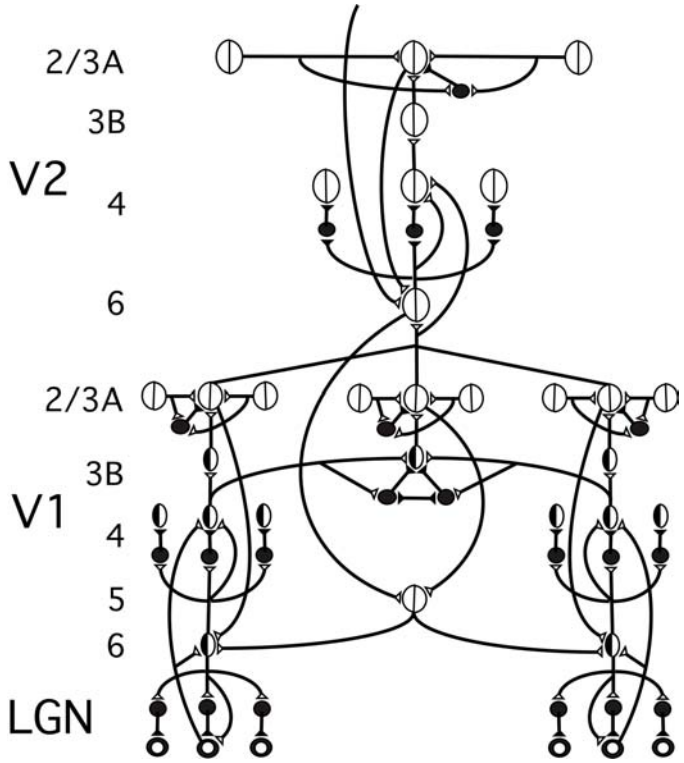

Figure 18: A 3D LAMINART model, including 3D boundary completion and attention, as well as the binocular and monocular interactions summarized in Figure 1. See text for details.

Due to the binocular fusion that occurs in layer 3B, the binocular boundaries that are formed in layers 3B and 2/3A can be positionally displaced, or shifted, relative to their monocular input signals to layers 6 and 4. Figure 17c suggests that these layer 2/3 boundaries feed signals back to layer 6 in order to select the winning groupings that are formed in layer 2/3. How can the positionally displaced binocular boundaries in layer 2/3A of Figure 17 contact the correct monocularly activated cells in layers 6 and 4, so as to complete the feedback loop 2/3A-to-6-to-4-to-3B-to-2/3A that can select the winning three-dimensional groupings? In particular, how can the feedback signal from a layer 2/3A cell that is positionally displaced with respect to its monocular inputs activate horizontal signals that can activate the correct layer 6 monocular sources?

We propose that horizontal connections that are known to occur in layer 5 (Callaway & Wiser, 1996) accomplish this. Feedback signals from layer 2/3A propagate vertically to layer 5, whose cells activate horizontal axons in this layer that contact the appropriate layer 6 cells. These layer 5-to-6 contacts are assumed to be selectively formed during development. Grossberg and Williamson (2001) have simulated how layer 2/3 connections and layer 6-to-4 connections may originate during development. The selective layer 5-to-6 contacts are proposed to form according to similar laws. In summary, inward horizontal layer 4-to-3B and 2/3A-to-2/3A connections are proposed to form binocular cells and their groupings, while outward layer 5-to-6 connections are proposed to close the feedback loops that help to select the correct three-dimensional groupings.

Once the generalization to 3D boundaries is made, top-down attentional modulation of these boundaries follows directly by using the same circuits as in Figure 17.

Appendix A: Model Equations

To make the equations easier to read, capital letters denote variables and lower case letters denote constants. Appendix A is most easily read in conjunction with Figure 1, which depicts the model. Each eye's stimulus was presented on a grid 55 units high and 70 units wide except Figures

11 and 12 which were presented on a 55 by 126 grid and Figure 14b which was presented on a 55 by 85 grid. In all simulations, white had a luminance value, in arbitrary units, of 2. In Figures 5-9, 11-12 and 15, the light gray bars (if any) had a luminance of 0.85 and the dark gray bars 0.68. In Figures 13 and 14 medium gray was represented by a luminance of 0.75, and black by 0.3. In Figure 16, the simulation of the Craik-O'Brian-Cornsweet effect, there was a luminance cusp which ranged from 0.4-0.9. Simulations were performed using the *Matlab*® software package. Analytical equilibrium solutions of the differential equations were used in all cases except for the V2 disparity filter equation (A15) and the V4 diffusion equation (A18), which could not be solved analytically, and were instead solved using Euler's method, and then solved again using a different step-size to verify the accuracy of the original solution.

LGN

The LGN cells obey membrane, or shunting, equations that receive input from the retina and are assumed to have circularly symmetric on-center, off-surround receptive fields. When these fields are approximately balanced, the network discounts the illuminant and contrast-normalizes its cell responses (Grossberg & Todorović, 1988). The LGN cell membrane potentials, $X_{ij}^{L/R}$, obey the following differential equation:

$$\frac{dX_{ij}^{L/R}}{dt} = -\varepsilon X_{ij}^{L/R} + \left(a - X_{ij}^{L/R}\right) I_{ij}^{L/R} - X_{ij}^{L/R} \sum_{p \neq i, q \neq j} g_{pqij} I_{pq}^{L/R}, \tag{A1}$$

where L/R designates that the cell belongs to the left or right monocular pathway, indices i and j denote the position of the input on the retina, ε is a constant (10^{-5}) that represents the rate of decay of the cell membrane potential, a is a constant (9.9) that represents the maximum membrane potential, $I_{ij}^{L/R}$ is the unnormalized luminance of the left or right retinal image, and g_{pqij} is a Gaussian kernel that represents the inhibitory off-surround:

$$g_{pqij} = \exp\left(-\frac{(p-i)^2 + (q-j)^2}{2\sigma^2}\right), \tag{A2}$$

where σ represents the size of the kernel (1.5). The steady-state cell membrane potentials of these cells are given by:

$$X_{ij}^{L/R} = \frac{aI_{ij}^{L/R}}{\varepsilon + \sum_{p,q} g_{pqij} I_{ij}^{L/R}}. \tag{A3}$$

The steady-state equation (A3) was used in the simulations. Below all equations that were solved at steady-state are given in their steady-state form.

V1 layer 4 simple cells

All cells in V1 layer 4 are modeled as monocular simple cells that are sensitive to either dark-light or light-dark contrast polarity, but not both, depending on their receptive field structure. At steady-state the membrane potentials, $S_{ij}^{H/V,L/R,+}$, of simple cells that respond to dark-light contrast polarity are given by:

$$S_{ij}^{H/V,L/R,+} = \sum_{p,q} k_{pq}^{H/V} \left[X_{i+p,j+q}^{L/R}\right]^+. \tag{A4}$$

where H/V designates that the cell responds to horizontal or vertical boundaries, + indicates that the simple cell responds to dark-light contrast polarity, $[x]^+ = \max(x,0)$, and $k_{pq}^{H/V}$ is a Gabor function representing the simple cell receptive field kernel:

$$k_{pq}^{H/V} = \phi \sin\left(\frac{2\Pi r}{\tau}\right)\exp\left[-\frac{1}{2}\left(\frac{p^2}{\sigma_p^2}+\frac{q^2}{\sigma_q^2}\right)\right], \tag{A5}$$

where ϕ , τ , σ_p , σ_q are constants (4.4, 3Π, 0.6, 0.6) representing the amplitude and dimensions of this kernel; $r = p$ for cells that respond to vertical boundaries; and $r = q$ for those that respond to horizontal boundaries.

