Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Physiology - Paris 97 (2003) 503–515
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

V1 mechanisms and some figure–ground and border effects

Li Zhaoping

Department of Psychology, University College London, Gower Street, London WC1E 6BT, UK

Abstract

V1 neurons have been observed to respond more strongly to figure than background regions. Within a figure region, the responses are usually stronger near figure boundaries (the border effect), than further inside the boundaries. Sometimes the medial axes of the figures (e.g., the vertical midline of a vertical figure strip) induce secondary, intermediate, response peaks (the medial axis effect). Related is the physiologically elusive ''cross-orientation facilitation'', the observation that a cell's response to a grating patch can be facilitated by an orthogonally oriented grating in the surround. Higher center feedbacks have been suggested to cause these figure–ground effects. It has been shown, using a V1 model, that the causes could be intra-cortical interactions within V1 that serve pre-attentive visual segmentation, particularly, object boundary detection. Furthermore, whereas the border effect is robust, the figure–ground effects in the interior of a figure, in particular, the medial axis effect, are by-products of the border effect and are predicted to diminish to zero for larger figures. This model prediction (of the figure size dependence) was subsequently confirmed physiologically, and supported by findings that the response modulations by texture surround do not depend on feedbacks from V2. In addition, the model explains the ''cross-orientation facilitation'' as caused by a dis-inhibition, to the cell responding to the center of the central grating, by the background grating. Furthermore, the elusiveness of this phenomena was accounted for by the insight that it depends critically on the size of the figure grating. The model is applied to understand some figure–ground effects and segmentation in psychophysics: in particular, that contrast discrimination threshold is lower within and at the center of a closed contour than that in the background, and that a very briefly presented vernier target can perceptually shine through a subsequently presented large grating centered at the same location.

Keywords: Segmentation; Contextual influences; Ripple effect; Closed contour; Shine through

1. Introduction

Segmenting figure from ground is one of the most important visual tasks, since it is seen as a pre-requisite for object recognition. While this topic has been studied extensively in computer vision and human psychophysics, physiological studies to probe the neural correlates of figure–ground segmentation in early visual cortex started only in recent years. In this paper, I review the relevant physiological observations on the ''figure–ground'' effects triggered by Lamme's finding that neural responses in V1 are higher to figures than background [20]. I will then relate them to physiological data on contextual and surround influences to cell responses in cortex. A V1 model is then used to demonstrate a proposal that V1 mechanisms, in particular, the intracortical interaction, are the causes of the physiological ''figure–ground effects''. Additional model predictions will be presented, and subsequent physiological data confirming model predictions will be reviewed. I will use the insights gained from the model to account for some figure–ground and segmentation effects observed psychophysically.

V1 is usually considered a low level visual area, its classical receptive fields (CRFs) are usually much smaller than the sizes of most figure surfaces. It is therefore exciting to find that neural responses in V1 are higher to figures than background [20,21,40]—the *figure–ground effect*. Further experiments revealed that the medial axis of a figure can sometimes induce even higher responses than the figure surface nearby [23]— the *medial axis effect* (see Fig. 1 for illustration). This effect is worth noting since, computationally, a convenient skeleton representation of a figure surface is suggested to be the medial axis transform [1], formally defined as the locus of the centers of the largest circles inside the figure region. It is a set of connected lines that are a

E-mail address: z.li@ucl.ac.uk (L. Zhaoping).

0928-4257/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jphysparis.2004.01.008

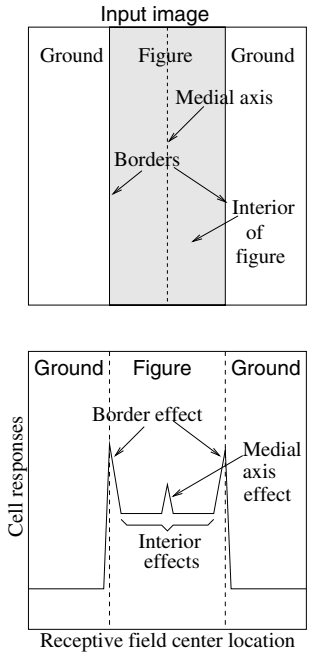

Fig. 1. Illustrating the figure–ground effect and its components: the border effect and interior effects, which includes the medial axis effect.

formal reduction of the shape of a surface (think of a stick figure for a man). The response differentiation between figure and ground becomes significant 80 ms after stimulus onset or 30–40 ms after the initial responses [20,22,23,40], late enough to allow contributions from higher visual areas. Furthermore, the figure–ground effects can be reduced by anesthesia or lesions in higher visual areas [21]. Hence, there was a common assumption that they mainly result from feedbacks from higher visual areas [20,23,40].

However, it is obviously important to consider how the figure–ground effects may result from boundary processing, a computational task more closely associated with V1. Indeed, another experiment [6] found that V1 cells robustly give higher responses to global borders between two texture regions, even under anesthesia. Furthermore, in the experiments showing the figure–ground effect [21,23,40], the response to the figure surface is usually highest near the figure boundary rather than anywhere further inside the boundary, including the medial axis. The differentiation between response levels to figure and ground appears earlier near figure boundaries and is significant at 10–15 ms [6] or 10–20 ms [22,23] after the initial responses, whereas it takes 30–40 ms after the initial responses to differentiate responses to figure interior from that to the ground [20,22,23,40]. The *figure–ground effect* thus consists of the *border effect* (Fig. 1), the response highlight to part of the figure near the boundary, and the *interior effects* (including the medial axis effect), the response highlights further inside the boundary.

In 1999, Li proposed [28] that V1 mechanisms are mainly responsible for these figure–ground effects observed physiologically, and that the interior effects, in particular, the medial axis effect, are by-products of the border effect. This proposal was inspired by the observations by Gallant et al. [6], as well as the following anatomical and physiological findings. Finite range intra-cortical interactions [5,7,11,34] cause the responses of a cell to be modulated by stimuli that are nearby, but outside its CRF. They are manifested in the contextual influences seen experimentally, which are mainly suppressive, though sometimes facilitatory. For instance, Knierim and van Essen [17] observed that a cell's response to an optimally oriented bar can be reduced by 80% when the bar is surrounded by similarly oriented bars near but outside the CRF. This is termed iso-orientation suppression. The surround suppression is weaker if the surround bars are oriented randomly, and is the weakest when the surround bars are oriented orthogonally to the central bar. A related observation is ''cross-orientation facilitation'', observed by Sillito et al. [37], that a V1 cell's response to a grating patch can be enhanced when the grating is surrounded by an orthogonally oriented grating. This facilitation effect was elusive as some subsequent attempts by other researchers failed to find it. Kapadia et al. [14] found that a V1 cell's response to a bar can be enhanced when contextual bars are aligned with the central bar to form a smooth line or contour—colinear facilitation. All these contextual influences, if caused by V1 mechanisms only, should be accounted for by the same V1 neural circuit of the intra-cortical interaction. The finite range interaction, mediated by axons extending a few millimeters laterally [7,34], i.e., linking CRFs separated by up to a few CRFs from each other, could propagate to make V1 cells sensitive to long range image features despite the locality of their CRFs.

Li's proposal was validated [28,30] by using a model of V1 whose parameters are chosen such that the model's responses to stimuli are consistent with the experimental data summarized above on intra-cortical interactions and contextual influences [25–27]. The model cells with nearby but not necessarily overlapping CRFs interact via intra-cortical connections. The model exhibited the border and interior effects, in particular the medial axis effect, and allowed to probe the dependence of these effects on size, shape, and texture features of the figures. It showed that whereas the border effect is robust, the interior, and, in particular, the medial axis, effects are by-products of the border effect. Furthermore, the interior effect is predicted to diminish as the figure size increases and the medial axis effect is predicted to be significant only for certain figure sizes. Figure size specificity of the medial axis effect was indeed evident in the original data [23]. Subsequently, new physiological data [35] confirmed the predicted diminishing response

enhancement to figure interiors of increasingly large figures. Meanwhile, it was shown that the surround modulations of V1 responses do not depend on V2 feedbacks [12].

