Border ownership from intracortical interactions in visual area V2

Li Zhaoping
Department of Psychology
University College London
London, United Kingdom
`z.li@ucl.ac.uk`, http://www.gatsby.ucl.ac.uk/∼zhaoping

Summary:
A border between two image regions normally belongs to only one of the regions; determining which one is essential for surface perception and figure-ground segmentation. Border ownership is signaled by a class of V2 neurons, even though its value depends on information coming from well outside their classical receptive fields. I use a model of V2 to show that this visual area is able to generate the ownership signal by itself, without requiring any top-down mechanism or external explicit labels for figures, T junctions or corners. In the model, neurons have spatially local classical receptive fields, are tuned to orientation, and receive information (from V1) about the location and orientation of borders. Border ownership signals that model physiological observations arise through finite-range, intra-areal interactions. Additional effects from surface features and attention are discussed. The model licenses testable predictions.

Running title: Border ownership by V2 mechanisms

1   Introduction

Separating figure from background in an image is one of the most important tasks in vision, and is commonly seen as a pre-requisite to object recognition. However, we still do not understand its neural basis. Figure and ground information in an image can be represented by assigning ownership of the border between two surfaces as belonging to one of the surfaces as the figure, which then occludes the other, the ground. The border is said to be owned by the assigned or occluding figure. The shapes of the surfaces and the resulting perception depend crucially on this allocation of border ownership (BOWN), as is well demonstrated by the famous example in Figure (1A)(Rubin 2001). Here, the bistable perception of either the flower vase or the mirrored, profile faces, depends on whether the contour in the image is interpreted as owned by the lighter or darker regions.

Many neurons in V2 have been observed to be selective to both the orientation and BOWN of contours (Zhou, Friedmann and von der Heydt 2000). That is, for instance, a cell preferring vertical contours responds more vigorously to a vertical contour segment within its classical receptive field (CRF) if this segment belongs to the surface to the left rather than the right side of this contour.

For example, the white square of Fig. 1B is more readily seen as a figure occluding a grey background than as a hole in a grey surface revealing an underlying white background. In BOWN terms, the bounding contour of the square is more readily owned by the white. In contrast, Fig. (1F) is generally interpreted as a grey square in a white

background. By construction, the left border of this square is at the same spatial location as the right border of the white square in Fig. (1B). Consider the CRF of a model cell shown by the ellipses in the figures. It is centered on this border location, but is much smaller than the whole border of the squares. If this cell is tuned to BOWN, its response to stimulus Fig. (1B) and Fig. (1F) should differ significantly, even though the stimulus patterns *within* its CRF are exactly the same for the two examples. It is the stimulus pattern in the context *outside* the CRF that determines the BOWNs and the neural responses.
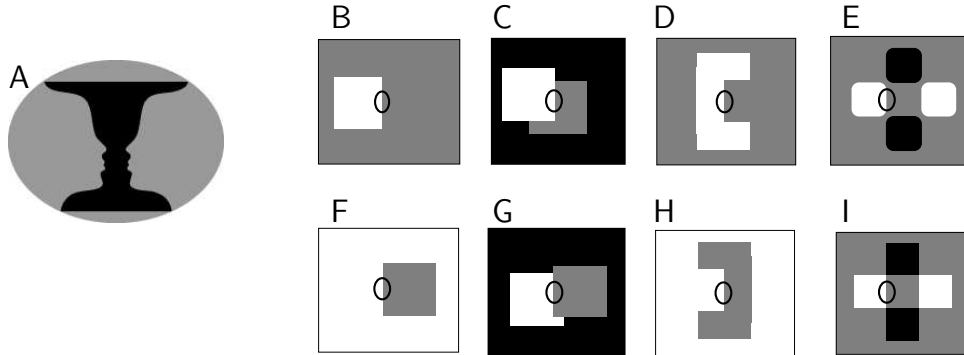

Figure 1: Border ownership examples. A: the bistable perception of flower vase or two faces depends on whether the borders between the luminance regions is assigned to the lighter or darker regions. B-I: schematics of stimulus patterns similar to those used in experiments by von der Heydt and colleagues. The oval depicts a classical receptive field (CRF) of an orientation tuned cell in V2, and is not part of the stimulus. In all these patterns, the stimulus within the CRF is the same, but the border within the CRF belongs to the figure to its left in the top row stimulus (B-E), and to the figure to its right in the bottom row (F-I). If the cell is tuned to prefer a border owned by a figure to the left of its CRF, then its responses to the top row stimuli will be higher than those to the bottom row, as observed in V2 cells.

Figure (1) shows that the same white-left to dark-right luminance contrast stimulus within the CRF can be generated by other Gestalt-based stimulus configurations and BOWN examples: by partial occlusion between white and gray surfaces (Fig. (1C) and (1G)); by a white or gray C-shaped figure in front of gray or white backgrounds (Fig. (1D) and (1H)); by a white semi-transparent rectangle on top of a black one (Fig. (1I)); and by four non-overlapping black or white figures on a gray background (Fig. (1E)). The stimulus patterns in Fig. 1B-I are similar to those used by von der Heydt and colleagues to demonstrate BOWN tuning in V2 neurons. They observed that a cell tuned to a figure to the left prefers Fig. (1BCDE) to Fig. (1FGHI) (Zhou, Friedman and von der Heydt 2000, Qiu, Macuda, and von der Heydt 2001, von der Heydt, Qiu, and He 2003, von der Heydt, Zhou, and Friedman 2000). As the size of figure square in Fig. 1B grows toward occupying the whole left half of the image, BOWN becomes ambiguous. Correspondingly, the BOWN signal in V2 neurons also weaken (Zhou et al 2000), i.e., the differences in responses to Fig.1B and Fig. 1F are reduced.

Von der Heydt and colleagues also observed that neural tunings to BOWN is not tightly constrained by, nor dependent on, any particular image cue such as surface luminance. BOWN tuning was also observed for figures defined by depth using random-dot stereograms (Qiu et al 2001), motion (von der Heydt et al 2003), or even for figures in line drawings (e.g., a square line drawing, Zhou et al 2000). Furthermore, BOWN tuning observed when luminance is the cue is consistent with that observed when disparity (Heydt et al 2000, Heydt 2003) or motion (von der Heydt et al 2003) is the cue, if a cell is tuned to BOWN using more than one cue. Furthermore, while surface feature differences between two neighboring surfaces can be qualitatively signaled by the feature contrast polarity of

the border, not all cells tuned to BOWN of a contour are tuned to luminance contrast polarity and vice versa (Zhou et al 2000), and many cells are tuned to BOWN irrespective of other image cues such as luminance contrast (von der Heydt 2003). Fig. 1G demonstrates that the gray figure is perceived as a whole non-occluded surface even though the contrast polarity changes along its border due to changes of the occluded region (see also Discussion and Fig. 5A). This implies that the contour information overrides the contrast polarity information to determine BOWN.

Tuning in BOWN is significantly weaker or even absent in V1 (Zhou et al 2000). Thus the question arises whether the context-dependent neural tuning to BOWN in V2 is generated by mechanisms within this area, or by top-down feedback from one or more higher visual areas, or by a combination of both top-down and local mechanisms (von der Heydt 2003). The insightful psychophysical observations of Nakayama, He and Shimojo (1995) weigh against higher areas. First, they demonstrated that manipulating stimuli to change perceived BOWNs dramatically affects behavior in rapid visual tasks like visual search. Second, in the case of "impossible objects", such as the famous Penrose "impossible" triangle, BOWN processing leads to perceptions of three-dimensional objects that can never be realized in real world, and that must therefore violate higher level object knowledge. A possible neural substrate for a local mechanism to process contextual information certainly exists in the form of the intra-cortical, lateral, neural connections within V2 that link cells with non-overlapping CRFs separated by a finite distance (Angelucci, Levitt, Walton, Hupe, Bullier, and Lund 2002).

Other evidences support contributions beyond V2. The first one comes from data on the latencies of BOWN signals, defined by the times (since initial responses) at which neural responses first differ between inputs with identical CRF stimulation, but different, contextually-defined, BOWN. In V2, BOWN latency does not seem to depend on the sizes of the figures such as the square in Fig. 1B. The essential contextual information needed to determine the ownership of a contour segment on a side of a square come from the ends of the side or other sides of the square. If finite range intra-cortical connections were used to propagate the contextual information, then larger squares would seem to require longer propagation times, thus requiring longer latencies. That this is not true for squares of sizes $3°$ and $8°$ has prompted the suggestion that central (global) feedbacks from visual areas with larger receptive fields are involved (Sugihara, Qiu, and von der Heydt 2003). Secondly, the bistable perception of the flower vase or faces can be influenced by voluntary attention, suggesting top-down modulation. Models have duly been constructed in which BOWN tuning results from a combination of V2 mechanisms and additional, presumably top-down, information such as labels for global figures, T junctions, and corners (unless one views that T junctions are corners can be detected by the end-stopped cells in V1 or V2, Heitger et al 1992, Heider et al 2000), all of which are very helpful to determine BOWN (Finkel and Sajda 1994, Craft, Schutze, Niebur, and von der Heydt 2004).