The cell membrane potentials of the simple cells with light-dark contrast polarity are the inverse of the previous cell membrane potentials:

$$S_{ij}^{H/V,L/R,-} = -S_{ij}^{H/V,L/R,+} = -\sum_{p,q} k_{pq}^{H/V}\left[X_{i+p,j+q}^{L/R}\right]^+ . \tag{A6}$$

Layer 3B monocular simple cells

At steady-state the membrane potentials, $B_{ij}^{H/V,L/R,+/-}$, of the layer 3B monocular cells are given by:

$$B_{ij}^{H/V,L/R,+/-} = 2\left[S_{ij}^{H/V,L/R,+/-}\right]^+ . \tag{A7}$$

where the multiplicative factor of 2 compensates for the fact that the monocular simple cells receive inputs from only one eye whereas the binocular simple cells, discussed in the next section, receive input from both eyes.

Layer 3B inhibitory cells

The layer 3B inhibitory cells, all responding only to vertical boundaries, receive excitatory input from layer 4 and inhibitory input from all other inhibitory interneurons that correspond to the same position and disparity. Their cell membrane potentials, $Q_{ijd}^{V,L/R,+/-}$, are determined at equilibrium by the following equations**:**

$$Q_{ijd}^{V,L,+/-} = \frac{1}{\gamma_2}\left(\left[S_{(i+s)j}^{V,L,+/-}\right]^+ - \beta\left(\left[Q_{ijd}^{V,R,+/-}\right]^+ + \left[Q_{ijd}^{V,R,-/+}\right]^+ + \left[Q_{ijd}^{V,L,-/+}\right]^+\right)\right), \tag{A8}$$

and

$$Q_{ijd}^{V,R,+/-} = \frac{1}{\gamma_2}\left(\left[S_{(i-s)j}^{V,R,+/-}\right]^+ - \beta\left(\left[Q_{ijd}^{V,L,+/-}\right]^+ + \left[Q_{ijd}^{V,L,-/+}\right]^+ + \left[Q_{ijd}^{V,R,-/+}\right]^+\right)\right), \tag{A9}$$

where γ_2 and β are constants (4.5, 4) representing the decay rate of the membrane potential and the strength of the inhibition, *d* is the disparity to which the model neuron is tuned and *s* is the allelotropic shift that depends on the disparity and is defined in Table 1.

Layer 3B binocular cells

The layer 3B binocular cells, all of which are vertically oriented and receive excitatory input from layer 4 and inhibitory input from the layer 3B inhibitory cells that correspond to the same position and disparity, have membrane potentials $B_{ijd}^{V,B,+/-}$. At equilibrium:

$$B_{ijd}^{V,B,+/-} = \frac{1}{\gamma_1}\left(\left[S_{(i+s)j}^{V,L,+/-}\right]^+ + \left[S_{(i-s)j}^{V,R,+/-}\right]^+ - \alpha\left(\left[Q_{ijd}^{V,L,+/-}\right]^+ + \left[Q_{ijd}^{V,L,-/+}\right]^+ + \left[Q_{ijd}^{V,R,+/-}\right]^+ + \left[Q_{ijd}^{V,R,-/+}\right]^+\right)\right), \tag{A10}$$

where γ_1 and α and are constants (0.29, 6) representing the rate of decay of the membrane potential and the strength of the inhibition. Appendix B proves that the exact values of α and γ_1 are not critical. Under mild constraints on these parameters, the binocular cells act like the "obligate cells" of Poggio (1991), responding only when their left and right inputs are approximately equal in magnitude. Equation (A10) was solved at equilibrium, using the theorem described in Appendix B to

speed up the simulations. Figure 19 shows that the calculated and simulated values are essentially identical.

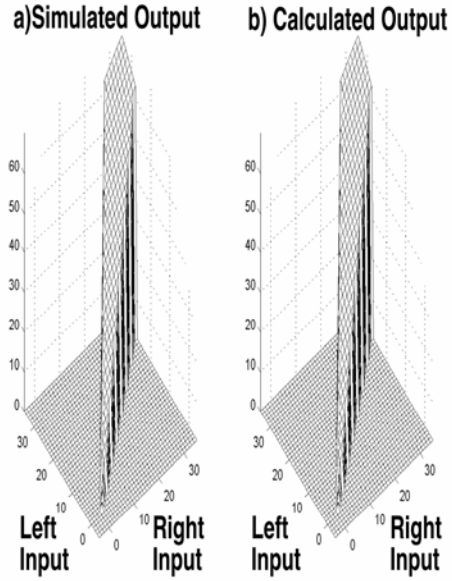

Figure 19

Figure 19: Output of a V1 layer 3B binocular simple cell using the parameter values of Appendix A, equation (A10). (a) Simulated; (b) Calculated.

Layer 2/3A monocular and binocular complex cells

V1 layer 2/3A consists of both monocular and binocular complex cells. These complex cells pool the cell membrane potentials of monocular/binocular layer 3B simple cells of like orientation and *both* contrast polarities at each position. At steady-state their membrane potentials, $C_{ijd}^{H/V,L/R/B}$, are given by:

$$C_{ijd}^{H/V,L/R/B} = \left[B_{ijd}^{H/V,L/R/B,+}\right]^{+} + \left[B_{ijd}^{H/V,L/R/B,-}\right]^{+}. \tag{A11}$$

V2 layer 4

In V2, virtually all cells are binocularly driven (Hubel & Livingstone, 1987) consistent with the model hypothesis that the left and right monocular inputs are combined in V2. The model assumes that this is done in layer 4. Since the monocular inputs do not yet have a depth associated with them, they are added to all depth planes along their respective lines-of-sight (cf. Section 2.3). At steady-state their membrane potentials, J_{ijd}^{H}, are given by:

$$J_{ijd}^{H} = \left[C_{(i+s)j}^{H,L} - \theta\right]^{+} + \left[C_{(i-s)j}^{H,R} - \theta\right]^{+}. \tag{A12}$$

where θ is a constant (1.42) representing the threshold of the V1 layer 2/3A cells and *s* is the allelotropic shift defined in Table 1. Similarly, at steady-state the cell membrane potentials of the vertically oriented layer 4 complex cells are given by:

$$J_{ijd}^{V} = \left[C_{ijd}^{V,B} - \theta\right]^{+} + \beta\left(\left[C_{(i+s)j}^{V,L} - \theta\right]^{+} + \left[C_{(i-s)j}^{V,R} - \theta\right]^{+}\right), \tag{A13}$$

where β is a constant (0.21) representing the strength of the monocular connections.