The insights provided by the model allowed an understanding of the elusive "cross-orientation facilitation" as dis-inhibition of the response to the center of the figure grating by the background grating. The model reveals that this effect can only be manifested within a small range of sizes of the figure grating, thus explaining its elusiveness in experimental investigations. Other related surround modulations, such as the extent of the surround summation and suppression as manifested in a V1 cell's responses to a grating [36] can also be accounted for.

Psychophysically, contrast detection tasks have been observed to be easier inside a closed contour, presumed as figure, than those in the background image regions [18]. A familiar dependence of this effect on the size of the contour region was also observed [19]. Recently, a "shine-through" phenomena, that a very briefly presented vernier target can be perceived as superposed on a subsequently presented grating, was also shown to depend on the size of the grating. I will show how the V1 model can also provide insights in these psychophysical phenomena.

In the rest of the paper, I will first describe the V1 model. Then the model is used as an organization guide to understand the neural mechanisms behind, and to provide a link between, the physiological and the psychophysical data outlined in this section.

2. Methods

The model contains arrays of model neuron units tuned to orientation and spatial location (see below). A unit (i, θ) has CRF center at location i and prefers orientation θ. An image is processed through the corresponding receptive fields to provide input to individual model units. The units interact with each other via lateral connections, using both monosynaptic facilitation and disynaptic inhibition through interneurons [7,11,34, 38]. Fig. 2 shows the elements of the model and their interactions. The lateral connections tend to link cells preferring similar orientations, giving orientation

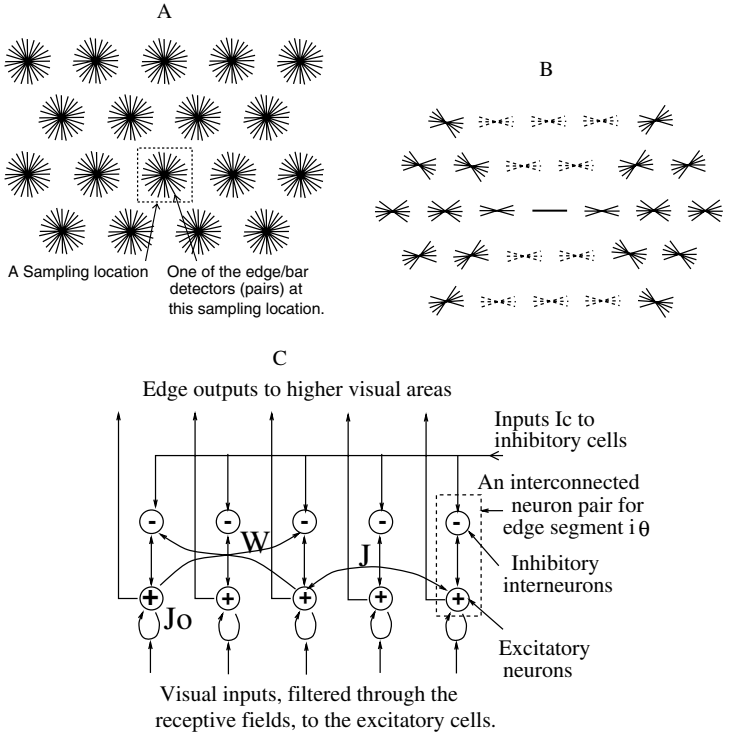

Fig. 2. (A) Visual inputs are sampled in a discrete grid of edge/bar detectors. Each grid point i has 12 neuron pairs, each tuned to a different orientation θ, spanning 180°. Each neural pair consists of an interconnected excitatory and inhibitory cells (see (C)). Two neural pairs $i\theta$ and $j\theta'$ at different grid points i and j can interact with each other via lateral connections. (B) A schematic of the lateral connection pattern. A bar at position i and oriented at θ denotes a corresponding neural pair $i\theta$. The pre-synaptic cell is marked by a thick solid bar, the postsynaptic targets are marked either by the thin solid bars or the thin dashed bars, depending on whether the pre-synaptic inputs are mono-synaptically facilitative via connections J or di-synaptically suppressive via connections W, respectively (see (C)). All targets are within a distance of several sampling grid spacings. The same connection pattern is translated and rotated for different CRF locations and orientations of the pre-synaptic cell. (C) An input bar at i and oriented at θ is directly processed by a corresponding pair of excitatory and inhibitory cells. Each cell models abstractly a local group of cells of the same type. The excitatory cell receives visual input and sends output to higher centers. The inhibitory cell is an interneuron. More details are provided in [25,27].

specific contextual influences including colinear facilitation and iso-orientation suppression. In addition, the model includes an activity normalization process (see below) which enables a general, orientation unspecific, finite range contextual suppression. The output from a model unit is its response to the stimulus within its CRF in the light of the intra-cortical interactions, and it thus depends on stimuli outside the CRF as well. The behavior of the model, i.e., the outputs from all units under visual inputs, thus depends strongly on the lateral connections, which are such that (1) they are consistent with the anatomical and physiological data [5,7,11,34] and (2) the model behavior exhibits the contextual influences (e.g., general and iso-orientation suppression and colinear facilitation) observed physiologically [14,17,37].

To focus on intra-cortical interactions, the model includes mainly layer 2–3 orientation selective cells. Cells influence each other via horizontal intra-cortical connections, transforming patterns of inputs to patterns of cell responses. At each location i there is a model V1 hypercolumn composed of 12 neuron pairs. Each pair (i, θ) consists of an excitatory (pyramidal) neuron and an inhibitory interneuron interconnected with each other, has CRF center i and a particular preferred orientation θ, and is called (the neural representation of) an edge or bar segment. Note that each model neuron models a local group of physiological neurons of similar properties, and that this group size depends on the neuron type. The excitatory cell receives the visual input; its output measures the response or salience of the edge segment and projects to higher visual areas. Based on experimental data [5,7,11,34], horizontal connections $J_{i\theta,j\theta'}$ (respectively $W_{i\theta,j\theta'}$) mediate contextual influences via monosynaptic excitation (respectively disynaptic inhibition) from pyramidal cells $j\theta'$ to $i\theta$ which have nearby but different CRF centers, $i \neq j$, and similar orientation preferences, $\theta \sim \theta'$. The monosynaptic excitatory connections J predominantly link colinear CRFs, whereas the disynaptic inhibitory connections W predominantly link non-colinear CRFs of similar orientations (Fig. 2(B)). The membrane potentials $x_{i\theta}$ and $y_{i\theta}$, of the excitatory and inhibitory cells, respectively, follow the equations (it is not necessary to understand these equations and their details in order to understand this paper, these details are included just for interested readers):

$$\dot{x}_{i\theta} = -\alpha_x x_{i\theta} - \sum_{\Delta\theta} \psi(\Delta\theta) g_y(y_{i,\theta+\Delta\theta}) + J_o g_x(x_{i\theta}) + \sum_{j\neq i,\theta'} J_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_{i\theta} + I_o$$

$$\dot{y}_{i\theta} = -\alpha_y y_{i\theta} + g_x(x_{i\theta}) + \sum_{j\neq i,\theta'} W_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_c$$