Understanding the capacity of local mechanisms in V2 to generate BOWN, particularly with size-independent latency, is therefore a key first step. In this paper, I describe a network model of V2 in which generic local and contextual contour (orientation) information (independent of the input cues responsible for the contours themselves) can feasibly be used by local V2 mechanisms to construct BOWN tuning consistent with all the physiological observations so far described. The additional contributions to BOWN tuning from surface feature (e.g., luminance) cues and top down factors, will be discussed in the Discussion, along with the relationship to previous work.

2 Results

Model neurons have CRFs and preferred orientations that sample the visual space evenly. For each (discrete) CRF location i and preferred orientation $0 \leq \theta < 180°$, there are two neurons with opposite preferred BOWNs; one indicating that the figure owning the contour segment at (i, θ) is to the *left* of the contour and the other that it is to the *right* (when one faces the direction θ). To represent BOWN graphically, we adopt a convention in which we show neurons with $360°$ tuning to direction rather than $180°$, with values differing by $180°$ favoring the *same* orientation, but opposite BOWN. Thus, each neuron is shown as being tuned to direction θ spanning 0 to $360°$, where $\theta = 0°$ is

3

the 3 o'clock (East) direction and θ increases counter-clock wise. In this representation, a neuron tuned to direction θ prefers orientation θ if $\theta < 180°$, and $\theta - 180°$ otherwise, while its preferred figure side is always to the *right* of the contour segment (i, θ) when facing direction θ.

A Visualizing model border elements and activities

a, A V2 neuron: prefers vertical, with figure to the right.

b, A V1 neuron input to two V2 neurons preferring opposite ownerships

c, An input pattern from V1

d, A V2 output pattern

B Examples of mutual excitation (a, b, c) and inhibition (d, e, f) between segments

a, **b,** **c,**

d, **e,** **f,**

C Example connections between cells receiving an input pattern (black contour segments) Length×width of the colored fins of the pre-synaptic cells ∝ connection strengths.

Excitatory connections

Post-synaptic

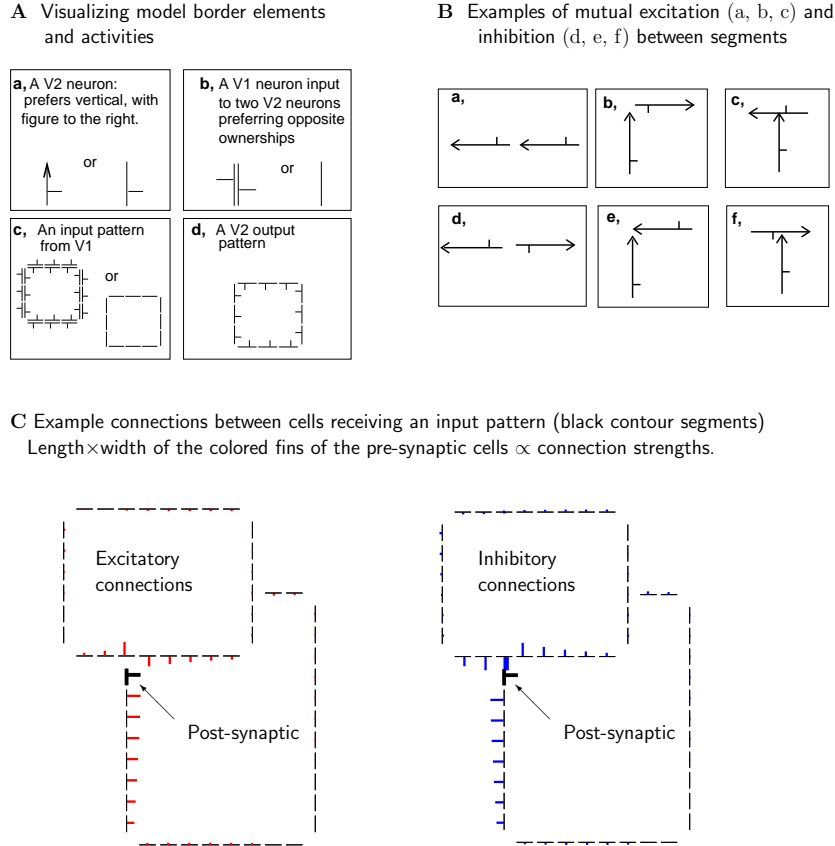Inhibitory connections

Post-synaptic

Figure 2: Visualizing the model elements. A: visualizing neural elements and activity patterns. a) A V2 neuron visualized by its preferred features: the CRF position, contour orientation, and preferred side of the owner figure as indicated by a fin, or equivalently, the direction of the border segment. b): input from a V1 neuron is directed equally to two V2 neurons (with the same CRF position and orientation preference) preferring two opposite BOWNs. c): an input image from V1 with no BOWN biases, and d) desired V2 responses (outputs) with BOWN biases. B: Examples of mutual excitation (via connections J in equation (1)) and mutual inhibition (via connections W in equation (2)) between pairs of model neural elements coding border segments (of specific orientations, BOWNs, and CRF positions) as shown. Apart from (c) and (f), mutual excitations (inhibition) are more likely and stronger between neurons signaling border segments that are more consistent (inconsistent) with belonging to a single figure. The strength of excitation decreases from a) to b) (see C). Mutual excitation between neurons is stronger when the corresponding two border segments could be linked by a rightward rather than a leftward turn from one segment to another (also see C), reflecting the *Gestalt* structure of the visual world for which object surfaces tend to be convex rather than concave. (c) and (f) are connections when the two border segments are consistent and inconsistent, respectively, with a T junction. C: The actual model connections between cells receiving inputs from an example occlusion pattern (thin black contour segments). The thick border segment mark the example post-synaptic cell. The connection strengths from the pre-synaptic cells scale with the length×width of the colored fins of the border segments.

4

Fig. 2 shows the model elements and the way we will visualize them. A neuron is represented by a directed edge (plotted with or without an arrow) at its CRF location i and pointing in its preferred direction θ. This edge has a fin directed towards the side of the figure the cell assumes owns the border (Fig. 2Aa). As mentioned, this is always the right side when facing θ. Thus a cell preferring a vertical orientation, $\theta = 90°$, could have preferred direction either $90°$ or $270°$, depending on whether it prefers a figure to the east or west of the border.

Model inputs are assumed to arise from population activity of topographically localized and orientation tuned V1 output neurons, representing images of contour segments (see Fig. (2Ac)), i.e., short oriented bars. Each input signal $I_{i\theta}$ is a graded firing rate representing the strength (e.g., contrast) of a contour segment (i, θ). However, to answer whether generic contour information is sufficient to determine BOWNs through context, the model inputs are abstracted as not specifying whether the underlying contour segments are derived from luminance or disparity cues, or are simply contour lines (even though the actual V1 does provide some such information to higher visual areas). Most critically, the inputs are not explicitly biased for the BOWN of their contour segments. Hence, unless otherwise stated, equal input strengths $I_{i\theta} = I_{i,\theta+180°}$ are always directed to two model cells (i, θ) and $(i, \theta + 180°)$ preferring the two possible BOWNs given (i, θ).

Fig. 2Ac shows an example model input pattern for a square contour, where $I_{i\theta}$ is non-zero for 3 pairs of directed border segments (i, θ) on each side of the square, and zero for all other border segments. V2 responses have BOWN biases, with unequal response (pyramidal) outputs $O_{i,\theta} \neq O_{i,\theta+180°}$ for border segments (i, θ) and $(i, \theta + 180°)$ of the same orientation and border location but opposite BOWN preference. This is visualized by a bar at location i and orientation θ, with a fin pointing to the dominant figure side. The thickness of the bars (i, θ) is proportional to $\max(O_{i,\theta}, O_{i,\theta+180°})$, the dominant output of the two pyramidals, and the length of the fin increases with $|O_{i,\theta} - O_{i,\theta+180°}| / \max(O_{i,\theta}, O_{i,\theta+180°})$, the fractional difference between the two responses $O_{i,\theta}$ and $O_{i,\theta+180°}$. Fig. 2Ad shows a possible V2 response pattern, preferring a square figure to a hole, to the V1 inputs shown in Fig. 2Ac.