Disparity (*d*)	**V. Near disparity**	**Near disparity**	**Zero disparity**	**Far disparity**	**V. Far disparity**
Allelotropic shift (*s*)	-8	-4	0	+4	+8

Table 1: The allelotropic shift (*s*) is the amount that the left and right monocular contours must be displaced to form a single fused binocular contour. It depends on the disparity. It is zero for matches in the fixation plane because these matches are between contours at retinal correspondence. Figure 3 illustrates the allelotropic shift and shows that a left monocular contour needs to be shifted more to the right for matches that are further from the observer, whereas a right monocular contour needs to be shifted in the opposite direction.

V2 layer 3B

Analogous to layer 4, at steady-state the cell membrane potentials, N_{ijd}^{H}, of the horizontally oriented layer 3B cells are given by:

$$N_{ijd}^{H} = \left[J_{ijd}^{H}\right]^{+}. \tag{A14}$$

V2 layer 3B contains the disparity filter (cf. Figure 3) in which each vertically oriented cell is inhibited by every other vertically oriented cell that shares either of its monocular inputs (represented by the solid lines) or corresponds to the same 2D position but a different depth (represented by the dashed line). The cell membrane potentials of the vertically oriented layer 3B cells are given by:

$$\frac{dN_{ijd}^{V}}{dt} = -N_{ijd}^{V} + \left[J_{ijd}^{V} - \delta^{V}\right]^{+} - \eta \sum_{d' \neq d} \left(m_{dd'} \left[N_{(i+s'-s)jd'}^{V}\right]^{+} + m_{dd'} \left[N_{(i+s-s')jd'}^{V}\right]^{+} + \mu \left[N_{ijd'}^{V}\right]^{+} \right), \tag{A15}$$

where *d* and *d'* represent disparities; *s* and *s'* are the corresponding allelotropic shifts, defined by Table 1; δ^{V} is a constant (0.15) that represents the threshold of a vertically oriented layer 4 cell, η is a constant (0.38) that scales the total inhibition that each cell receives, μ is a constant (0.1) that represents the inhibition from boundaries directly in front or behind, and $m_{dd'}$ represents the inhibition from all other neurons that share an input, as detailed in Table 2.

	V. Near	**Near**	**Zero**	**Far**	**V. Far**
V. Near	-	3	5	3	2
Near	0.4	-	2.8	1.5	0.4
Zero	0.2	1.3	-	1.3	0.2
Far	0.4	1.5	2.8	-	0.4
V. Far	2	3	5	3	-

Table 2: The inhibition coefficients $m_{dd'}$. Each neuron is inhibited by every other neuron that shares either of its inputs by an amount that depends on the disparities of the inhibited and inhibiting neurons (cf. Figure 3). See text for further discussion of parameter choices.

The disparity filter is robust in that its behavior is stable across a range of parameter values. The key features of the disparity filter, as illustrated in Table 2, is that it is symmetrical about the fixation plane (i.e., the near and far disparity planes equally inhibit and are equally inhibited by the zero disparity plane) and that it favors the zero disparity plane in that this plane inhibits the near and far disparity planes more than they inhibit it. Equation (A15) was solved using Euler's method and then resolved with a different step size to check the accuracy of the solution.

V2 layer 2/3A complex cells

The V2 layer 2/3A cells receive input from V2 layer 3B. In the present model, they merely scale this input, but see Grossberg (1999) for a discussion of their function in perceptual grouping. Analogous to layer 4, their steady-state cell membrane potentials, $T_{ijd}^{H/V}$, are given by:

$$T_{ijd}^{H/V} = 50 * \left[N_{ijd}^{H/V}\right]^{+}, \tag{A16}$$

where the multiplicative factor of 50 allows the V2 plots in Figures 5-9 and 11-16 to be plotted to the same scale as the V1 plots.

V4

V4 receives lightness signals from the LGN via the V1 blobs and the V2 thin stripes, and boundary signals via the V1 interblob regions and the V2 pale stripes. It combines the monocular lightness signals from the two eyes that correspond to the same 3D location. Its lightness input, Z_{ijd}, is given by:

$$Z_{ijd} = \left[X^{L}_{(i+s)j}\right]^{+} + \left[X^{R}_{(i-s)j}\right]^{+}, \tag{A17}$$

where *i, j* are positional indices, *d* represents the disparity to which the model neuron is tuned and *s* the allelotropic shift defined in Table 1. The on-center off-surround receptive field structure of the LGN ensures that these lightness signals are only present at the borders of surfaces, where the lightness signals represent the ratio of the luminances on either side of the border. In V4, these lightness signals fill-in by a process whose lateral spread is gated, and thereby contained, by boundary signals. Following Grossberg and Todorović (1988), the V4 cell membrane potentials, W_{ijd}, are modeled by a diffusion equation that is solved at steady-state:

$$W_{ijd} = \frac{Z_{ijd} + \sum_{p,q \in n_{ij}} W_{pqd} P_{pqijd}}{1 + \sum_{p,q \in n_{ij}} P_{pqijd}}. \tag{A18}$$

Equation (A18) was solved iteratively until equilibrium was achieved. Diffusion of potential occurs between the nearest-neighbor locations n_{ij} of (*i,j*):

$$n_{ij} = \{(i, j-1), (i-1, j), (i+1, j), (i, j+1)\}. \tag{A19}$$

The gating coefficients, P_{pqijd}, in (A18) represent the inhibition of the diffusion by the boundary signals $T_{pqd}^{H/V}$ and $T_{ijd}^{H/V}$. They are defined by:

$$P_{pqijd} = \frac{f}{1 + h\left(T^{H}_{(i-0.5)(j+0.5)d} + T^{V}_{(i-0.5)(j+0.5)d} + Y^{H}_{(i-0.5)(j-0.5)d} + Y^{V}_{(i-0.5)(j-0.5)d}\right)}, \quad \text{if p = } i\text{-1 and q = } j, \tag{A20}$$

$$P_{pqijd} = \frac{f}{1 + h\left(T^{H}_{(i+0.5)(j+0.5)d} + T^{V}_{(i+0.5)(j+0.5)d} + Y^{H}_{(i+0.5)(j-0.5)d} + Y^{V}_{(i+0.5)(j-0.5)d}\right)}, \quad \text{if p = } i\text{+1 and q = } j, \tag{A21}$$

$$P_{pqijd} = \frac{f}{1 + h\left(T^{H}_{(i-0.5)(j-0.5)d} + T^{V}_{(i-0.5)(j-0.5)d} + Y^{H}_{(i+0.5)(j-0.5)d} + Y^{V}_{(i+0.5)(j-0.5)d}\right)}, \quad \text{if p} = i \text{ and q} = j\text{-}1, \tag{A22}$$

$$P_{pqijd} = \frac{f}{1 + h\left(T^{H}_{(i-0.5)(j+0.5)d} + T^{V}_{(i-0.5)(j+0.5)d} + Y^{H}_{(i+0.5)(j+0.5)d} + Y^{V}_{(i+0.5)(j+0.5)d}\right)}, \quad \text{if p} = i \text{ and q} = j\text{+}1, \tag{A23}$$

As shown by these equations, the boundary lattice is offset by [0.5 0.5] relative to the lightness lattice, corresponding to the idea that these two processing streams are spatially displaced with respect to one another in the cortical map. Constants f and h (1000, 10000) represent the diffusion rate and the strength of the gating inhibition. Spurious lattice edge effects were avoided by using the wrap-round technique according to which the last element of a row/column is adjacent to the first element of the same row/column.