where $\alpha_x x_{i\theta}$ and $\alpha_y y_{i\theta}$ model the decay to resting potential, $g_x(x)$ and $g_y(y)$ are sigmoid-like functions modeling cells' firing rates in response to membrane potentials x and y, respectively, $\psi(\Delta\theta)$ is the spread of inhibition within a hypercolumn, $J_o g_x(x_{i\theta})$ is self excitation, I_c and I_o are background inputs, including noise. In addition, I_o for each unit $i\theta$ includes a suppressive, orientation unspecific, local activity normalization (suppressive) input (proportional to $-[\sum_{|j-i|\leqslant 2,\theta'} g_x(x_{j\theta'})]^2$, after the model by Heeger [8]) that is contributed by unit activities from a local neighborhood within 2 grid spacings. Finite range inhibition, which are orientation diffusive or unspecific, mediated by the inhibitory basket cells [15,16] could be the neural basis for the normalization. Visual input $I_{i\theta}$ persists after onset. The initial neural responses are mainly driven by feed-forward inputs $I_{i\theta}$. The activities are then modified by the contextual influences, which become apparent after about one membrane time constant. Since the horizontal connections are of finite range, if two neurons have no direct (mono-synaptic or disynaptic) connection between them but are both connected to a third neuron, they can exert contextual influences upon each other only after about two membrane constants after stimulus onset. In other words, contextual influences can propagate a long distance given long enough time and active intermediate neurons. Temporal averages of $g_x(x_{i\theta})$ are used as the model's output.

This V1 model had been previously constructed [25–27] to account for physiologically observed contextual influences, including iso-orientation suppression, weaker suppression under cross-orientation or random orientation surround [17], facilitation under colinear flanks [14], and enhanced responses to texture borders [6]. The model behavior is also consistent with human visual behavior such as contour saliency enhancement, texture segmentation, and visual searches, when the model is applied on global stimuli such as closed curves against random backgrounds, texture regions, or visual targets among distractors [26,27,29]. Our claim of consistency between our model and human psychophysics is under the assumption that V1 response, at least at its initial phase after stimulus onset (before the top–down feedback is significant), corresponds to the visual saliency of the underlying stimulus ([31]; note that stimulus saliency depends on, but is not the same as, stimulus contrast. See [10]). In this paper, we apply the model to figure–ground stimuli or those stimuli relevant to our topic concerned. Various strengths and weaknesses of the model on topics less related to that of this paper are discussed in more detail in Refs. [26–29]. The plotted region in each picture of this paper, of a visual stimuli or the response pattern, is often a small region of an extended image. The detailed model parameters (e.g., the synaptic weights, available in [26]; for readers interested in reproducing the results) are the same for all simulated examples except when noted.

3. Results

Fig. 3 shows that the model exhibits the figure–ground (border and interior), medial axis, and the border effects [28]. The highest responses are to the figure borders or the whole of a small figure against background. The responses to the medial axis are enhanced, but not so greatly as at the borders. These different response levels are to input bars of the same contrast, and are therefore solely due to contextual influences. The border effect is highly significant within a distance of about 2 texture element spacings from the border in the example of Fig. 3(C), let us call this the *effective border region*. It is mainly caused by the fact that the texture elements at the border experience weaker iso-orientation suppression because they have fewer iso-orientation neighbors than the texture elements far from the border. When the figure is sufficiently small (e.g., $\leqslant 4$ texture element spacings), all parts of it belong to the effective border region and are thereby awarded higher responses (Figs. 3(A) and 4(A)). Indeed, physiological experiments demonstrating the figure–ground (interior) effects [20–22,40] employ small figures. In larger figures, the stronger responses to the effective border region cause extra iso-orientation suppression to the adjacent (figure) texture bars outside the effective border region, let us call this the *border suppression region*. This is significant

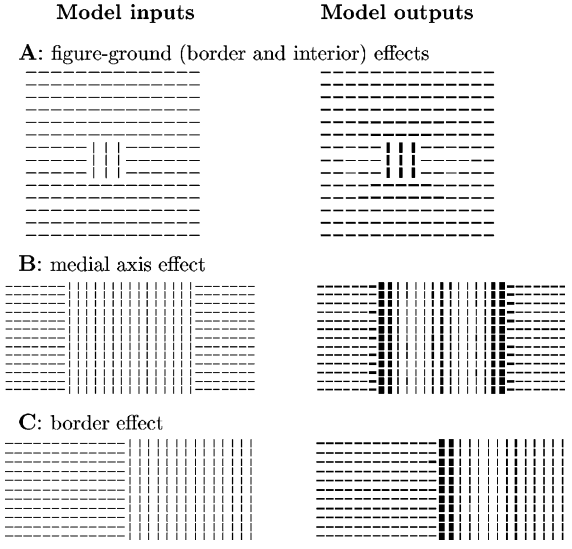

Fig. 3. Model behavior on the figure–ground, medial axis, and border effects. Figure and ground are defined by the orientation of the texture bars. The model inputs are composed of texture bars of equal contrast, the CRFs of the model units have roughly the same size as the bars; the model outputs plot the texture bars with their thicknesses proportional to the neural responses to the bars in the respective images (as in other figures of this paper). In (A) the average response to figure is more than 2 times that of the average response to background. In (B) the ratios of response levels to figure border, figure axis, and background are 4.3:2.2:1. In (C) the border response is 4 times that of the background response.

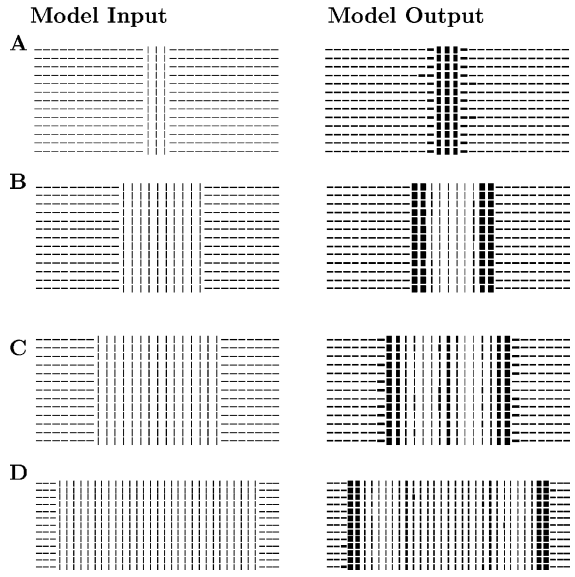

Fig. 4. Dependence on the size of the figure. The figure–ground effect is most evident only for small figures (A), and the medial axis effect is most evident (C) only for figures of appropriately finite sizes.

and is visible at the right side of the border in Fig. 3(C). This region can reach no further than the longest length of the lateral connections (mediating the suppression) from the effective border region. The (figure) texture bars further still from the border and adjacent to the border suppression region not only escape the stronger suppression from the border, but also experience weaker iso-orientation suppression from the weakened texture bars in the nearby border suppression region. As a result, a second and significant response or saliency peak appears—the ripple effect, at about 9 texture element spacings to the right of the texture border in Fig. 3(C). Let us call the distance (e.g., 9 texture element spaces in Fig. 3(C)) from the border to the secondary ripple the *ripple wavelength*, which should be of the same order of magnitude as the longest length of the lateral connections.