The model neurons not sharing the same CRF location excite or suppress each other through mono-synaptic excitation or di-synaptic inhibition (via interneurons) through intracortical connections that are characteristic of V2 (Angelucci et al 2002). The connections link nearby, and not necessarily overlapping CRFs. Their nature and strengths depend on the CRF locations and preferred orientations and BOWNs of the pre- and post-synaptic cells. Two neurons facilitate or suppress each other's activities if the two corresponding border segments and BOWNs are consistent or inconsistent with being owned by a single figure surface. For example, if two neurons represent two nearby and co-aligned contour segments, they facilitate each other if they prefer the same figure side (Fig. 2Ba) and inhibit each other otherwise (Fig. 2Bd). The connections also satisfy other Gestalt grouping principles. For instance, in the visual world, borders of figures tend to be convex relative to the figure, i.e., object surfaces tend to have rounded or oblique corners rather than spiky or acute corners. In other words, when walking along the border with the figure on the right side, the border tends to turn right towards the figure rather than turning left away from it. This convexity bias is represented in the model connections as stronger facilitation between neurons representing two border segments related to each other through a right turn (Fig. 2Bb, Fig. 2C) than through a left turn. Neurons are connected inhibitorily if they represent border segments that could plausibly be part of two borders colliding

towards each other at some point (Fig. 2Be) or departing from each other at a point (Fig. 2Bd). Furthermore, while the model does not have neurons explicitly detecting T junctions, two neurons for two border segments which are plausibly the top and stem, respectively, of a T junction facilitate each other if the neuron for the T top has a BOWN preference consistent with an occluding surface (Fig. 2Bc), otherwise, suppressive connections are designed between the two neurons (Fig. 2Bf). A consequence of this interaction is that cells should manifest end-stopping properties. Meanwhile, the connection strengths also decay with distances between linked cells. All these Gestalt grouping features of the connection patterns can be visualized in Fig. 2C in an example input pattern. Finally, the connection strengths are such that if there is no visual stimulus within the CRF of a neuron, no single contextual stimulus bar segment is sufficient to evoke any sustained responses from it. Note that the model connection structure requires many links between cells tuned to very different, e.g., orthogonal, orientations. This is in contrast to what is believed to be in V1, where interconnected neurons tend to prefer similar orientations. Indeed, it is easier to observe in V2 than in V1 correlated firings between neurons preferring very different orientations (Tamura et al, 1996). Interactions between nearby, orthgonally oriented elements have also been observed psychophysically (Yu et al., 2002; Popple, 2003).

Since BOWN latency is diagnostic, the model has to address finite-time conduction. A pre-synaptic neuron's output arrives at its post-synaptic target after an axonal conduction latency or delay. The existing data from various experiments still leave much uncertainty about the true range of the conduction latency or velocity (Grinvald, Lieke, Frostig, Hildesheim 1994 Bringuier, Chavane, Glaeser, Fregnac 1999, Girard, Hupe, Bullier 2001, Bair, Cavanaugh, Movshon 2003), and it is possible that the conduction velocity vary with the axon length (see Discussion). I henceforth assume in this paper that the conduction latency between the pre- and post- synaptic cells is randomly in the range of 8-10 milliseconds (ms), (or, for neurons with a membrane time constant of 10ms, between $80 - 100\%$ of the membrane time constant), regardless of the distance between the connected cells (Girard et al 2001, see Discussion section for the data and considerations behind this assumption). Neural signal integration time at the post-synaptic cell is additionally modelled through the membrane time constant in the model neuron. Procedures section lists all the model details and parameters necessary to reproduce the results, and the same model parameters are used in the simulations of all examples in this paper.

Fig. (3) demonstrates that the model responds in accordance with psychophysical observations. The model's initial response (Fig. 3B) reflects the feedforward visual input (Fig. 3A), which contains no BOWN bias, and the random input noise perturbing each model neuron. Later, systematic differential responses to the two possible assignments of figure and ground emerge (Fig. 3C), and grow towards an asymptotic level as the activities evolve (Fig. 3D). Through intracortical interactions, border segments in the context of a given cell's CRF bias its responses, in a way that depends on their preferred BOWNs. The cell's response in turn influences the responses of the contextual segments.

Physiologically (Zhou et al 2000, Sugihara et al 2003), the latency of the BOWN signal is defined by the time at which significant differences appear in the responses of a neuron to two border stimuli with the same local features in the CRF, but opposite BOWNs (such as the border within the oval in Fig. 1B and Fig.1F). In our model, the latency of the BOWN tuning is equivalently defined *using a single stimulus,* as follows. Given one border stimulus

A Model input   **B** Model initial response at t=0.6   **C** Response at t = 2.4   **D** Response at t = 6.3

E Temporal responses for square size = 10 as in A-D   **F** Temporal responses for square size = 20   **G** Temporal responses for square size = 30

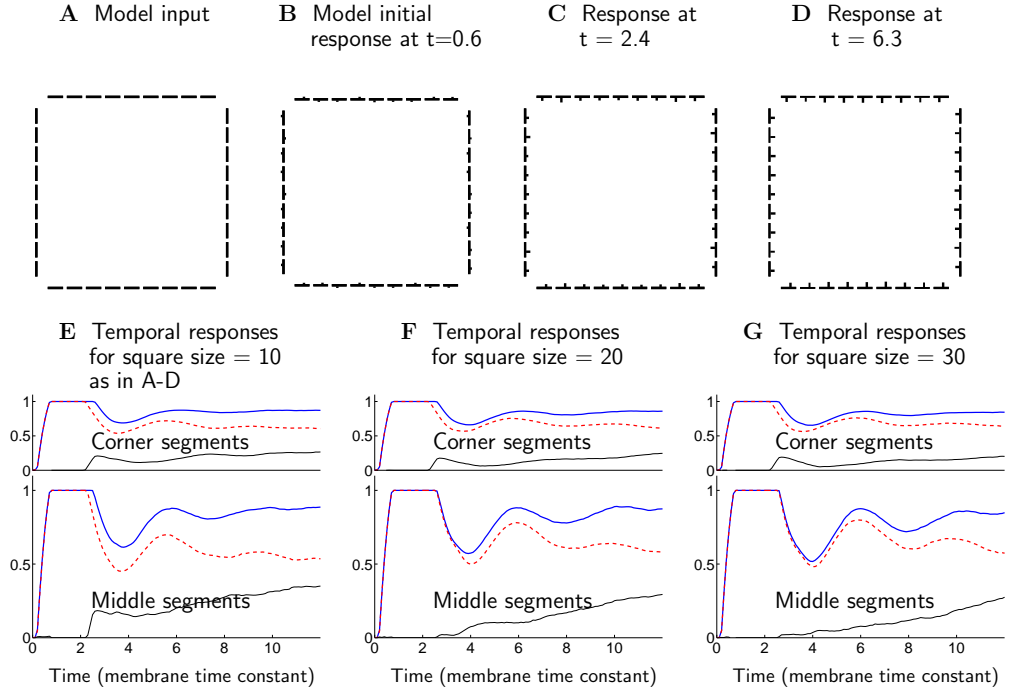Corner segments   Middle segments

Time (membrane time constant)

Figure 3: Model's input-output. Stimulus onset at time = 0, all times are in the units of membrane time constant. A: The input is a square contour frame, with equal strength inputs for both options of BOWN for each contour segment. B: Initial model response at time = 0.6 reflects the feedforward input in A. The BOWN signals in the border segments are weak and inconsistent between border segments, reflecting the random noise added in the model (otherwise no BOWN bias in any border segment would appear before the contextual influences arise.) C : Model responses at time = 2.4 start to consistently favor the choice of the inside of the square as being the figure. D: Model responses at a later stage have stronger BOWN signals. Visualization of responses in A-D is as follows. For each border segment, let O_1 and O_2 be the responses to this segment from two neurons preferring opposite BOWN, and assume $O_1 > O_2$. The thickness of each border bar is plotted roughly proportional to O_1, with the fin pointing to the dominant figure side, and the fin length, increasing with the BOWN bias $O_1 - O_2$, is $0.1 + 0.9(O_1 - O_2)/(O_1 + 0.05)$ as a fraction of the border bar length. E: Time courses of the responses corresponding to B-D. In each subplot, the blue solid curve on top is the response from the neuron(s) prefering the square as the figure, the red dashed curve that of the neuron(s) preferring the opposite BOWN, and the black solid (thin) curve the difference between them, i.e., the BOWN strength. The top subplot is the average response from all (8) border segments at the corners of the squares. The bottom subplot is the average response from four border segments in the middle of the four sides of the square. F,G: same as E, but for larger square sizes, indicated by the length of one side of the square. Note different scales of the vertical axes between the top and bottom subplots, and that responses of all model neurons are bounded between 0 and 1. Note: BOWN latencies are about 2-3 membrane time constant in all square sizes, regardless of whether the border segment is a corner or middle border segment; BOWN signal is weaker in larger squares; and the BOWN signal strength initially increases more quickly in time for the corner than the middle segments in F and G.

pattern , e.g., Fig. 1B, we examine the temporal responses of the two neurons which prefer two opposite BOWNs, but have the same CRF and preferred orientation matching the border, e.g., the oval in Fig. 1B. The latency is defined by the time when the responses in these two neurons start to differ significantly. The strength of the BOWN signal may be seen as the difference between the two responses.