Appendix B: Proof of the Obligate theorem

The obligate property is proved below. The proof shows that binocular simple cells in layer 3B can be activated only if they receive approximately equal inputs from both left and right eye monocular simple cells in layer 4. The constraints that determine cell firing depend upon the ratios of left and right monocular cell activity. This property explains the ratio constraint on stereoscopic fusion that is illustrated in Figure 10. The proof also shows that the activities converge exponentially to unique equilibrium activities in response to any combination of constant monocular inputs. The obligate property is caused by a balance between excitatory inputs from layer 4 monocular simple cells and inhibitory inputs from layer 3B inhibitory interneurons. The interneurons are themselves activated by layer 4 monocular simple cells and mutually inhibit each other, in addition to inhibiting the binocular simple cells.

The obligate property is proved below for layer 3B binocular cells with dark-light contrast polarity ($B^{V,B,+}_{ijd}$). By symmetry, the same proof holds for cells with light-dark contrast polarity ($B^{V,B,-}_{ijd}$).

Obligate Theorem

Consider the system:

$$\frac{dB^{V,B,+}_{ijd}}{dt} = -\gamma_1 B^{V,B,+}_{ijd} + \left(\left[S^{V,L,+}_{(i+s)j}\right]^+ + \left[S^{V,R,+}_{(i-s)j}\right]^+\right) - \alpha\left(\left[Q^{V,L,+}_{ijd}\right]^+ + \left[Q^{V,L,-}_{ijd}\right]^+ + \left[Q^{V,R,+}_{ijd}\right]^+ + \left[Q^{V,R,-}_{ijd}\right]^+\right), \tag{B1}$$

$$\frac{dQ^{V,L,+}_{ijd}}{dt} = -\gamma_2 Q^{V,L,+}_{ijd} + \left[S^{V,L,+}_{(i+s)j}\right]^+ - \beta\left(\left[Q^{V,R,+}_{ijd}\right]^+ + \left[Q^{V,R,-}_{ijd}\right]^+ + \left[Q^{V,L,-}_{ijd}\right]^+\right), \tag{B2}$$

$$\frac{dQ^{V,R,+}_{ijd}}{dt} = -\gamma_2 Q^{V,R,+}_{ijd} + \left[S^{V,R,+}_{(i-s)j}\right]^+ - \beta\left(\left[Q^{V,L,+}_{ijd}\right]^+ + \left[Q^{V,L,-}_{ijd}\right]^+ + \left[Q^{V,R,-}_{ijd}\right]^+\right), \tag{B3}$$

$$\frac{dQ^{V,L,-}_{ijd}}{dt} = -\gamma_2 Q^{V,L,-}_{ijd} + \left[S^{V,L,-}_{(i+s)j}\right]^+ - \beta\left(\left[Q^{V,R,-}_{ijd}\right]^+ + \left[Q^{V,R,+}_{ijd}\right]^+ + \left[Q^{V,L,+}_{ijd}\right]^+\right), \tag{B4}$$

and

$$\frac{dQ^{V,R,-}_{ijd}}{dt} = -\gamma_2 Q^{V,R,-}_{ijd} + \left[S^{V,R,-}_{(i-s)j}\right]^+ - \beta\left(\left[Q^{V,L,-}_{ijd}\right]^+ + \left[Q^{V,L,+}_{ijd}\right]^+ + \left[Q^{V,R,+}_{ijd}\right]^+\right), \tag{B5}$$

where $S^{V,L,+}_{(i+s)j}$ and $S^{V,R,+}_{(i-s)j}$ are monocular simple cell activities that are defined by (A4) and (A6), $0 < \gamma_1$ and

$$0 < \beta < \gamma_2 < \alpha < \gamma_2 + \beta \tag{B6}$$

Under these conditions, the system converges exponentially to the unique equilibria specified by (1), (2), (3) and (4) provided that the inputs are constant.

(1) if $0 < S^{V,L,+}_{(i+s)j}, S^{V,R,+}_{(i-s)j};\ S^{V,L,-}_{(i+s)j}, S^{V,R,-}_{(i-s)j} < 0$; $\dfrac{\beta}{\gamma_2} \le \dfrac{S^{V,L,+}_{(i+s)j}}{S^{V,R,+}_{(i-s)j}}$; and $\dfrac{\beta}{\gamma_2} \le \dfrac{S^{V,R,+}_{(i-s)j}}{S^{V,L,+}_{(i+s)j}}$,

then at equilibrium $B^{V,B,+}_{ijd} = \dfrac{1}{\gamma_1}\left(1 - \dfrac{\alpha}{\gamma_2 + \beta}\right)\left(S^{V,L,+}_{(i+s)j} + S^{V,R,+}_{(i-s)j}\right)$; (B7)

(2) if $0 < S^{V,L,+}_{(i+s)j}, S^{V,R,+}_{(i-s)j};\ S^{V,L,-}_{(i+s)j}, S^{V,R,-}_{(i-s)j} < 0$; and $\dfrac{S^{V,R,+}_{(i-s)j}}{S^{V,L,+}_{(i+s)j}} < \dfrac{\beta}{\gamma_2}$,

then at equilibrium $B^{V,B,+}_{ijd} = \dfrac{1}{\gamma_1}\left(S^{V,R,+}_{(i-s)j} + \left(1 - \dfrac{\alpha}{\gamma_2}\right)S^{V,L,+}_{(i+s)j}\right)$; (B8)

(3) if $0 < S^{V,L,+}_{(i+s)j}, S^{V,R,+}_{(i-s)j};\ S^{V,L,-}_{(i+s)j}, S^{V,R,-}_{(i-s)j} < 0$; and $\dfrac{S^{V,L,+}_{(i+s)j}}{S^{V,R,+}_{(i-s)j}} < \dfrac{\beta}{\gamma_2}$,

then at equilibrium $B^{V,B,+}_{ijd} = \dfrac{1}{\gamma_1}\left(S^{V,L,+}_{(i+s)j} + \left(1 - \dfrac{\alpha}{\gamma_2}\right)S^{V,R,+}_{(i-s)j}\right)$; (B9)

(4) for all other values of $S^{V,L,+}_{(i+s)j}, S^{V,R,+}_{(i-s)j}, S^{V,L,-}_{(i+s)j}, S^{V,R,-}_{(i-s)j}$,

at equilibrium $B^{V,B,+}_{ijd} \le 0$. (B10)

Figures 18a and 18b show the simulated and calculated outputs for the above system for the parameter values summarized in Appendix A.