Note that the sizes of the effective border region, the border suppression region, and the ripple wavelength shown in our examples are for illustrative purposes and can differ quantitatively from their physiological counterparts, which should depend on the stimulus scale, as well as on the actual lengths and strengths of the lateral connections. Rockland and Lund [34] found that the horizontal connections are 2–3 mm in length in the primates, while Gilbert and Wiesel found corresponding lengths in the cats to be up to 4 mm [7] or 6–8 mm [41]. Since cells tend to connect to each other when they have similar receptive field sizes [32], and assuming that 1 mm of cortical distance corresponds to roughly one receptive field size, we estimate that horizontal connections link cells displaced by 2–3 receptive field sizes for monkeys and, depending on which experimental data to rely on,

4, or 6–8, or 2–5 [32] receptive field sizes for cats. (In comparison, the horizontal connection length in our model is nine times the receptive field size.) These length values correspond to the ripple wavelength. The sizes of the border region and the border suppression regions should be scaled down from these values. Given a visual stimulus, one can obtain its dominant scale and spatial frequency value to arrive at the receptive field sizes of the cells that are most excited by the stimulus. The size of the ripple wavelength (or border suppression region) can then be estimated as the corresponding multiples, 2–3 times for monkeys and 2–8 times for cats, of the receptive field size. Note that all these estimates should be taken with caution, since the length of the horizontal connections is not the only determining factor, the strengths of the connections for all distances up to the maximum length is the actual underlying factor dictating the outcome.

Fig. 4 summarizes the model dependence on the sizes of the figures. Fig. 4(B) shows that the response to the center of the figure is smallest when the figure size is such that the center is in the border suppression regions from both borders. When the size of the figure is about twice the ripple wavelength, the ripples or the secondary saliency peaks from the two opposite borders superpose at the center of the figure to give the medial axis effect in Fig. 4(C) (the same as Fig. 3(B)). This reinforces the saliency peak at this medial axis since it has two border suppression regions (from two opposite borders), one on each side of it, as its context. For even larger figures, the medial axis effect diminishes (Fig. 4(D)), and the response level to the center approaches that to the background (Fig. 6(B)) for very large figures. Physiologically, the medial axis effect is also observed only for certain figure sizes, about four to six times of the cell's receptive field in monkeys [23], for a given cell. This agrees with the fact that the lengths of the horizontal connections, i.e., the length of the ripple wavelength, in the monkeys is about 2–3 mm [34], i.e., two to three times of receptive field size.

Fig. 5 demonstrates that the shapes and texture elements of the figure also play important roles. For instance, Figs. 5(A) and (E) and 4(C) show that the position of the border highlight is biased towards the texture region where texture elements are parallel to the border, whether or not this texture region is the designated figure. This is also observed physiologically (see Fig. 12 in [23]). This observations can be understood as follows: for these most highlighted border elements, not only do they have fewer iso-orientation neighbors to suppress them, they in addition have more colinear neighbors (than the border elements in the neighboring texture) to facilitate them. In fact, such a bias or asymmetry at the border can lead to a psychophysical bias in the perceived location of the texture border [33], simply because the response highlights do not distribute symmetrically across the border. Additionally, the colinear facilitation causes a texture border more salient when the texture elements are parallel to the border, as observed psychophysically [39]. Analogously, the medial axis effect is also stronger in a texture region when the medial axis is parallel, rather than orthogonal, to the texture element. This is also observed in physiological data (see Fig. 12 in [23]). The more complex spatial distribution of the border effect and the medial axis effect in Fig. 5(B) and (C) can now be understood as results of the number of contextual neighbors that are iso-oriented and/or colinear, whether or not the texture region concerned is designated as figure.

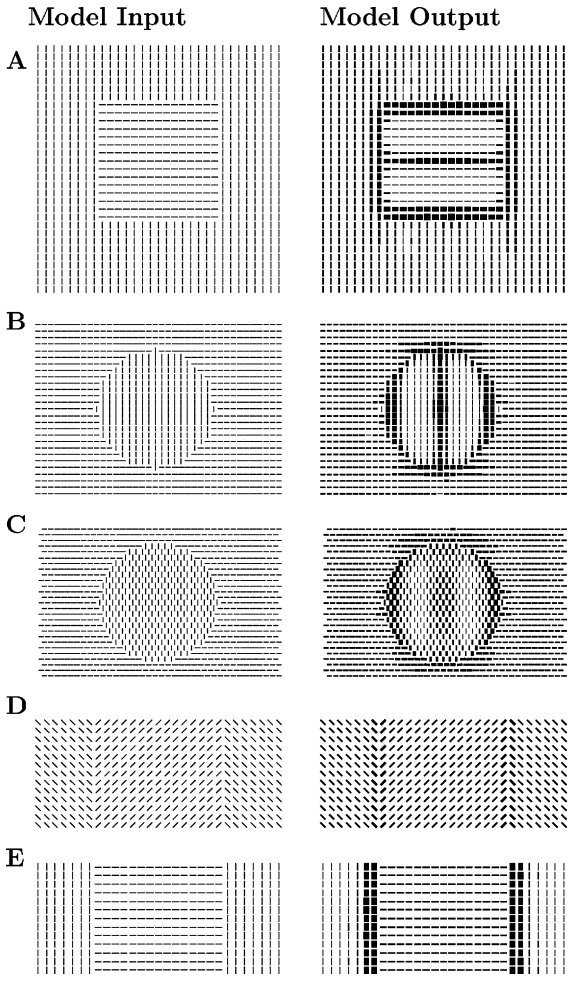

Fig. 5. Dependence on the shape of the figure and the texture features.

It is now apparent that the effects arising from the texture border should stay the same qualitatively when no neighboring or background texture regions exist, since the most significant contextual influences on a cell arise from cells responding to the same texture region. Our insight can then be applied to understand a V1 cell's response to increasingly large grating patches, as a grating patch can be viewed as a regular texture (without

background or surround textures) while the cell's CRF is centered at the center of the grating patch. For small grating sizes (first 2 grating patches in Fig. 6(A)), the cell's CRF is within the effective border region of the whole patch, and thus gives vigorous responses. When the grating is large enough so that the cell's CRF is at the border suppression region of the grating boundary the third to sixth grating patches in Fig. 6(A), its response is very much suppressed. The suppression diminishes as the grating patch becomes larger, the cell's response can be strong again when the CRF is at the location of the secondary ripple from the grating boundary, when the grating radius is roughly the ripple wavelength (in this example at a grating radius around 9 grid units or CRF sizes). As the grating size increases further, the response will asymptote towards a level corresponding to a response to an infinitely large grating. Accordingly, a curve of response vs. grating size can be obtained, as in Fig. 6(C). The summation zone and suppression zone observed physiologically on the size tuning of the cell's responses [36] can be identified in this curve for grating radius smaller than the ripple wavelength. The model however predicts a second rise, a stronger response at a grating radius equal to the ripple wavelength. This could be possibly missed by physiological experiments when very large grating sizes are not tested.