The physiological data show that for square figures, BOWN arises about 30 milliseconds after the initial neuron responses (Sugihara et al 2003), and this latency is the the same for squares of sizes 3° and 8°. From anatomical

data (Angelucci et al 2002), it is reasonable to assume, that $3°$ corresponds roughly to the longest length of lateral connections in V2. In our model, we examine responses to three squares whose sides are 1, 2, and 3 times, as long as the longest intra-cortical connection length (which is 10) in the model (Fig. 3). For these squares, it is instructive to consider a border segment in the middle of a figure side, since this is furthest from the corners of the square, which is the only location in the image where there is any BOWN bias (arising from convexity). For the smallest square, the lateral connections link the middle border segment to the closest corners directly; for the largest square, no corner is within reach. Fig. (3E,F,G) show that the BOWN signal arises roughly with the same latency, between 2-3 membrane time constants (i.e., 20-30 ms if membrane time constant is 10 ms) after stimulus onset or initial responses, for all three square sizes, even though the strengths of the BOWN signals are weaker (in the middle border segments) for larger squares, as observed physiologically (Zhou et al 2000). Fig. (3F,G) show that for larger squares, the BOWN signal increases more quickly with time initially for the corner border segments than the middle border segments, even though the two BOWN latencies are comparable.

These simulation results suggest that top-down feedback is not necessary to create size invariant latencies, at least within a range of figure sizes. This may seem counter-intuitive, given the finite lengths of connections and the non-zero axonal conduction delays. However, note that the BOWN latency in the model is in any case about 2-3 times the axonal conduction delay and the membrane time constant. Within this temporal window, the contextual information at the corners of the figure surface could propagate to a distance up to 2-3 times as long as the longest intracortical connections. It does suggest, however, that the size invariance of BOWN latency may not hold for much larger figures, a prediction that could be physiologically tested (see Discussion).

Fig. (4A-D) demonstrate four other examples of the model's response to inputs resembling those used in physiological experiments, including figure occlusion, a C-shaped figure, figure transparency, and four squares (see Fig. (1CDEGHI)). Just as observed in physiology, the model exhibits appropriate BOWN tuning.

In Fig. (4A), the occluding square owns the occluding borders. That the two T junctions "1" and "2" provide the essential cue for the occlusion in cases such as this has led to suggestions that T junction detectors or labels are needed to determine figure-ground relationships. Indeed this has been the basis of previous models (Finkel and Sajda 1994, Craft et al 2004). In our model, there is no explicit T junction detector, rather the information about T junctions is implicit in the lateral interactions.

In Fig. (4B), if the context of the borders "1", "2", "3" (the '⊓' shape) were to be removed, the BOWN responses to these three borders would be the opposite of that shown, due to the local convexity bias among the three borders. The intracortical interactions, though of limited range, are able to cooperatively or collectively process contextual information from all the other borders, even ones that are far away. In the simulation, the latencies (from initial response) of the BOWN signals for these borders "1", "2", "3" are about twice as long as those of other borders whose ownership can more straightforwardly be determined. This is as if BOWNs for the ambiguous borders are determined after those of the less ambiguous ones. It is also apparent that the BOWN signal strength is weaker for the ambiguous borders.

Fig. (4)C and Fig. (4)D are response to stimuli adapted from the stimulus patterns Fig. 1I and Fig. 1E, which are like those used in physiological experiments (Qiu and von der Heydt 2003). The two stimulus patterns differ only
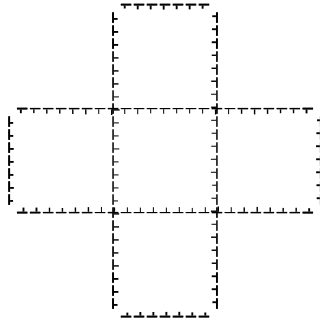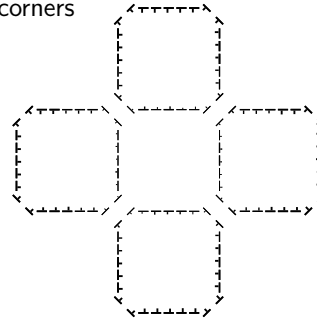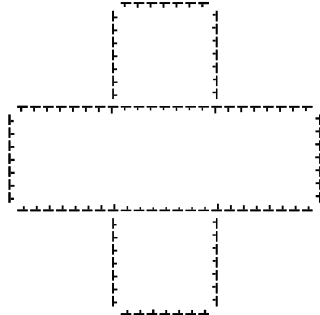
A Occluding square
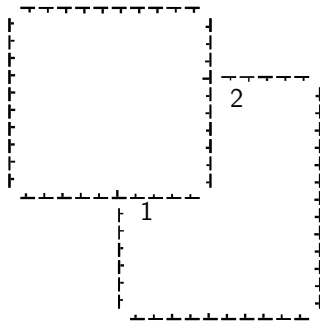
B C-shaped figure
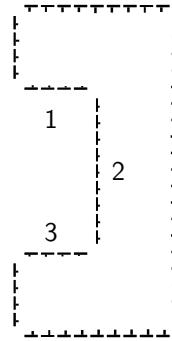
C Two transparent rectangles

D Four squares, rounded corners

E Two occluding rectangles
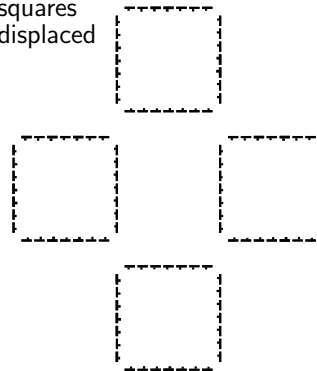
F Four squares displaced

Figure 4: Model responses to stimuli used in physiological experiments (A-D), and those not yet tried physiologically (E and F). Same format as in Fig. 3B-D. The pictures show model outputs after the initial transients, using stimulus patterns analogous to Fig. 3A for which there is no border ownership bias. E and F are further variations of the stimulus from those of C and D. Again, ownership tuning of cells is visualized by the fins of the contour segments. The number markings show particular borders and junctions that are described in the text.

by a few contour segments, effectively rounding the corners of the four squares. Nevertheless, the perceived BOWNs of the central borders are dramatically different, leading to changes in surface perception. Fig. (4C) is a transparent rectangle on top of another one and Fig. (4D) involves four identical square shapes with rounded corners. In both

physiology (Qiu and von der Heydt 2003) and the model, the neuronal responses are consistent with the perceived ownership signals. Note that, for the central border segments, the alternative contextual borders provide different and conflicting BOWN biases. Hence, the BOWN decisions for the central borders depend on the relative strengths of the conflicting contextual biases. Clearly, small changes in the stimulus from the inputs to Fig. (4C) to the inputs to Fig. (4D) reversed the relative order of the strengths of the conflicting biases. The model latencies of the BOWN signals for the central borders are again longer (especially in Fig. 4DE) than those for the borders in the periphery. The conflicts inherent in the contextual influences are likely to account for much of the longer latencies.

Fig. (4EF) show model simulations of two other variants of the stimulus corresponding to Fig. (4C), which have not yet been tried physiologically. Fig. (4E) changes the transparent occluder in Fig. (4C) into an opaque occluder, by removing border segments in the occluded region. Fig. (4F) displaces the top and bottom squares slightly away from the other two squares. Both examples further demonstrate that the slight stimulus variations can lead to dramatic changes in perception of the surfaces and BOWNs, and that these changes are readily matched by the model.

3   Discussion

We have shown that intracortical interactions in a V2 model suffice to explain the physiologically observed neural tuning to border ownership in this area, without invoking top-down feedback, explicit signals (such as labels for T junctions), or other mechanisms beyond V2. We demonstrated this in a model which involves only the essential elements relevant for this question: intracortical connections for contextual influences and analog, non-spiking, neuron models.

Based on many insightful psychophysical observations, Nakayama, He, and Shimojo (1995) had previously argued that low level, bottom-up, mechanisms should control surface perception and border ownership. However, influences over surface perception (such as the bi-stable perception of the flower vase and faces) of voluntary control suggest at least some role for top-down mechanisms. Physiological data cannot decide the extent of top-down involvement conclusively, since the short latency of the border ownership signal favors low level mechanisms (Zhou et al 2000), while the invariance of the latencies to figure size favors higher level mechanisms (Sugihara et al 2003). Our model suggests that it is indeed plausible that top-down mechanisms only play a modulatory role in the perception of border ownership.