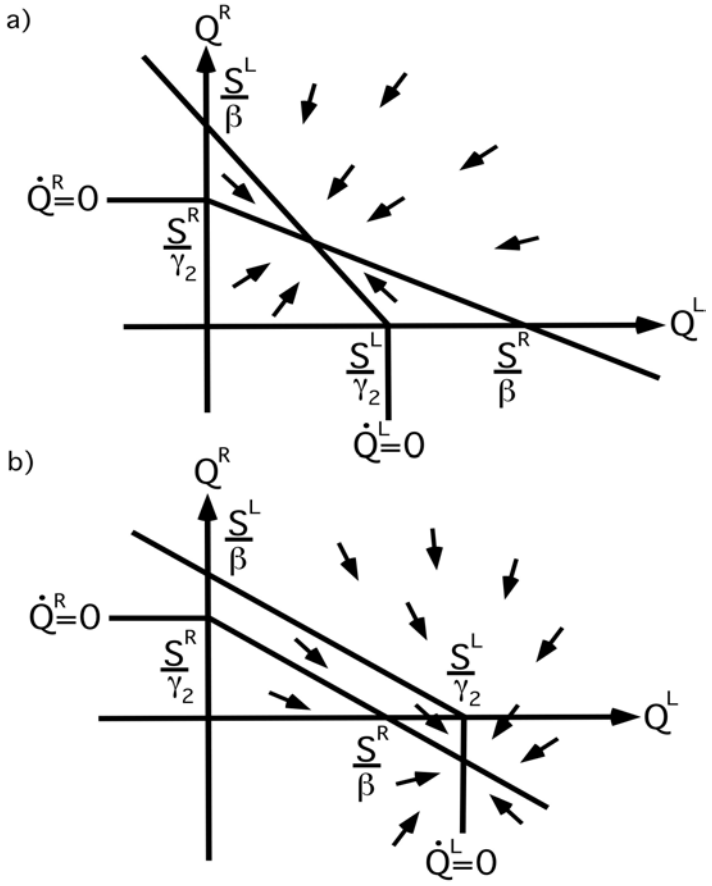

Figure 20: Phase-plots used in Appendix B: (a) Case 1; (b) Case 2. In both cases the $\frac{dQ_{ijd}^{V,R,+}}{dt}=0$ nullcline crosses the ordinate and the abscissa at $\frac{S_{(i-s)j}^{V,R,+}}{\gamma_2}$ and $\frac{S_{(i-s)j}^{V,R,+}}{\beta}$, while the $\frac{dQ_{ijd}^{V,L,+}}{dt}=0$ nullcline crosses them at $\frac{S_{(i+s)j}^{V,L,+}}{\beta}$ and $\frac{S_{(i+s)j}^{V,L,+}}{\gamma_2}$.

Proof

First note by (A6) that, out of the four possible inputs $S_{(i+s)j}^{V,L,+}$, $S_{(i-s)j}^{V,R,+}$, $S_{(i+s)j}^{V,L,-}$, and $S_{(i-s)j}^{V,R,-}$, at most only two can be positive. This greatly simplifies the subsequent analysis.

Case 1: $0<S_{(i+s)j}^{V,L,+},S_{(i-s)j}^{V,R,+};S_{(i+s)j}^{V,L,-},S_{(i-s)j}^{V,R,-}<0;\frac{\beta}{\gamma_2}\le\frac{S_{(i+s)j}^{V,L,+}}{S_{(i-s)j}^{V,R,+}}$; and $\frac{\beta}{\gamma_2}\le\frac{S_{(i-s)j}^{V,R,+}}{S_{(i+s)j}^{V,L,+}}$. (B11)

Under these conditions, (B4) and (B5) imply that, for sufficiently large t,

$$Q_{ijd}^{V,L,-},Q_{ijd}^{V,R,-}\le 0. \tag{B12}$$

By (B12), and recalling for this case $0<S_{(i+s)j}^{V,L,+},S_{(i-s)j}^{V,R,+}$, (B2) and (B3) can be approximated at large times:

$$\frac{dQ_{ijd}^{V,L,+}}{dt}=-\gamma_2 Q_{ijd}^{V,L,+}+S_{(i+s)j}^{V,L,+}-\beta\left[Q_{ijd}^{V,R,+}\right]^+, \tag{B13}$$

and

$$\frac{dQ_{ijd}^{V,R,+}}{dt} = -\gamma_2 Q_{ijd}^{V,R,+} + S_{(i-s)j}^{V,R,+} - \beta\left[Q_{ijd}^{V,L,+}\right]^+ . \tag{B14}$$

Equations (B13) and (B14) are used to draw the phase-plane plot shown in Figure 20a. Equation (B11) implies:

$$0 < \frac{S_{(i-s)j}^{V,R,+}}{\gamma_2} \le \frac{S_{(i+s)j}^{V,L,+}}{\beta} \tag{B15}$$

and

$$0 < \frac{S_{(i+s)j}^{V,L,+}}{\gamma_2} \le \frac{S_{(i-s)j}^{V,R,+}}{\beta} . \tag{B16}$$

From these equations and where the nullclines intersect the axes in Figure 20a, it follows that the nullclines must cross each other at a point where

$$0 \le Q_{ijd}^{V,L,+}, Q_{ijd}^{V,R,+} . \tag{B17}$$

This allows us to remove the rectification in (B13) and (B14) which in turn allows us to perform local analysis on the linear system

$$J = \begin{bmatrix} -\gamma_2 & -\beta \\ -\beta & -\gamma_2 \end{bmatrix}. \tag{B18}$$

The eigenvalues are $-\gamma_2 \pm \beta$. By (B6), $\beta < \gamma_2$, and so both eigenvalues are negative. Thus the crossing of the nullclines represents a unique equilibrium point to which the system exponentially converges. This equilibrium point can be found by adding (B13) and (B14) and recalling that for this case $Q_{ijd}^{V,L,+}$ and $Q_{ijd}^{V,R,+}$ are nonnegative:

$$\frac{d}{dt}\left(Q_{ijd}^{V,L,+} + Q_{ijd}^{V,R,+}\right) = -(\gamma_2 + \beta)\left(Q_{ijd}^{V,L,+} + Q_{ijd}^{V,R,+}\right) + \left(S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,+}\right). \tag{B19}$$

The equilibrium point is:

$$Q_{ijd}^{V,L,+} + Q_{ijd}^{V,R,+} = \frac{1}{\gamma_2 + \beta}\left(S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,+}\right). \tag{B20}$$

By (B12) and (B20), and recalling for this case $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+}$, we see that at large times (B1) is approximated by:

$$\frac{dB_{ijd}^{V,B,+}}{dt} = -\gamma_1 B_{ijd}^{V,B,+} + \left(1 - \frac{\alpha}{\gamma_2 + \beta}\right)\left(S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,+}\right). \tag{B21}$$

Equation (B21) then exponentially converges to:

$$B_{ijd}^{V,B,+} = \frac{1}{\gamma_1}\left(1 - \frac{\alpha}{\gamma_2 + \beta}\right)\left(S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,+}\right). \tag{B22}$$

Case 2: $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+};\ S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,-} < 0$ and $\dfrac{S_{(i-s)j}^{V,R,+}}{S_{(i+s)j}^{V,L,+}} < \dfrac{\beta}{\gamma_2}$. (B23)