Now we are ready to explain the elusive phenomena of "cross-orientation facilitation". In this, the neural response to a figure patch (a grating) is enhanced when

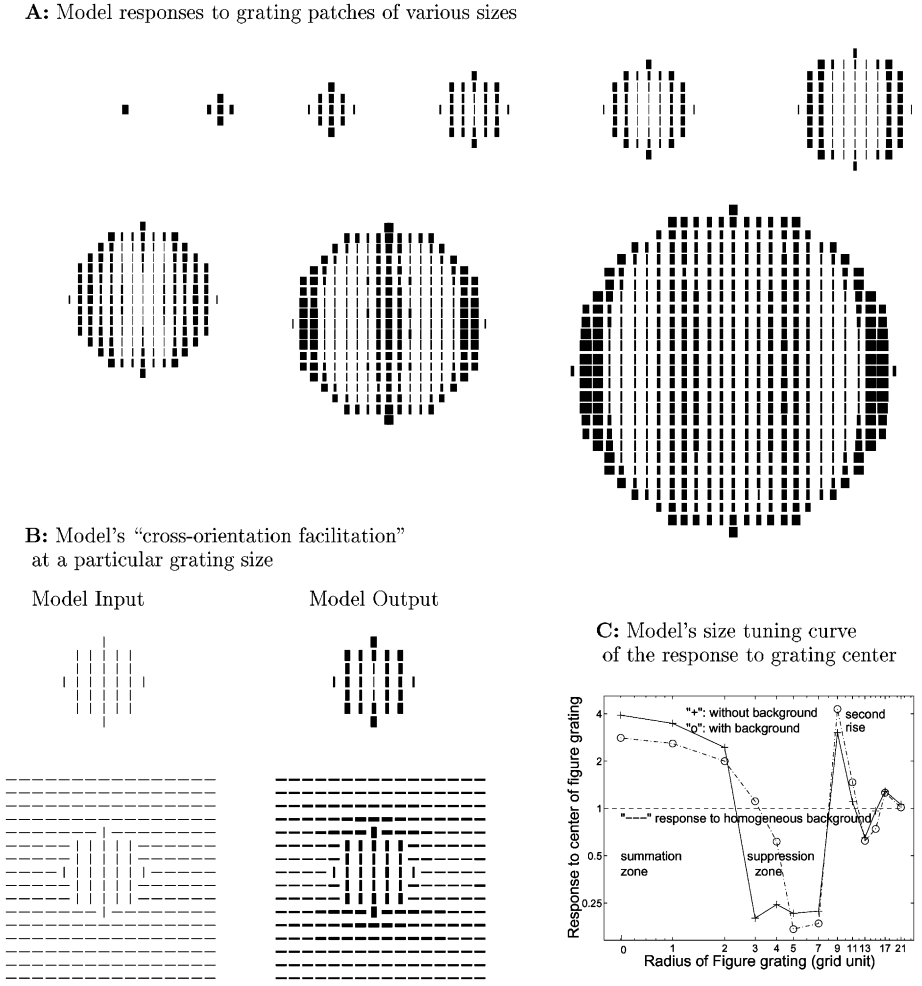

Fig. 6. (A) Model's response to grating patches (without background) of various sizes. All gratings differ only in size but not in spatial frequency, to fit plots in picture, the large gratings are not plotted in scale. (B) Cross-orientation dis-inhibition—the suppression of the figure center by the figure border is reduced by the general suppression of the figure border by the background of an orthogonal grating. (C) Response (normalized to the response level to homogeneous background) to the center of the figure with and without the background (of an orthogonal grating) for different figure radii. Note that a patch radius indicated in the plot is the extension from the center of the CRF to the periphery of the patch, measured in the unit of the distance between neighboring nodes in the model grid. Hence, a radius = 0 in the plot does not mean a grating patch of size zero, but a patch that includes only one grid node, and hence has zero extension to any other grid nodes. The example in (B) is at figure radius = 3, and Fig. 5(B) is at figure radius = 9 when the colinear facilitation at the medial axis is itself overwhelming. For very large figures, the response level to the center of the figure approaches that to homogeneous background.

the figure is surrounded by a background of an orthogonal grating. Although this has been observed by Sillito et al. [37], other researchers have only found suppression from the cross-orientation background. According to our model, when the center of a lone grating is in the border suppression region, it evoked responses in a cell will be suppressed by the neighboring cells which are responding vigorously to the border of this grating. When a background grating is added to the stimulus, its evoked responses provide a general, orientation unspecific, suppression to the neighboring cells responding to the central grating—in particular, cells responding to the border of the central grating will be suppressed. This suppression of the border highlights by the surround grating dis-inhibits the border suppression region at the center of the central grating, and thereby produce an apparent "cross-orientation facilitation" (see Fig. 6(B)). Therefore, this phenomena should be experimentally elusive since a critical stimulus parameter is figure size, which should be such that the figure center is at the border suppression region (see Fig. 6(C)). In our model which has horizontal connections up to 9 times the receptive field size, the circular figure size for "cross-orientation facilitation" has a radius about 3–4 times the receptive field. Since monkeys and cats have horizontal connections up to 3 and 3–8 mm respectively [7,32, 34,41] physiological "cross-orientation facilitation" is most likely seen in figures with radius roughly 0.5–1.5 times of the cell's receptive field for monkeys (or 0.5–4 times for cats). While Sillito et al. [37] showed a figure in which a cell exhibited "cross-orientation facilitation" when the size of the central grating is optimal (i.e., the grating leads to maximum summation when presented alone), they showed in a later and more detailed study that cells tend to exhibit "cross-orientation facilitation" when the central grating is larger than the CRF [13,24]. Another model [2] provided an alternative account of the "cross-orientation facilitation" by considering the neural dynamics between excitatory and inhibitory cells. However, this account cannot explain the size dependency of the phenomena since the model concerned includes only the orientation dimension but not the spatial dimension (thus the size parameter) in the visual stimulus.

A closed contour is also very salient because of the colinear facilitation between the contour segments [4,14,18,25]. It can thus influence the image area that it surrounds in just the way that the border of a figure does to the figure region enclosed. Consider the case that the background (both inside and outside the contour) consists of randomly oriented bars. The contour effect is weaker than the figure border effect because the general, orientation unspecific, suppression from the contour to the random background is weaker than the iso-orientation suppression from the border to the figure interior of an iso-orientation figure region [17]. The contour effect is further submerged by the noisy contextual facilitation and suppression arising from the accidental alignments between the random background bars (Fig. 7). However, the contour effect can be made evident by averaging out the randomness (with many trials of different random background stimulus), and it can then be seen to depend on the size of the contour just as the border effect depends on the size of the figure. The center of a closed contour is less salient in an appropriately small contour but more salient in an appropriately large one (Fig. 7). Let there be a stimulus bar at the center of the contour. When this bar is less salient, it evokes a weaker V1 response. A contrast increment of

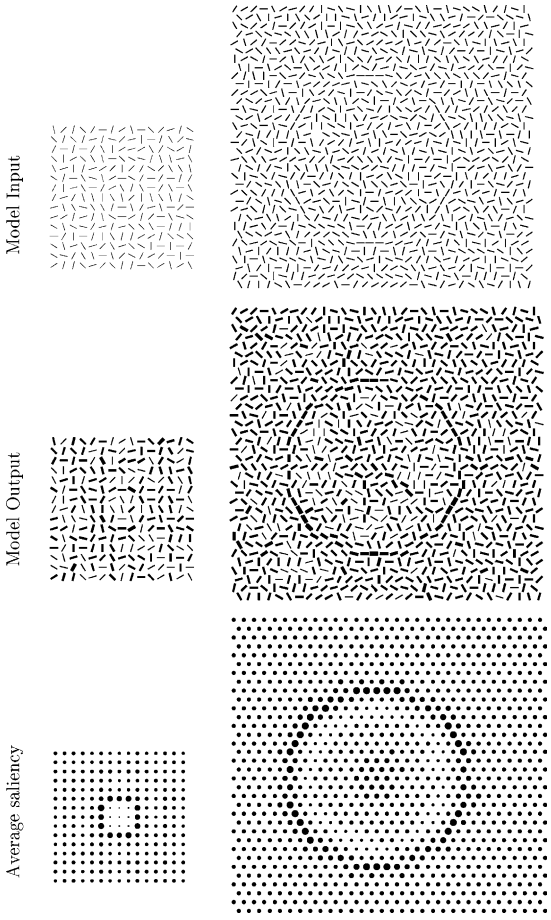

Fig. 7. Closed contours (circles) in backgrounds of randomly oriented bars can influence the saliencies within them. Model inputs show two examples, each contains randomly oriented bars except those that connect to form a circle near the image center. In the smaller image, the circle is not as conspicuous because it is quite small, its diameter is only 4 grid units long. Model outputs are shown for the corresponding example inputs. Note that bars that form the circles are more likely to induce higher outputs due to their alignment with the contextual bars. The saliency maps are for the corresponding stimuli types, averaged over various inputs with the same contour elements but different random backgrounds. At each grid point, the saliency value is visualized by the radius of the plotted circle, which is proportional to the average model outputs at that spatial location.

this bar would cause a more significant change in the V1 response (than that if the baseline response to the bar were stronger). Assuming that contrast discrimination of a stimulus is carried out by detecting significant changes in the evoked neural responses relative to neural noise (which usually scales with the responses), we conclude that it is easier to perform contrast discrimination on a bar in the suppressed center of a contour. This is just like in Weber's law where it is easier to discriminate contrast when the baseline contrast is low. Indeed, psychophysically, contrast discrimination threshold are low at the center of a closed contour, in a way that depends on the size of the contour (relative to the length of the (gabor) bars) [18,19].