To model the latency of BOWN signal and its dependence on figure sizes, one important model parameter is the axonal conduction latency between linked V2 cells. However, most indirect experimental measurements of this latency, coming from examining temporal characteristics of neural responses to visual stimulation at various distances (Grinvald et al, 1994 Bringuier et al 1999, Bair et al, 2003), obtain latency values (and inferred conduction speed) that include membrane integration time of neurons (even potentially including relay neurons) as well as axonal conduction time. Girard et al (2001) measured the latency more directly by activating axons (using a stimulating electrode) and recording (with a separate electrode) the resulting orthodromically-propagated spikes a small distance away. They found intra-cortical conduction latencies to range between 1 and 10 milliseconds, with about 90% of the

10

A A gray surface occluding a white surface in a black background

Image Input   Model Input   Model Output

B examples when surface luminance or attention make a difference

a, image input and model output for transparency

b, image input and model outputs for 4 white squares

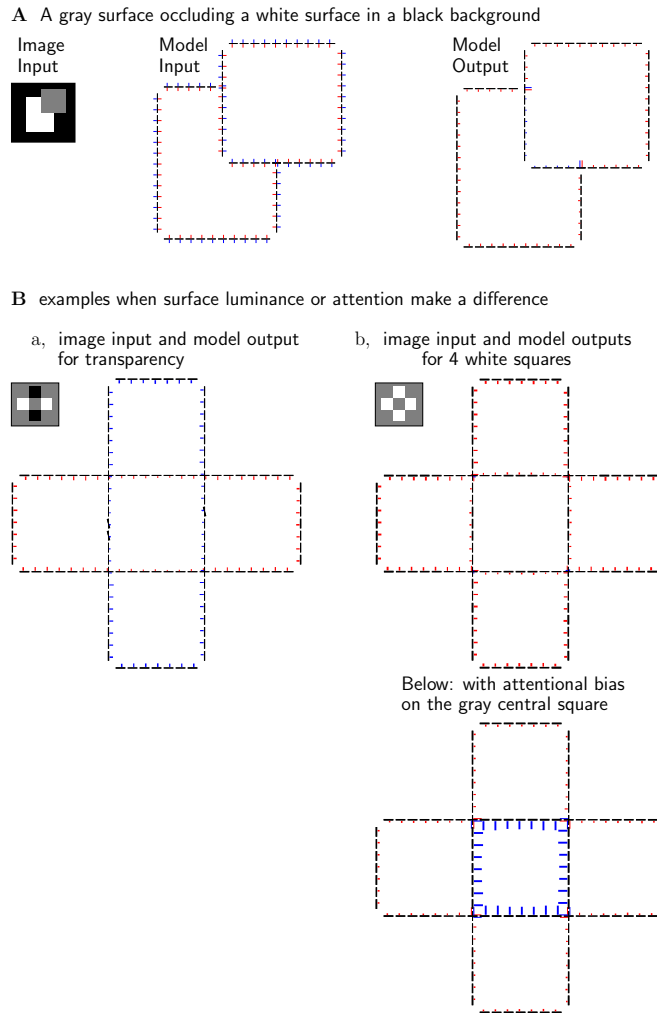Below: with attentional bias on the gray central square

Figure 5: Modeling contributions of surface features and top down control in BOWN processing. The model is augmented by replacing each original neuron by two neurons preferring the same BOWN but opposite contrast polarities. A border segment of a given contrast polarity gives equal input to two neurons preferring the opposite BOWNs and corresponding contrast polarities, e.g., a vertical border between left-black and right-white provides equal input to a cell preferring a lighter figure to the right of the border and another cell preferring a darker figure to the left of the border. For visualization, a red fin indicates that the cell prefers a contrast polarity such that the preferred figure side has higher luminance (or value of whatever other feature types) than the non-preferred figure side, a blue fin indicates otherwise. The neural connections are the same as before except that facilitatory (suppressive) connections of the type shown in Fig2Ba, 2Bb,2Bd, and 2Be, are reduced to 45% of the original strength if the two linked cells prefer the opposite (same) contrast polarities. The strengths of the connection types shown in Fig. 2B(c, f) stay the same regardless of preferred contrast polarities of the cells. Everything else about the model is the same as before. A: BOWN processing survives changes of contrast polarity along the border of the grey figure. At the T junctions, where the contrast polarity of the tops of the junction is ambiguous, inputs are provided with half strength to all four cells preferring each of the four possible combinations of the preferred BOWNs and polarities. B: when surface luminance and attention do matter. Changing the black squares in (a) to white squares in (b) makes different model responses consistent with perception. Note that in (ii), the borders at the intersecting corners of the squares have ambiguous polarities, hence, inputs are half-ed and provided to all cell types preferring any BOWN and polarity. In (b) bottom, attention is modelled by providing 20% additional inputs to neurons preferring the grey central square as the figure. All model outputs are plotted in the same format as in Fig. 3B-D, except for the color of the fins.

latencies below 7 millisecond (Fig. 3 of Girard et al 2001). Further and unpublished analysis of the data (Pascal Girard, private communication, 2004) show that this conduction latency does *not* seem to depend on the conduction

11

distance, and that the conduction speed shows a trend of growing linearly with conduction distance (in the range, 0.5-5.3 millimeters, of the distances observed). One should further note the following. (1) In the experiment of Girard et al (2001), if polysynaptic activations are excluded, then the stimulating and recording electrodes are likely *inbetween* pre- and post- synaptic cells, and thus the latencies obtained should be *shorter* than the actual conduction latencies between the pre- and post- synaptic cells. (2) The longer latencies obtained by Girard et al (2001) may possibly arise from polysynaptic activation. Since longer axons seem to conduct faster, it is not unreasonable to assume that the intra-cortical connections have the property that the conduction latencies are about the same regardless of the distances between linked cells, such that integration of all contextual information is roughly synchronous. It is under all these considerations that the model uses a random conduction latency between 0.8 to 1.0 membrane time constant, corresponding to 8-10 ms for membrane time constant of 10 ms, between pre- and post-synaptic cells regardless of separation between the cells.

While the model does confirm that the latencies of the BOWN signals are roughly invariant with sizes of simple surfaces like squares, it suggests that this invariance may be limited to a finite range of sizes. Longer latencies (in addition to weaker BOWN tuning) are observed in the model when figure sizes are larger than those shown in Fig. 3. Of course, different cells in the visual cortex have different sizes of receptive fields. If intra-cortical connections also have some degree of scale invariance such that longer connections link cells of larger receptive fields, size invariance of the BOWN signals could hold in the real visual cortex better than in the currently single-scale model. The model also show deterioration of size invariance in the BOWN latency as the input strength of the border segments decreases. In particular, shorter BOWN latencies are observed for smaller figures when input is weaker. Weaker inputs also cause the BOWN latency to be shorter at the ends than that in the middle of the sides of squares. Future experimental data on how BOWN latency and sensitivity vary with a wider range of figure sizes and with input strength should hopefully guide further development and revision of the computational model.

To explain the existing physiological data (using stimuli as summarized in Fig. 1B-I) about BOWN tuning, it seems that a minimal model without surface feature information (such as luminance) suffices. This does not prevent surface feature information from playing any role in surface perception (Metelli 1974, He and Ooi 1998). For instance, the bistable perception of vase vs. faces in Fig. 1A seems to be heavily influenced by the difference between the luminance value of the vase and face regions. Indeed, V1 does provide surface (e.g., luminance) information to V2, and this information should contribute to perception of the surface. There are even some cells in V2 whose responses to borders are influenced more or less significantly by the underlying surface luminance or the contrast polarity at the border (Zhou et al 2000).

The observations (Zhou et al 2000, Qiu et al 2001, von der Heyt et al 2003) that border ownership tuning in individual V2 cells is consistent under different stimulus cues, such as luminance contrast, depth, and motion, suggests that the model will generalize to cases in which additional surface information is integrated. A simple implementation would be to make model cell be tuned additionally to the feature (e.g., luminance) contrast polarity of the borders. So each cell in the original model is now replaced by two cells preferring the same BOWN, orientation, and CRF, but opposite contrast polarities. Hence, a border of a given contrast polarity would provide input (from V1) to cells preferring this polarity, again regardless of the cells' preferred BOWNs. Intracortical connections between

cells would in turn be modulated by the preferred contrast polarities of the linked cells, such that mutual facilitation (or suppression) between cells tuned to consistent (or inconsistent) BOWNs, as in Fig. 2B(a-b) (or Fig. 2B(d-e)), would be somewhat weaker when the two connected cells prefer opposite (or same) contrast polarities.