First note that (B12)-(B14) apply to this case, allowing us to draw the phase-plane plot shown in Figure 20b. From where the nullclines intersect the axes and from (B23), we see that the nullclines must cross each other at a point where

$$Q_{ijd}^{V,R,+} < 0 \tag{B24}$$

and

$$Q_{ijd}^{V,L,+} > 0 . \tag{B25}$$

Equations (B24) and (B25) imply that (B13) and (B14) can be rewritten as:

$$\frac{dQ_{ijd}^{V,L,+}}{dt} = -\gamma_2 Q_{ijd}^{V,L,+} + S_{(i+s)j}^{V,L,+} \tag{B26}$$

and

$$\frac{dQ_{ijd}^{V,R,+}}{dt} = -\gamma_2 Q_{ijd}^{V,R,+} + S_{(i-s)j}^{V,R,+} - \beta Q_{ijd}^{V,L,+}. \tag{B27}$$

Linear analysis of

$$J = \begin{bmatrix} -\gamma_2 & 0 \\ -\beta & -\gamma_2 \end{bmatrix} \tag{B28}$$

yields only a single eigenvalue, $-\gamma_2$. Because this eigenvalue is negative, the intersection of nullclines represents an equilibrium point to which the system converges. Solving (B26) at equilibrium yields:

$$Q_{ijd}^{V,L,+} = \frac{S_{(i+s)j}^{V,L,+}}{\gamma_2}. \tag{B29}$$

By (B12), (B24) and (B29), and recalling that for this case $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+}$, it follows that at large times, (B1) is approximated by:

$$\frac{dB_{ijd}^{V,B,+}}{dt} = -\gamma_1 B_{ijd}^{V,B,+} + \left(1 - \frac{\alpha}{\gamma_2}\right) S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,+}, \tag{B30}$$

which has the equilibrium solution:

$$B_{ijd}^{V,B,+} = \frac{1}{\gamma_1}\left(\left(1 - \frac{\alpha}{\gamma_2}\right) S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,+}\right). \tag{B31}$$

Case 3: $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+}$; $0 < S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,-}$ and $\frac{S_{(i+s)j}^{V,L,+}}{S_{(i-s)j}^{V,R,+}} < \frac{\beta}{\gamma_2}$. (B32)

By symmetry with Case 2:

$$B_{ijd}^{V,B,+} = \frac{1}{\gamma_1}\left(S_{(i+s)j}^{V,L,+} + \left(1 - \frac{\alpha}{\gamma_2}\right) S_{(i-s)j}^{V,R,+}\right). \tag{B33}$$

Case 4: $S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+} < 0$ and $S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,-} < 0$. (B34)

From (B1) we see that at equilibrium:

$$B_{ijd}^{V,B,+} \le 0. \tag{B35}$$

Case 5: $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,-}$; $S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,+} < 0$; $\frac{\beta}{\gamma_2} \le \frac{S_{(i+s)j}^{V,L,+}}{S_{(i-s)j}^{V,R,-}}$; and $\frac{\beta}{\gamma_2} \le \frac{S_{(i-s)j}^{V,R,-}}{S_{(i+s)j}^{V,L,+}}$. (B36)

By analogy with Case 1, in particular (B12) and (B20), at equilibrium:

$$Q_{ijd}^{V,L,-}, Q_{ijd}^{V,R,+} \le 0 \tag{B37}$$

and

$$Q_{ijd}^{V,L,+} + Q_{ijd}^{V,R,-} = \frac{1}{\gamma_2 + \beta}\left(S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,-}\right). \tag{B38}$$

Using (B37) and (B38), and recalling that for this case $0 < S_{(i+s)j}^{V,L,+}$ and $S_{(i-s)j}^{V,R,+} < 0$, we see that at large times (B1) is approximated by:

$$\frac{dB_{ijd}^{V,B,+}}{dt} = -\gamma_1 B_{ijd}^{V,B,+} + S_{(i+s)j}^{V,L,+} - \frac{\alpha}{\gamma_2+\beta}\left(S_{(i+s)j}^{V,L,+} + S_{(i-s)j}^{V,R,-}\right). \tag{B39}$$

Its equilibrium solution is:

$$B_{ijd}^{V,B,+} = \frac{1}{\gamma_1}\left(\left(1-\frac{\alpha}{\gamma_2+\beta}\right)S_{(i+s)j}^{V,L,+} - \frac{\alpha}{\gamma_2+\beta}S_{(i-s)j}^{V,R,-}\right). \tag{B40}$$

Recalling that for this case $\frac{\beta}{\gamma_2}S_{(i+s)j}^{V,L,+} \le S_{(i-s)j}^{V,R,-}$, we convert (B40) into an inequality:

$$B_{ijd}^{V,B,+} \le \frac{1}{\gamma_1}\left(\left(1-\frac{\alpha}{\gamma_2+\beta}\right)S_{(i+s)j}^{V,L,+} - \frac{\alpha}{\gamma_2+\beta}\frac{\beta}{\gamma_2}S_{(i+s)j}^{V,L,+}\right). \tag{B41}$$

From (B6) we see that $\gamma_2 < \alpha$. This allows (B41) to be rewritten

$$B_{ijd}^{V,B,+} < \frac{1}{\gamma_1}\left(\left(1-\frac{\gamma_2}{\gamma_2+\beta}\right)S_{(i+s)j}^{V,L,+} - \frac{\gamma_2}{\gamma_2+\beta}\frac{\beta}{\gamma_2}S_{(i+s)j}^{V,L,+}\right). \tag{B42}$$

After factorization and cancellation we find:

$$B_{ijd}^{V,B,+} < 0. \tag{B43}$$

Case 6: $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,-}$; $S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,+} < 0$; and $\frac{S_{(i-s)j}^{V,R,-}}{S_{(i+s)j}^{V,L,+}} < \frac{\beta}{\gamma_2}$. (B44)

By analogy with Cases 1 and 2, in particular (B12), (B24) and (B29), at equilibrium:

$$Q_{ijd}^{V,L,-}, Q_{ijd}^{V,R,+} \le 0, \tag{B45}$$

$$Q_{ijd}^{V,R,-} < 0, \tag{B46}$$

and

$$Q_{ijd}^{V,L,+} = \frac{S_{(i+s)j}^{V,L,+}}{\gamma_2}. \tag{B47}$$

Using these equations, and recalling that for this case $0 < S_{(i+s)j}^{V,L,+}$ and $S_{(i-s)j}^{V,R,+} < 0$, we see that at large times (B1) is approximated by:

$$\frac{dB_{ijd}^{V,B,+}}{dt} = -\gamma_1 B_{ijd}^{V,B,+} + S_{(i+s)j}^{V,L,+} - \frac{\alpha}{\gamma_2}S_{(i+s)j}^{V,L,+}. \tag{B48}$$

This converges to:

$$B_{ijd}^{V,B,+} = \frac{1}{\gamma_1}\left(1-\frac{\alpha}{\gamma_2}\right)S_{(i+s)j}^{V,L,+}. \tag{B49}$$

From (B6) we see $\gamma_2 < \alpha$. Thus (B49) implies

$$B_{ijd}^{V,B,+} < 0. \tag{B50}$$

Case 7: $0 < S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,-}$; $S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,+} < 0$; and $\frac{S_{(i+s)j}^{V,L,+}}{S_{(i-s)j}^{V,R,-}} < \frac{\beta}{\gamma_2}$. (B51)