4. Discussion

Our model suggested and predicted that (1) V1 mechanisms can account for the particular kinds of figure–ground effects observed in the physiological experiments by Lamme [20], Zipser et al. [40], Lee et al. [23], and Lamme et al. [22], including interior effects, in particular, the medial axis effect, and the border effect, observed physiologically, (2) the interior effects, including the medial axis effect, are weaker than the border effect, and, most importantly, (3) the interior effects are products of the border effect and are only seen for certain figure sizes. By comparison, the border effect is robust. The model makes the testable prediction that the figure–ground interior effect away from borders should disappear when the figure is large enough. We therefore suggest that feedback from higher visual areas is not necessary for these effects of figure–ground and medial axis observed in these particular experiments, although, of course, we cannot exclude the possibility that it also contributes.

Shortly after the model predictions were published [28,30], they were confirmed by physiological experiments. In particular, Rossi et al. [35] showed that the figure–ground effects in V1 were only present when the figure is small enough, or when the CRF of the recorded cell is close enough to the texture border, and that the V1 neurons appear to signal texture boundaries rather than figures per se. Húpe et al. [12] also showed the supporting evidence that the response modulations in the V1 cells by texture surround do not depend on feedbacks from V2.

Our findings are consistent with the observation that the border effect has a shorter latency (10–20 ms after the initial response) [6,22,23] than the interior effects (30–40 ms after the initial response) [20,22,23,40]. If influences from given contextual activities take 10–20 ms to build up, the initial border effect should arise at about this latency after the initial feed-forward-driven cell responses. Since the interior effects depend on the border effect, it should take additional 10–20 ms to become evident. The border effect in our model also has a relatively shorter latency for the same reason.

Computationally, marking the border is sufficient for the purpose of segmenting figure from ground. Since it is often a subjective decision as to which is figure and which is ground, it is neither necessary nor desirable to highlight one arbitrary region against another, at least under pre-attentive conditions. Our model of V1 was originally proposed to account for pre-attentive contour enhancement and texture segmentation [25–27]. Contextual influences were proposed to detect locations where homogeneity in the input image breaks down, and make these locations more salient by stronger responses to them. These highlights mark candidate locations for boundaries of image region (or object surface), for smooth contours and small figures against backgrounds, serving pre-attentive segmentation.

4.1. From borders to objects: understanding shine through and inheritance

We can deduce from the above arguments that any local salient peak marks a border of a (homogeneous) surface or object, or the whole of an object of a sufficiently small size. Hence, a single saliency peak in a whole image should mark a small object, while two saliency peaks separated along a spatial dimension in a whole image could mark the two borders of an extended (one dimensional) surface (that extends infinitely in the other spatial dimension). Furthermore, three separate saliency peaks could suggest the presence of a small object in addition to an extended background surface. We can apply this concept to understand the recently observed psychophysical phenomena called inheritance and shine through [9]. It is observed that, when a vernier stimulus (with a horizontal offset) is presented for a short time (20 ms) and followed immediately (for 300 ms) by a (vertical) line grating composed of an upper and a lower halfs aligned with each other, the vernier is invisible but the whole grating is perceived to have the vernier offset between its upper and lower halfs. This phenomena is termed inheritance, and is present when the grating has no more than 5 lines or grating elements. However, when the grating is larger, with more than 7 grating elements, the percept becomes a vernier superposed on, or shining through, the grating which does not show any offset (see Fig. 8). The discrimination of the offset direction, in inheritance and shine through, depends on the size of the grating and is poorest for a five element grating. When simulated with the corresponding stimuli, our model produces response patterns consistent with the psychophysical phenomena, see Fig. 8. In particular, when the input grating has fewer elements, e.g., three elements, the model responded with a single saliency peak (among the 3 grating elements) in each