As expected, this augmented model, including the surface features, helps to enhance the BOWN signals in some cases, such as to stimulus Fig. 1DH. Fig. 5 demonstrates the behavior of this augmented model in more interesting cases when surface (luminance) feature information is overridden by contour information, or affects the resulting perception. Consistent with perception, Fig. 5A shows that the grey surface continues to be perceived by the model as the occluding figure, even though the contrast polarity changes along its border as it occludes two surfaces of different luminances. Fig. 5B demonstrates that the perception of a white transparent rectangle occluding a black one can be changed to the perception of four white figures when the luminance of the two black surfaces is changed to white, again consistent with human perception. Fig. 5Bb additionally demonstrates that a central grey square flanked by the four white squares can be perceived as figure rather than background by an attentional bias, speculatively modelled as an 20% additional input to cells tuned to the biased BOWN along the border of this grey square (this is based on physiological observations that attentional differences lead to sensitivity differences in cortical neurons to their inputs in cortical areas V2, MT, etc. (Motter 1993, Treue and Maunsell 1999)). While Fig. 5 demonstrates the possible additional powers of the model with additional mechanisms, these mechanisms are only coarsely and speculatively modelled here and are not within the main purpose of this paper. They require further study, especially when they are more precisely pinned down through physiological experiments. The main goal in this paper is to answer the feasibility question of whether V2 mechanisms suffice for generating BOWN tuning. For the same reason, the model is a minimal network using only the essential elements relevant for our question: intracortical connections for contextual influences and a simple neuron model, without extraneous details such as neural spikes and ion channels.

Ours is the first model to generate BOWN tuning in such a minimal manner. Previous models of border ownership utilize additional external signals such as T-junctions (Finkel and Sajda 1994) and L junctions (corners) (Kikuchi and Akashi 2001), or use higher level neural units representing global figures (Craft et al 2004). These additional external mechanisms essentially *answer* the BOWN question by themselves. Hence, these models do not indicate whether or not V2 mechanisms alone can determine BOWN. Interestingly, cells for the global figures used in Craft et al (2004), termed grouping cells, are multiscale, and if they exist as predicted by the authors, within or beyond V2 for the BOWN processing by reciprocal connections with the BOWN tuned cells, they would enable size invariance of the BOWN latency over a wider range of figure sizes. A recent model (Nishimura and Sakai 2004) has generated limited BOWN tuning, using surround suppression and facilitation for contextual influences using luminance cues. It can account for some (Fig. (1BCDFH)) but not others (Fig. (1EGI)) of the physiologically tested cases of BOWN tuning. Relying heavily on luminance cues, it would be difficult for it to generalize to contour images, or explain the persistence of BOWN in the face of inconsistent border contrast polarities (Fig. 5A), or to generate dramatic changes in perceptual border ownership on the basis of small changes to an input stimulus, as observed in Fig. (4C-F). There are also models (e.g., Li 1998) of contour integration or enhancement by intracortical mechanisms in V1, and a V2 model (Zhaoping 2002) of grouping and segmentation of surfaces by depths, but these also do not generate BOWN

tuning. Grossberg (1994) has suggested a number of interesting models of border and surface processing, though these also do not address BOWN tuning.

Although this model is simple and minimal, it makes non-trivial and directly testable predictions. In addition to the prediction on how BOWN latency could vary with a wider range of figure sizes and with input strength, there are the following predictions. In particular, BOWN latency is predicted to be longer for the border segments whose multiple contextual regions provide opposite or conflicting ownership biases, such as the segments marked in Fig. (4B), and for border segments near the insections of several surfaces, such as in Fig. (4C-F). These longer latencies for the borders indicated in Fig. (4B-F) suggest that local or global surface complexity slows down the cooperative computation of border ownerships.
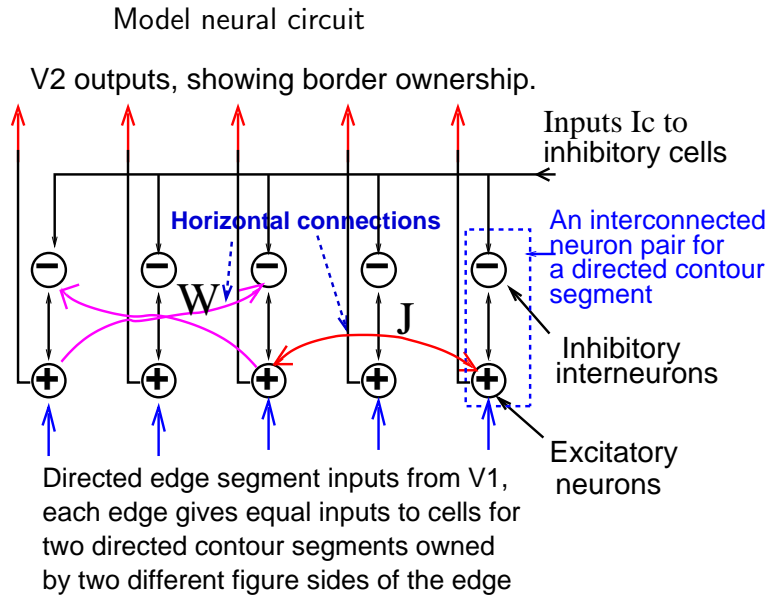
4 Procedures

Figure 6: Model neural circuit. An excitatory pyramidal (principal) cell codes for one specific CRF location and preferred orientation/BOWN. The lateral connections J and W mediate monosynaptic excitation (J) or di-synaptic inhibition (W) between the principal cells. Identical V1 inputs are always directed to two principal cells of the same CRF location and preferred orientation but opposite preferred BOWNs (such as in Fig. 2(Ab) and 2(Ac)). Lateral interactions in the model cause different output strengths (e.g., Fig. 2(Ad)) from these two cells.

We present full technical details and quantitative model parameters in this section. (General understanding of this paper does not require following most of the technical details here.) For each (discrete) CRF location i (evenly spaced on a 2-d Manhattan grid with wrap around boundary conditions) and preferred border direction θ (24 possible values evenly spaced within $(0, 2\pi)$), there is a principal pyramidal cell with a state variable $x_{i\theta}$

modeling membrane potential, and activation or firing rate $g_x(x_{i\theta})$ which is a sigmoid-like function of $x_{i\theta}$. This pyramidal cell is paired with an inhibitory interneuron which has membrane potential $y_{i\theta}$, a firing rate $g_y(y_{i\theta})$, receives excitatory drive $g_x(x_{i\theta})$ from the principal cell, and gives inhibitory feedback $g_y(y_{i\theta})$ in return (Fig. 6). Monosynaptic excitation from one pyramidal cell (j, θ') to another (i, θ) is $J_{i\theta, j\theta'} g_x(x_{j\theta'})$ where $J_{i\theta, j\theta'}$ models the synaptic connection strength. Similarly, interneuron (i, θ) receives excitation $W_{i\theta, j\theta'} g_x(x_{j\theta'})$ from pyramidal cell (j, θ'), mediating di-synaptic inhibition from pyramidal cell (j, θ') to (i, θ). Each model neuron models a local group of cells with similar properties and feature preferences (Amari 1972). A model cell's output, i.e., $g_x(x_{i\theta})$, may be seen as modeling the total firing rate (within a scale factor) of spikes generated from all cells in this local cell group. The model outputs as plotted in the figures are from the pyramidal cells, i.e., $O_{i\theta} = g_x(x_{i\theta})$.

The visual input from V1 for a contour segment at discrete location i and orientation $0 \leq \theta < \pi$ is $I_{i\theta}$. It models the total outputs of a local group of V1 cells having similar tuning properties, with a graded value ranging from zero to a maximum evoking, correspondingly, null to saturating responses of the principal cell for a single isolated contour segment. It is received by two principal (pyramidal, excitatory) neurons which share CRF location i but prefer directions θ and $\theta + \pi$ respectively. For notational convenience, inputs are treated as also directional, spanning 0 to 2π, with the constraint $I_{i\theta} = I_{i,\theta+\pi}$ for all (i, θ) unless otherwise stated. Due to the finite orientation tuning width of the cells, a directed edge (i, θ) in an input image such as Fig. 3A provides input $I_{i\theta'} = \hat{I} e^{-|\theta'-\theta|/(\pi/8)}$ to cells $(i\theta')$ preferring $\theta' \approx \theta$. All simulation examples shown in this paper use $\hat{I} = 3.5$, corresponding to medium-high contrast input, though other values could also be used.