By analogy with Case 2, in particular (B12), (B24) and (B29), at equilibrium:

$$Q_{ijd}^{V,L,-}, Q_{ijd}^{V,R,+} \le 0, \tag{B52}$$

$$Q_{ijd}^{V,L,+} < 0, \tag{B53}$$

and

$$Q_{ijd}^{V,R,-} = \frac{S_{(i-s)j}^{V,R,-}}{\gamma_2}. \tag{B54}$$

Using these equations, and recalling that for this case $0 < S_{(i+s)j}^{V,L,+}$ and $S_{(i-s)j}^{V,R,+} < 0$, we see that at large times (B1) is approximated by:

$$\frac{dB_{ijd}^{V,B,+}}{dt} = -\gamma_1 B_{ijd}^{V,B,+} + S_{(i+s)j}^{V,L,+} - \frac{\alpha}{\gamma_2} S_{(i-s)j}^{V,R,-}. \tag{B55}$$

This converges to:

$$B_{ijd}^{V,B,+} = \frac{1}{\gamma_1}\left(S_{(i+s)j}^{V,L,+} - \frac{\alpha}{\gamma_2} S_{(i+s)j}^{V,R,-} \right). \tag{B56}$$

Recalling that for this case $\frac{\gamma_2}{\beta} S_{(i+s)j}^{V,L,+} < S_{(i-s)j}^{V,R,-}$, we convert (B56) into an inequality:

$$B_{ijd}^{V,B,+} < \frac{1}{\gamma_1}\left(1 - \frac{\alpha}{\beta}\right) S_{(i+s)j}^{V,L,+}. \tag{B57}$$

From (B6) we see that $\beta < \alpha$. Thus (B57) implies:

$$B_{ijd}^{V,B,+} < 0. \tag{B58}$$

Case 8: $S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,-} < 0$ and $0 < S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,+}$ (B59)

By analogy with Cases 5-7:

$$B_{ijd}^{V,B,+} < 0. \tag{B60}$$

Case 9: $S_{(i+s)j}^{V,L,+}, S_{(i-s)j}^{V,R,+}, S_{(i+s)j}^{V,L,-}, S_{(i-s)j}^{V,R,-} = 0$. (B61)

By inspection, (B2), (B3), (B4) and (B5) imply that at equilibrium,

$$Q_{ijd}^{V,L,+}, Q_{ijd}^{V,R,+}, Q_{ijd}^{V,L,-}, Q_{ijd}^{V,R,-} = 0. \tag{B62}$$

By (B62), (B1) implies that at equilibrium,

$$B_{ijd}^{V,B,+} = 0. \tag{B63}$$

As we have now considered all possible cases, the theorem is proved.

References

Anzai, A., Bearse, M. A., Freeman, R. D., & Cai, D. (1995). Contrast coding by cells in the cat's striate cortex: Monocular vs. binocular detection. *Visual Neuroscience*, *12,* 77-93.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86,* 2054-2068.

Bakin, J. S., Nakayama, K., & Gilbert, C. D. (2000). Visual responses in monkey area V1 and V2 to three-dimensional surface configurations. *The Journal of Neuroscience*, *20,* 8188-8198.

Callaway, E. M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annual Review of Neuroscience*, *21,* 47-74.

Callaway, E. M., & Wiser, A. K. (1996). Contributions of individual layer 2-5 spiny neurons to local circuits in macaque primary visual cortex. *Visual Neuroscience, 13,* 907-922.

Cogan, A. I., Kontsevich, L. L., Lomakin, A. J., Halpern, D. L., & Blake, R. (1995). Binocular disparity processing with opposite-contrast stimuli. *Perception*, *24,* 33–47.

Cumming, B. G. (2002). Receptive field structure and disparity tuning in primate V1. *Vision ScienceS Society (abstracts)*, 288.

Cumming, B. G., & Parker, A. J. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *The Journal of Neuroscience*, *20,* 4758-4767.

Dow, B. M. (1974). Function classes of cells and their laminar distribution in monkey visual cortex. *Journal of Neurophysiology*, *37,* 927-946.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research*, *36,* 1839-1857.

Frisby, J. P. (2001). Limited understanding of Panum's limiting case. *Perception*, *30,* 1151-1152.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., & Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106,* 795-834.

Gillam, B., Blackburn, S., & Cook, M. (1995). Panum's limiting case: double fusion, convergence error, or 'da Vinci stereopsis'. *Perception*, *24,* 333-346.

Gillam, B., Blackburn, S., & Nakayama, K. (1999). Stereopsis based on monocular gaps: Metrical encoding of depth and slant without matching contours. *Vision Research*, *39,* 493-502.

Grimson, W. E. (1981). A computer implementation of a theory of human stereo vision. *Philosophical Transactions of the Royal Society (B), 292,* 217-253.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87*, 1-51.

Grossberg, S. (1987). Cortical dynamics of three-dimensional form, color, and brightness perception: II. Binocular theory. *Perception & Psychophysics*, *41,* 117-158.

Grossberg, S. (1994). 3D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, *55*, 48-120.

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional figures. *Psychological Review*, *104,* 618-658.

Grossberg, S. (1999a). How does the cerebral cortex work? Learning, attention and grouping by the laminar circuits of visual cortex, *Spatial Vision, 12*, 163-186.

Grossberg, S. (1999b). The link between brain learning, attention, and consciousness. *Consciousness and Cognition, 8,* 1-44.

Grossberg, S., Hwang, S., and Mingolla, E. (2002). Thalamocortical dynamics of the McCollough effect: Boundary-surface alignment through perceptual learning. *Vision Research,* in press.

Grossberg, S., & Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research, 39,* 3796-3816

Grossberg, S., & McLoughlin, N. (1997). Cortical dynamics of three-dimensional surface perception: Binocular and half-occluded scenic images. *Neural Networks*, *10,* 1583- 1605.

Grossberg, S. and Mingolla, E. (1985). Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations. *Perception and Psychophysics, 38,* 141-147.

Grossberg, S., Mingolla, E., & Ross, W. D. (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neuroscience, 20,* 106-111.

Grossberg, S., Mingolla, E., & Williamson, J. (1995). Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. *Neural Networks, 8,* 1005-1028.

Grossberg, S., & Pessoa, L. (1998). Texture segregation, surface representation, and figure-ground separation *Vision Research, 38,* 2657-2684.

Grossberg, S., & Raizada, R. D.. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research, 40,* 1413-1432.

Grossberg, S., & Todorović D. (1988). Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. *Perception and Psychophysics, 43,* 241-277.

Grossberg, S., & Williamson, J.R. (2001). A neural model of how horizontal and interlaminar connections of visual cortex develop into adult circuits that carry out perceptual groupings and learning. *Cerebral Cortex, 11,* 37-58.

Howard, I. P., & Rogers, B. J. (1995). *Binocular Vision and Stereopsis.* New York: Oxford University Press.

Howe, P. D. L., & Grossberg, S. (2001). Laminar cortical circuits for stereopsis and surface depth perception. *Society for Neuroscience Abstracts*, 164.17.