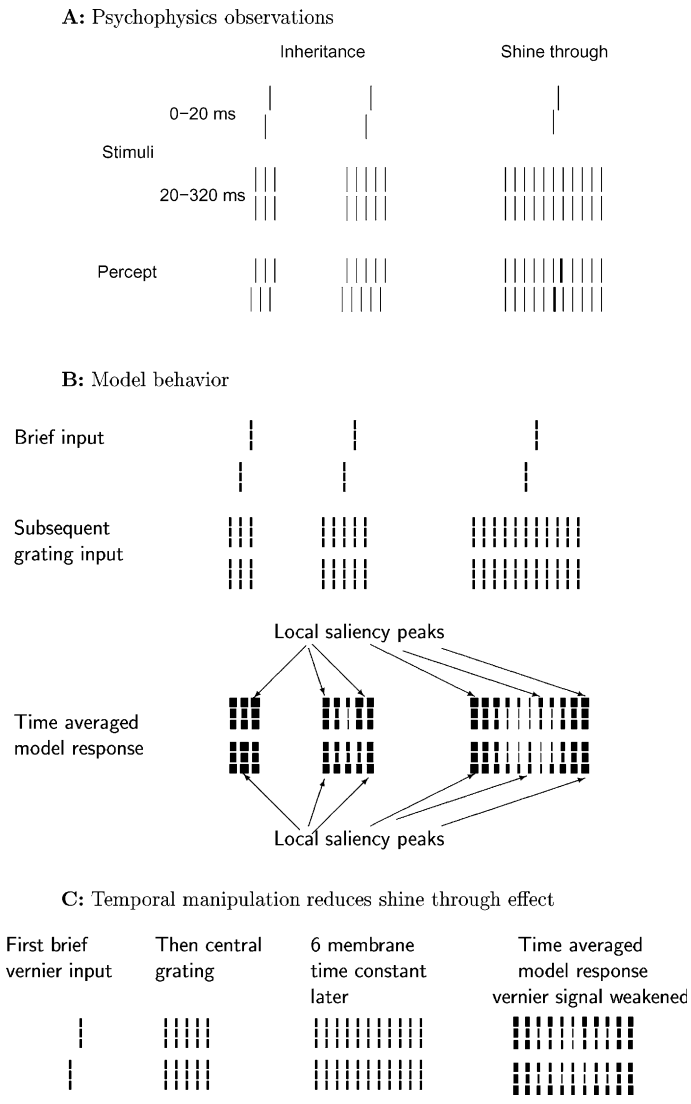

Fig. 8. (A) Inheritance and shine through observed psychophysically by Herzog and Koch [9], a briefly presented vernier followed by a grating gives the percept of inheritance when the grating has no more than five elements, but a percept of shine through when the grating has no less than seven elements. The percept of the vernier offset, whether in inheritance or shine through, is weakest when there are only 5 grating elements. (B) The model simulation of inheritance and shine through with 3, 5, and 11 grating elements. Each bar in psychophysics is simulated as composed of three shorter segments, each excites an underlying cell with a corresponding CRF. If fewer segments were used for each bar in the simulation, the shine through phenomena would weaken in the model behavior, consistent with the psychophysical observation that the phenomena were observed mainly for long enough bars in the stimuli (Herzog, private communication 2002). When there are only 3 or 5 grating elements, the model responded with only one or two (local) saliency peaks in the horizontal dimension in both the top and bottom half of the gratings, suggesting a percept of a single small object or grating (for the three element grating) or extended grating surface (for the five element grating), since each saliency peak signals the border of a single object. The horizontal offset between the saliency peaks is thus assigned to the offset of the whole single (inferred or perceived) object, i.e., the grating. Note that the top and bottom saliency peaks in the case of the three element grating are offset horizontally in the direction of the original vernier, and this offset is assigned to the perceived single object (grating). The horizontal offset between the top and bottom saliency peaks in the case of the five element grating are quite ambiguous, consistent with the poor discrimination performance of the offsets observed psychophysically. When there are 11 grating elements, 3 saliency peaks arise in both the top and bottom halfs of the grating, giving the percept of two objects: a grating without offset and a (superposed) single vernier with the correct offset. C: If the central five elements of the grating was onset 6 membrane time constants earlier than the peripheral grating elements, the saliency peaks evoked by the vernier are much less prominent, giving reduced shine through as observed psychophysically.

(upper or lower) half of the response pattern. Consequently, a single, small, object (grating) is inferred (in the horizontal dimension) in each half of the image, and the horizontal offset between the upper and lower saliency peaks were assigned (inherited) as between the correspondingly inferred objects which are the upper and lower halfs of the grating. When the input grating has five elements, two saliency peaks appeared in each (upper and lower) half of the response pattern, leading to a percept of a single extended grating with two bor-

ders in each half of the image. The horizontal offset between the upper and lower saliency peaks are ambiguous in the model response, and this is consistent with the observation that offset discrimination is poorest at this grating size [9]. When the input grating has no less than seven elements, e.g., 11 elements, the model responded with three local saliency peaks in each half of the response pattern. Thus, in each half of the image, the inferred percept is a single small object corresponding to the central saliency peak, superposed on an extended grating with two borders corresponding to the left and right saliency peaks. Since the upper and lower saliency peaks evoked by the borders of the gratings are aligned vertically, the grating does not appear offset between the two halfs. Meanwhile, the unambiguous horizontal offset between the upper and lower central saliency peaks is inferred as the horizontal offset of the vernier superposed on the grating.

The mechanisms of how the subjects decode the offset and assign it to the vernier or to the whole grating are supposedly carried out in higher visual areas. There, the spatial and temporal pattern of the visual stimulus, such as the offset value, would be decoded from the neural responses. However, regardless of the actual mechanisms, the decoding should be consistent, at least in the maximum likelihood sense, with the inference on the number of objects in the scene. This means, if there are multiple hypotheses about the visual scene for a given neural response pattern, the most likely hypothesis will be the perceptual outcome. Furthermore, the number of objects in the most likely hypothesis should be consistent with the number of saliency peaks in the neural responses. Hence, when an offset is detected and a single (homogeneous) object is inferred in the scene, the only consistent solution is to assign the offset to all parts of the object (otherwise it is not an homogeneous object by definition), i.e., to all the elements of the grating. This means, the subjects should not be able to tell whether the offset belongs to the vernier or is actually present in the grating. This is consistent with experimental data [9]. In the shine through case, when three saliency peaks are observed in each (upper and lower) half of the response pattern, two different hypotheses (among others) about the scene are possible. One is an offset vernier superposed on an non-offset grating—shine through. Another is to interpret the saliency peak evoked by the vernier as the border between two (left and right) adjacent gratings in both the upper and lower halfs of the scene. The maximum likelihood decoding would clearly favor the first hypothesis, since the second one would require an accidental or low likelihood vertical alignment between the two left borders of the two (upper and lower) gratings on the left and between the two right borders of the two gratings on the right.

We can understand the neural responses in the inheritance and shine-through phenomena by the V1 mechanisms as simulated in our model. Under a vernier stimulus, the responding neurons are subject to colinear facilitation but little iso-orientation suppression from each other due to the particular horizontal connection pattern (see Fig. 2(B)). Twenty milliseconds of the vernier exposure is sufficient for the colinear facilitation to take effect [14] and hence the response to this vernier should be quite strong, even after the initial transient to the stimulus onset. Under a grating stimulus, neurons responding to the neighboring grating elements inhibit each other by iso-orientation suppression. Many of the neurons initially excited by the vernier continue to be active by participating in responses to the subsequent grating. They have a head start in their activities compared to other responding neurons, and are thus biased to be the winners in the mutual iso-orientation suppression battle between neurons responding to neighboring grating elements. When the grating is small enough, all parts of it belong to the border region, and a single saliency peak arises from each half of the response pattern. However, the location of the saliency peaks in response to the grating are now biased to be the location of the vernier components. This results in inheritance. When the grating has five elements, the location of the vernier falls in the border suppression region of the grating borders. The neural activities initiated by the vernier are now strongly suppressed. Hence, the subjects have difficulties discriminating the vernier offset. When the grating is large enough, the location of the vernier is beyond the suppression region of the borders. Thus, among the neurons responding to the grating region near the original vernier but away from the grating borders, the activity imbalance initiated by the head start responses to the vernier can be sustained. This results in local saliency peaks corresponding to the vernier, and a perceptual shine through. Note that the vernier bars should be long enough to induce sufficient colinear facilitation, which contributes to the strength of the activity head start. In our example, each vernier bar is as long as three times the length of the receptive field. Our model predicts that inheritance and shine through will not be as effective if the vernier segments are too short (two times the length of the CRF in our model). This prediction agrees qualitatively with experimental data (Herzog, unpublished data, private communication 2002). Note that our prediction of the vernier length is with respect to the scale of the stimulus, i.e., relative to the width of the vernier segment. In the experiment, the vernier double bars were 21 arc minutes [9]. The lengths could be longer or shorter when the whole stimulus pattern is scaled up or down. A simpler V1 model [3], omitting the vertical spatial dimension and orientation tuning, can also account for much of the phenomena, although it cannot account for the dependence on the length of the vernier since the model is only one dimensional (horizontal). We should note that

contextual influences take about 10–20 ms to build up [14,17]. Hence, changing the temporal characters of the stimulus will affect perception. For instance, (after the vernier) if the central five elements of the grating onset before the periphery elements, the vernier signal can be suppressed by the strong borders of the central (five element) grating. This suppression is manifest once the onset asynchrony between the central and peripheral grating elements exceeds 10–20 ms, the time for contextual influence to take effect. This should reduce or eliminate the head start signal from the vernier by the time the peripheral grating elements are presented. Indeed, Herzog and Koch [9] observed deteriorating shine through performance as the onset asynchrony increases from 10 to 60 ms. Simulation results from our model agree with their observations, see Fig. 8(C).

It is apparent that the current version of the V1 model is very minimal. In particular, the model behavior can be quite different from reality. One example is the following. While the medial axis effect and the figure interior (non-medial axis) effect can exist simultaneously (i.e., given a single stimulus pattern) in physiology [23], they do not do so in our model given a single figure size (see Fig. 4), at least when looking at the temporally averaged model responses. (Due to temporally desynchronized nature of the responses to different parts of the stimulus pattern, simultaneous mexial axis and interior effects are possible in the model within particular time windows when the responses to the background are low). A very probable cause for this is that the strengths of the horizontal connections depend on the distances between the linked cells in a very different manner in the model from that in reality. The spatially very sparse sampling in our model also prevented the model from analyzing how V1's behavior changes with small changes in the vernier offset in the shine through and inheritance effects.