Given population inputs $I_{i\theta}$, the neural activities at time t evolve according to the equations:

$$
\begin{aligned}
\dot{x}_{i\theta}(t) = & -\alpha_x x_{i\theta}(t) - g_y(y_{i,\theta}(t)) - \sum_{\Delta\theta \neq 0} \psi(\Delta\theta) g_y(y_{i,\theta+\Delta\theta}(t)) \\
& + \sum_{j \neq i, \theta'} J_{i\theta, j\theta'} g_x(x_{j\theta'}(t - \tau^J_{i\theta, j\theta'})) + I_{i\theta}(t) + I_o + N^x_{i\theta}(t) + I^{normalization}_{i\theta}(t) \quad (1)
\end{aligned}
$$

$$
\dot{y}_{i\theta}(t) = -\alpha_y y_{i\theta}(t) + g_x(x_{i\theta}(t)) + \sum_{j \neq i, \theta'} W_{i\theta, j\theta'} g_x(x_{j\theta'}(t - \tau^W_{i\theta, j\theta'})) + I_c + N^y_{i\theta}(t) \quad (2)
$$

In these equations, $\alpha_x x_{i\theta}$ and $\alpha_y y_{i\theta}$ model the decay to resting potentials, implying membrane time constants of $1/\alpha_x$ and $1/\alpha_y$ for the pyramidal and the interneuron, respectively; For the interneurons, $\alpha_y y_{i\theta}$ could also include the effect of mutual inhibition between the local interneurons that constitute a model interneuron; $\psi(\Delta\theta) \leq 1$ is the spread of inhibition within a hypercolumn (cells of the same i but different θ); I_o and I_c are static background inputs, $N^x_{i\theta}$ and $N^y_{i\theta}$ model fluctuating neural noises, $I^{normalization}_{i\theta}$ models suppressive inputs due to the normalization of local activities (Heeger 1992)), and $\tau^J_{i\theta, j\theta'}$ and $\tau^W_{i\theta, j\theta'}$ model axonal conduction delays between linked cells.

The model parameters used in the equations are as follows.

$$\alpha_x = \alpha_y = 1, \text{ hence all time here is in the unit of the membrane time constant}$$

$$g_x(x) = \begin{cases} 0 & \text{if } x < x_{th} = 1 \\ (x - x_{th}) & \text{if } x_{th} \leq x \leq x_{sat} = 2 \\ (x_{sat} - x_{th}) & \text{if } x > x_{sat}. \quad \text{Note that the maximum response level is 1.} \end{cases}$$

$$g_y(y) = \begin{cases} 0 & \text{if } y < y_{th1} = 1 \\ g_1(y - y_{th1}) & \text{if } y_{th1} \leq y \leq y_{th2} = 1.2, \text{ where } g_1 = 0.21 \\ g_2(y - y_{th2}) + g_1(y_{th2} - y_{th1}) & \text{if } y_{th2} \leq y \leq y_{sat} = 300, \text{ where } g_2 = 2.5 \\ g_2(y_{sat} - y_{th2}) + g_1(y_{th2} - y_{th1}) & \text{if } y > y_{sat} \end{cases}$$

$$\psi(\theta) = \begin{cases} 0.8 & \text{if } |\theta| = \pi/12 \\ 0.1 & \text{if } |\theta| = \pi/6 \\ 0 & \text{otherwise} \end{cases}$$

The background static inputs are $(I_o, I_c) = (0.1, 1.0)$; $N_{i\theta}^x$ and $N_{i\theta}^y$ are zero-mean noises independent between neurons, fluctuating with an average amplitude of 0.2 and a temporal correlation length roughly 10% of the membrane time constant; and $I_{i\theta}^{normalization} = -(\hat{a}_i)^2/128$ depends on the sum $\hat{a}_i$ of pyramidal activities (outputs) in the vicinity of i, defined as all spatial locations displaced from i by no more than 2 grid units horizontally and vertically. This normalization input provides increasing suppression to a pyramidal's activity as the overall pyramidal activities in the vicinity increases, thus helps to ensure stability of the network. Each axonal conduction latency $\tau_{i\theta,j\theta'}^{J,W}$ is a random number within the range $(0.8, 1)/\alpha_x$, and for simplicity, does not change with the post-synaptic cell.

To describe the synaptic weights, we need some notation. Let β be the direction of the spatial displacement $j - i$ (spatial distance is in the unit of the grid) from one cell $i\theta$ to another $j\theta'$, $d = |i - j|$, and $0 \leq \theta, \theta' < 2\pi$. Let $\theta_1 = \phi(\theta, \beta)$ and $\theta_2 = \phi(\beta, \theta')$, where $\phi(x, y) = x - y$, or $x - y + 2\pi$, or $x - y - 2\pi$ for $-\pi < x - y \leq \pi$, or $x - y \leq -\pi$, or $x - y > \pi$, respectively. Denoting $\text{sign}(x) = 1$ for $x > 0$ and $\text{sign}(x) = -1$ otherwise, define $(\theta_1', \theta_2') = (\text{sign}(\theta_1)|\pi - |\theta_1||, \text{sign}(\theta_2)|\pi - |\theta_2||)$. Then

$$(\theta_a, \theta_b) \equiv \begin{cases} (\theta_1, \theta_2) & \text{if } |\theta_1| + |\theta_2| \leq |\theta_1'| + |\theta_2'| \\ (\theta_1', \theta_2') & \text{otherwise} \end{cases}$$

Now θ_a and θ_b describe the directional angle between the two border segments (i, θ) and $(j\theta')$ and the spatial displacement $j - i$. The directional angles are positive or negative if a right or left turn of no more than half a cycle brings the border segments align with $j - i$ or $i - j$. Define $\theta_\pm' \equiv \theta_a \pm \theta_b$, let $\theta_\pm = \theta_\pm'$, or $2\pi - \theta_\pm'$, or $-2\pi - \theta_\pm'$, for

$-\pi \leq \theta'_\pm \leq \pi$, or $\theta'_\pm > \pi$, or $\theta'_\pm < -\pi$, respectively.

$$
J_{i\theta,j\theta'} = \begin{cases}
(11/108)e^{-(3-2.5\text{sign}(\theta_+))|\theta_+|/(5\pi)-2\theta_-^2/\pi^2}f_1(d) & \text{if } |\theta_a| \leq \pi/11, |\theta_b| \leq \pi/11 \\
(11/81)e^{-3\theta_+/(5\pi)-2\theta_-^2/\pi^2}f_2(d) & \text{otherwise, if } \theta_a, \theta_b \geq 0, \theta_+ < \pi/2.01 \\
(11/81)e^{-(9\theta_+/(8\pi))^2-2\theta_-^2/\pi^2}f_2(d) & \text{otherwise, if } \theta_a, \theta_b \geq 0, \theta_+ \geq \pi/2.01 \\
& \quad |\theta_-| < \pi/2.01 \\
(11/81)e^{-(9\theta_+/(8\pi))^2-0.5(\theta_-/(\pi/2))^6}f_2(d) & \text{otherwise, if } \theta_a, \theta_b \geq 0, \theta_+ \geq \pi/2.01 \\
& \quad |\theta_-| \geq \pi/2.01 \\
(11/81)e^{-4(\theta_+/\pi)^2-9\theta_-^2/\pi^2}f_2(d) & \text{otherwise, if } \theta_a, \theta_b \leq 0 \\
(11/81)e^{11.5\text{sign}(\theta_+)\theta_+^2/\pi^2-14\theta_-^2/\pi^2}f_2(d) & \text{otherwise, if } \theta_a \cdot \theta_b \leq 0, |\theta_-| < \pi/2.01 \\
(11/81)e^{11.5\text{sign}(\theta_+)\theta_+^2/\pi^2-(15/4)(2\theta_-/\pi)^6}f_2(d) & \text{otherwise, if } \theta_a \cdot \theta_b \leq 0, |\theta_-| \geq \pi/2.01
\end{cases}
$$

where $f_1(d) = e^{-(d/9)^2}$, $f_2(d) = e^{-d/5}$, and $f_1(d) = f_2(d) = 0$ for $d > 10$ and $d = 0$. This, though cumbersome, is no more than a piece-wise parameterization of the lateral connections with changes in spatial configuration between the underlying border segments, as qualitatively described in Fig. 2B(a, b, d, e). Additionally, the connection strength decays with distance between linked cells, vanishes for distance larger than 10, and is a translation invariant quantity depending only on θ, θ', and the relative displacement $j - i$. Meanwhile, the connections onto the interneurons are

$$
W_{i\theta,j\theta'} = c(J_{i(\theta+\pi)\%(2\pi),j\theta'} + J_{i\theta,j(\theta'+\pi)\%(2\pi)})/J_{i,0,i+1_x,0}
$$

where $x\%(2\pi) = x$ if $x < 2\pi$ and $x\%(2\pi) = x - 2\pi$ otherwise, $i + 1_x$ is the grid position one unit displaced from i horizontally, and $c = 0.02646$ usually except, when (θ_a, θ_b) as defined above for the two border segments $(i\theta)$ and $(j(\theta' + \pi)\%(2\pi))$ satisfy $|\theta_a|, |\theta_b| \leq \pi/11$ (i.e., these two segments are roughly aligned), in which case $c = 0.0147$. All of these synaptic weights describe only the connection types qualitatively indicated in Fig. 2B(a, b, d, e). If, however, the two border segments for the two cells are close enough to each other, and are near perpendicular to each other like a T junction, the synaptic weights take different values. Such a spatial relationship is judged by if the two segments satisfy either

$$
0 < d \leq 2, \quad |\theta_a d| < 0.5, \quad \text{and} \quad \pi/3.1 < |\theta_b| < 2\pi/3.1 \tag{3}
$$

or

$$
0 < d \leq 2, \quad |\theta_b d| < 0.5, \quad \text{and} \quad \pi/3.1 < |\theta_a| < 2\pi/3.1 \tag{4}
$$