Howe, P. D. L., & Watanabe, T. (2002). A new test of the disparity energy model of stereopsis. Submitted for publication.

Hubel, D. H., & Livingstone, M. S. (1987). Segregation of form, color, and stereopsis in primate area 18. *The Journal of Neuroscience*, *7,* 3378-3415.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195,* 215-243.

Julesz, B. (1971). *Foundations of Cyclopean Perception.* Chicago: The University of Chicago Press.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science.* Chicago: University of Chicago Press (4th Ed.).

Katz, L. C., Gilbert, C. D., & Wiesel, T. N. (1989). Local circuits and ocular dominance columns in monkey striate cortex. *The Journal of Neuroscience, 9,* 1389-1399.

Kelly, F. J., & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics, 62*, 1596-1619.

Krol, J. D., & van de Grind, W. A. (1983). Depth from dichoptic edges depends on vergence tuning. *Perception, 12,* 425-438.

Livingstone, M. S., & Hubel, S. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *The Journal of Neuroscience*, *4,* 309-356.

Marr, D., & Poggio, T. (1976). Cooperative computation of stereo disparity. *Science, 194,* 283-287.

McKee, S. P., Bravo, M. J., Smallman, H. S., & Legge, G. E. (1995). The 'uniqueness constraint' and binocular masking. *Perception*, *24,* 49-65.

McKee, S. P., Bravo, M. J., Taylor, D. G., & Legge, G. E. (1994). Stereo matching precedes dichoptic masking. *Vision Research*, *34,* 1047-1060.

McLoughlin, N., & Grossberg (1998). Cortical computation of stereo disparity. *Vision Research*, *38,* 91-99.

Merigan, W. H., & Maunsell, J. H. R. (1993). How parallel are the primate visual pathways. *Annual Review of Neuroscience, 16,* 369-402.

Nakamura, K., & Colby, C. L. (2000a). Visual, saccade-related, and cognitive activation of single neurons in monkey extrastriate area V3A. *Journal of Neurophysiology, 84,* 677-692.

Nakamura, K., & Colby, C. L. (2000b). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences, 99,* 4026-4031.

Nakamura, H., Kuroda, T., Wakita, M., Kusunoki, M., Kato, A., Mikami, A., Sakata, H., & Itoh, K. (2001). From three-dimensional space vision to prehensile hand movements: The lateral intraparietal area links the area V3A and the anterior intraparietal area in macaque. *The Journal of Neuroscience*, *21*, 8174-8187.

Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Research*, *30,* 1811-1825.

Ohzawa, I. (1998). Mechanisms of stereoscopic vision: The disparity energy model. *Current Opinion in Neurobiology*, *8,* 509-515.

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.

Panum, P. L. (1858). *Physiologische Untersuchungen ueber das Sehen mit zwei Augen. Kiel:* Schwerssche Buchhandlung), translated by C Hubscher 1940 (Hanover, NH: Dartmouth Eye Institute).

Pessoa, L., Mingolla, E., & Neumann, H. (1995). A contrast- and luminance-driven multiscale network model of brightness perception. *Vision Research*, 35, 2201-2223.

Peterhans, E. (1997). Functional organization of area V2 in the awake monkey. *Cerebral Cortex*, *12,* 335-357.

Poggio, G. F. (1972). Spatial properties of neurons in striate cortex of unanesthetized macaque monkey. *Investigative Ophthalmology*, *11,* 369-377.

Poggio, G. F. (1991). Physiological basis of stereoscopic vision. In: *Vision and Visual Dysfunction. Binocular vision* (pp. 224-238). Boston, MA: CRC.

Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, *40*, 1392-1405.

Poggio, G. F. & Talbot, W. H. (1981). Mechanisms of static and dynamic stereopsis in foveal cortex of the rhesus monkey. *Journal of Physiology, 315*, 469-492.

Pollen, D. A., Przybyszewshi, A. W., Rubin, M. A. and Foote, W. (2002). Spatial receptive field organization of macaque V4 neurons. *Cerebral Cortex, 12*, 601-616.

Raizada, R., & Grossberg, S. (2001). Context-sensitive bindings by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, in press.

Read, J. C. A, Cumming, B. C., & Parker, A. J. (2002). Simple cells can show non-linear binocular combination. *Vision ScienceS Society (abstract)*, 287.

Rockland, K. S., & Virga, A. (1990). Organization of individual cortical axons projecting from area V1 (area 17) to V2 (area 18) in the macaque monkey. *Visual Neuroscience*, 4, 11-28.

Roe, A. W., & Ts'o, D. Y. (1995). Visual topography in primate V2: Multiple representation across functional stripes. *The Journal of Neuroscience*, *15,* 3689-3715.

Roe, A. W., & Ts'o, D. Y. (1997). The functional architecture of area V2 in the macaque monkey. *Cerebral Cortex*, *12,* 295-333.

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *Journal of Neurophysiology*, *39,* 1288-1319.

Schiller, P. H., Logothetis, N. K., & Charles, E. R. (1990a). Role of the color-opponent and broad-band channels in vision. *Visual Neuroscience*, *5,* 321-346.

Schiller, P. H., Logothetis, N. K., & Charles, E. R. (1990b). Functions of the colour-opponent and broad-band channels of the visual system. *Nature*, *343,* 68-70.

Schiller, P. H. and Malpeli, J. G. (1978). Functional specificity of lateral geniculate nucleus laminae of the rhesus monkey. *Journal of Neurophysiology, 41,* 788-797.

Schor, C. and Heckmann, T. (1989). Interocular differences in contrast and spatial frequency: effects on stereopsis and fusion, *Vision Research, 29,* 837-847.

Smallman, H. S., & McKee, S. P. (1995). A contrast ratio constraint on stereo matching. *Proceedings of the Royal Society of London B*, *260,* 265-271.

Smith, E. L., Chino, Y., Ni, J., & Cheng, H. (1997). Binocular combination of contrast signals by striate cortical neurons in the monkey. *Journal of Neurophysiology, 78,* 366-382.

Swaminathan, G., & Grossberg, S. (2001). Laminar cortical circuits for the perception of slanted and curved 3D surfaces. *Society for Neuroscience Abstracts*, 619.49

Tootell, R. B. H., Mendola, J. D., Hadjikhani, N. K., Ledden, P. J., Liu, A. K., Reppas, J. B, Sereno, M. I., & Dale, A. M. (1997). Functional analysis of V3A and related areas in human visual cortex. *The Journal of Neuroscience*, *17,* 7060-7078.

Tsao, D. Y., & Livingstone, M. S. (In Press). Spatiotemporal maps of disparity-selective simple cells in macaque V1. *Neuron*

von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, *40,* 1955-1967.

Wallach, H. (1976). *On Perception* (pp. 8). New York: Quadrangle/The New York Times Book Co.

Wang, Z. , Wu, X., Ni, R., & Wang, Y. (2001). Double fusion does not occur in Panum's limiting case: evidence from orientation disparity. *Perception*, *30*, 1143-1149.

Xiao, Y., Zych, A., & Felleman, D. J. (1999). Segregation and convergence of functionally defined V2 thin stripe and interstripe compartment projections to area V4 of macaques. *Cerebral Cortex, 9,* 792-804.