When condition (3) is satisfied, the connections are zero unless $\theta_b < 0$, then

$$
J_{i\theta,j\theta'} = f_T(d)e^{-2|\theta_a d|}, \quad W_{i\theta,j(\theta'+\pi)\%(2\pi)} = 0.0588f_T(d)e^{-2|\theta_a d|}e^{-20|\pi/2+\theta_b|/\pi}/J_{i,0,i+1_x,0}.
$$

Here $f_T(d) = (11/90)e^{-d/6}$. While if condition (4) is satisfied, the connections are zero unless $\theta_a < 0$, then

$$
J_{i\theta,j\theta'} = 3f_T(d)e^{-2|\theta_b d|}, \quad W_{i(\theta+\pi)\%(2\pi),j\theta'} = 0.0588f_T(d)e^{-2|\theta_b d|}e^{-20|\pi/2+\theta_a|/\pi}/J_{i,0,i+1_x,0},
$$

These connections are schematically shown in Fig. 2Bc and 2Bf.

17

The quantitative values for the lateral connections are designed such that the desired contextual influence for BOWN are achieved. In particular, this requires achieving BOWN tuning for a simple square figure and for one surface occluding another. Some mathematical analysis for stability of the recurrent networks is necessary to ensure that the network is well behaved, analogous to those for a recurrent model of primary visual cortex (Li and Dayan 1999, Li, 2001). However, due to the known mathematical difficulties in differential equations with time delays, the analysis is approximate and has to be supported by simulations. While the technical details are not important for the aim of this paper, it is suffices to mention that the model parameters are roughly robust once a desirable parameter region is reached, since insignificant changes of the parameters do not destroy the overall model behavior in simulations. For instance, the model behavior is not too sensitive to small changes in how the synaptic weights decay with distance, given a fixed value of the total synaptic weight integrated over distance and the longest length of the connections. This robustness is demonstrated by the fact that the same model parameters are used in the simulations of all the stimulus examples shown in this paper (with additional parameters described in the caption of Fig. 5 when including additional model features such as surface luminance and attentional influences), and random dynamic noise (as described above) in neural inputs does not destroy the desired network behavior. Further analysis of this model network is expected to be extensive, given the dynamic complexities involved in recurrent networks of such type. However, future physiological and anatomical data should hopefully help to constrain the network parameters and thus reduce the difficulties.

Acknowledgements

References

[1] Amari S-I. (1972) Characteristics of random nets of analog neuron-like elements. *IEEE trans. systems. man. and cybernetics.* 2(5), p643-657.

[2] Angelucci A, Levitt JB, Walton EJ, Hupe JM, Bullier J, Lund JS (2002) Circuits for local and global signal integration in primary visual cortex *J. Neurosci.* 22(19):8633-46

[3] Bair W, Cavanaugh JR, Movshon JA. (2003). Time course and time-distance relationships for surround suppression in macaque V1 neurons. *J. Neurosci.* 23(20):7690-701.

[4] Bringuier V., Chavane F., Glaeser L., Fregnac F. (1999) Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* 283(5402):695-9

[5] Craft E. Schutze H. Niebur E. & von der Heydt R. (2004) A physiologically inspired model of border ownership assignment Vision Science Society, Abstract, Florida, USA Journal of Vision, 4(8), 728a, http://journalofvision.org/4/8/728/, doi:10.1167/4.8.728. See also manuscript submitted for publication.

[6] Finkel L.H. & Sajda P. (1994) Constructing visual perception *Am. Scientist* 82:224-237

[7] Girard P, Hupe JM, Bullier J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J Neurophysiol.* 85(3):1328-31.

[8] Grinvald A, Lieke EE, Frostig RD, Hildesheim R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *J Neurosci.* 14(5 Pt 1):2545-68.

[9] Grossberg S. (1994) 3-D vision and figure-ground separation by visual cortex. *Percept. Psychophys.* 55(1):48-121

[10] He ZJ and Ooi TL. (1998) Illusory-contour formation affected by luminance contrast polarity. *Perception* 27(3):313-35.

[11] Heider B, Meskenaite V, Peterhans E. (2000) Anatomy and physiology of a neural mechanism defining depth order and contrast polarity at illusory contours. *Eur J Neurosci.* 12(11):4117-30.

[12] . Heitger F, Rosenthaler L, von der Heydt R, Peterhans E, Kubler O. (1992) Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Res.* 32(5):963-81.

[13] Kikuchi M. and Akashi Y. (2001). A model of border-ownership coding in early vision *Lecture Notes in Computer Science 2130*, Artificial Neural Networks-ICANN2001, eds.: G. Dorffner, H. Bishop and K. Hornik, Springer, pp.1069-1074.

[14] Li Z. (1998) A neural model of contour integration in the primary visual cortex" *Neural Computation* 10, 903-940.

[15] Li, Z. and Dayan P. (1999) Computational differences between asymmetrical and symmetrical networks" *Network: Computation in Neural Systems* Vol. 10, 1, 59-77.

[16] Li, Z. (2001) Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. *Neural Computation*, 13/8, p.1749-1780, 2001.

[17] Metelli F. (1974) Scientific American 230(4) 90-98.

[18] Motter B.C. (1993) Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *J. Neurophysiol.* 70(3):909-19.

[19] Nakayama, K., He, Z., and Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In Kosslyn, S. and Osherson, D., editors, Visual cognition: An invitation to cognitive science, Vol. 2 (2nd ed.), pages 1–70. MIT Press, Cambridge, MA.

[20] Nishimura H. Sakai K. (2004) Determination of border ownership based on the surround context of contrast. *Neurocomputing*, Vol. 58-60, p843-848.

[21] Popple AV. Context effects on texture border localization bias (2003) *Vision Res.* 43(7):739-43.

[22] Qiu F.T., Macuda T. J. & von der Heydt (2001) Neural coding of border ownership: Figural and stereoscopic cues Soc. of. Neurosci. Abstract, 27, 165.6.

[23] Qiu F.T. & von der Heydt R. (2003) Interaction of border ownership and transparency in the monkey visual cortex. Vision Science Society, abstract 115, Florida, USA Journal of vision 3(9), abstract 115, p115a, http://journalofvision.org/3/9/115/

[24] Rubin E. (2001). Figure and ground. in *Visual Perception Essential Readings* ed. Yantis, S. 225-230. Psychology Press, Philadelphia.

[25] Sugihara T. Qiu F.T. and von der Heydt R. (2003) Border-ownership Coding in Monkay area V2: Dynamics of image context integration. Soc. of. Neurosci. Abstract, 819.12.

[26] Tamura H., Sato H., Katsuyama N., Hata Y., Tsumoto T. Less segragated processing of visual information in V2 than in V1 of the monkey visual cortex. (1999) *Eur. J. Neurosci.* 8(2):300-9.

[27] Treue S. and Maunsell, JH. (1999)). Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *J. Neurosci.* 19(17): 7591-602.

[28] von der Heydt R. Zhou H. Friedman HS (2000) Representation of stereoscopic edges in monkey visual cortex *Vision Res.* 40(15):1955-67

[29] von der Heydt R. Qiu F. T. & He. Z.J. (2003) Neural Mechanisms in border ownership assignment: motion parallax and Gestalt Cues. Vision Science Society, abstract 666, Florida, USA Journal of Vision, 3(9), abstract 666, p666a. http://journalofvision.org/3/9/666/

[30] von der Heydt R. (2003) Neural coding of border ownership: implications for the theory of figure-ground perception In *Perceptual Organization in vision: Behavioral and Neural Perspectives* Behrmann M, Kimchi R, Olson CR, eds), Lawrance Erlbaum Associates.

[31] Yu C, Klein SA, Levi DM. Facilitation of contrast detection by cross-oriented surround stimuli and its psychophysical mechanisms (2002), *J Vis.* 2(3):243-55

[32] Zhaoping L. (2002) Pre-attentive segmentation and correspondence in stereo *Philos Trans R Soc Lond B Biol Sci.* 357(1428):1877-83.

[33] Zhou. H., Friedman, H.S., & von der Heydt, R. (2000) Coding of border ownership in monkey visual cortex *J. Neuroscience* 20, 6594-6611