To summarize, figure–ground effects observed physiologically in V1 are the byproducts of the robust border effect. The border effect arises from the computational need to signal or highlight salient image locations, in particular, the border between image regions. The underlying neural mechanism is the intra-cortical interactions that causes the neural response of individual V1 cells to depend on both the direct input in the CRF and the contextual stimuli nearby. These neural mechanisms are manifested in various phenomena, including figure–ground and medial axis effects. Fig. 5 suggests that, even if one only considers the region border, the degree of highlights depends on the border properties, e.g., the alignment of the texture elements with the region border, rather than whether a particular region is assigned "figure" or "ground". However, it is objective and desirable always to highlight a very small region against a larger region, in our model by the pre-attentive mechanisms in V1. The higher V1 responses make the corresponding locations more salient, and the V1 can thus produce a saliency map of the input [31]. The higher saliency of a smaller region against a homogeneous background could be the reason why smaller regions tend to be treated as the figures against larger backgrounds. This framework of highlighting the object boundaries or small objects for segmentation can be applied to understand some seemingly complex psychophysical phenomena such as shine through and inheritance. The same framework has also been applied successfully to understand how the ease or difficulty of a visual search task depends on image features and spatial configurations, assuming that the ease of a search is determined by the degree that the target of the search is highlighted relative to that of the distractors or background [27,31]. The "figure–ground effects" observed in V1, and various other byproducts of the V1's computation for a saliency map for pre-attentive segmentation, are especially helpful to diagnose the underlying intracortical interactions.

Acknowledgements

I am very grateful to Peter Dayan, Michael Herzog, and two anonymous reviewers for careful readings of various versions of the manuscript and for their very helpful comments. This work is supported by the Gatsby Foundation.

References

[1] H. Blum, Biological shape and visual science, J. Theor. Biol. 38 (1973) 205–287.
[2] V. Dragoi, M. Sur, Dynamic properties of recurrent inhibition in primary visual cortex: contrast and orientation dependence of contextual effects, J. Neurophysiol. 83 (2) (2000) 1019–1030.
[3] M.H. Herzog, U. Ernst, A. Etzold, C. Eurich, Local interactions in neural networks explain global effects in the masking of visual stimuli, Neural Comput. 15 (9) (2003) 2091–2113.
[4] D.J. Field, A. Hayes, R.F. Hess, Contour integration by the human visual system: evidence for a local 'associat ion field', Vision Res. 33 (2) (1993) 173–193.
[5] D. Fitzpatrick, The functional organization of local circuits in visual cortex: insights from the study of tree shrew striate cortex, Cereb. Cortex 6 (3) (1996) 329–341.
[6] J.L. Gallant, D.C. van Essen, H.C. Nothdurft, Two-dimensional and three-dimensional texture processing in visual cortex of the macaque monkey, in: T. Papathomas, C. Chubb, A. Gorea, E. Kowler (Eds.), Early Vision and Beyond, MIT press, 1995, pp. 89–98.
[7] C.D. Gilbert, T.N. Wiesel, Clustered intrinsic connections in cat visual cortex, J. Neurosci. 3 (5) (1983) 1116–1133.
[8] D.J. Heeger, Normalization of cell responses in cat striate cortex, Visual Neurosci. 9 (1992) 181–197.
[9] M.H. Herzog, C. Koch, Seeing properties of an invisible element: feature inheritance and shine-through, Proc. Natl. Acad. Sci. USA 98 (2001) 4271–4275.
[10] R.F. Hess, S.C. Dakin, D.J. Field, The role of "contrast enhancement" in the detection and appearance of visual contours, Vision Res. 38 (6) (1998) 783–787.

[11] J.A. Hirsch, C.D. Gilbert, Synaptic physiology of horizontal connections in the cat's visual cortex, J. Neurosci. 11 (6) (1991) 1800–1809.
[12] J.-M. Húpe, A.C. James, P. Girard, J. Bullier, Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2, J. Neurophysiol. 85 (1) (2001) 146–163.
[13] H.E. Jones, W. Wang, A.M. Sillito, Spatial organization and magnitude of orientation contrast interactions in primate V1, J. Neurophysiol. 88 (5) (2002) 2796–2808.
[14] M.K. Kapadia, M. Ito, C.D. Gilbert, G. Westheimer, Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys, Neuron 15 (4) (1995) 843–856.
[15] Z.F. Kisvarday, D.-S. Kim, U.T. Eysel, T. Bonhoeffer, Relationship between lateral inhibitory connections and the topograph of the orientation map in cat visual cortex, Euro. J. Neurosci. 6 (1994) 1619–1632.
[16] Z.F. Kisvarday, E. Togh, M. Rausch, U.T. Eysel, Orientation-specific relationship between populations of excitatory and inhibitory lateral connections in the visual cortex of the cat, Cerebral Cortex 7 (1997) 605–618.
[17] J.J. Knierim, D.C. van Essen, Neuronal responses to static texture patterns ion area V1 of the alert macaque monkeys, J. Neurophysiol. 67 (1992) 961–980.
[18] I. Kovacs, B. Julesz, A closed curve is mach more than an incomplete one: effect of closure in figure–ground segmentation, Proc. Natl. Acad. Sci. USA 90 (1993) 7495–7497.
[19] I. Kovacs, B. Julesz, Perceptual sensitivity maps within globally defined visual shapes, Nature 370 (1994) 644–646.
[20] V.A. Lamme, The neurophysiology of figure–ground segregation in primary visual cortex, J. Neurosci. 15 (2) (1995) 1605–1615.
[21] V.A.F. Lamme, K. Zipser, H. Spekreijse, Figure–ground signals in V1 depend on consciousness and feedback from extra-striate areas, Soc. Neuroscience Abstract 603.1, 1997.
[22] V.A. Lamme, V. Rodriguez-Rodriguez, H. Spekreijse, Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey, Cereb. Cortex 9 (4) (1999) 406–413.
[23] T.S Lee, D. Mumford, R. Romero, V.A.F. Lamme, The role of the primary visual cortex in higher level vision, Vision Res. 38 (1998) 2429–2454.
[24] Z. Li, H.E. Jones, W. Wang, A.M. Sillito, Modeling orientation and scale dependent surround suppression and facilitation in striate cortex, Presented at Annual Meeting of Society for Neuroscience, Abstract Number 285.19, San Diego, 2001.
[25] Z. Li, A neural model of contour integration in the primary visual cortex, Neural Comput. 10 (4) (1998) 903–940.
[26] Z. Li, Visual segmentation by contextual influences via intracortical interactions in primary visual cortex, Network, Comput. Neural Syst. 10 (1999) 187–212.
[27] Z. Li, Contextual influences in V1 as a basis for pop out and asymmetry in visual search, Proc. Natl. Acad. Sci. USA 96 (1999) 10530–10535.
[28] Z. Li, Can V1 mechanisms account for figure–ground and medial axis effects?, in: S.A. Solla, T.K. Leen, K.-R. Muller (Eds.), Advances in Neural Information Processing Systems 12, MIT Press, Cambridge, MA, 2000, pp. 136–142.
[29] Z. Li, Pre-attentive segmentation in the primary visual cortex, Spatial Vision 13 (2000) 25–50.
[30] Z. Li, J. Hertz, Multiple zones of contextual surround for V1 receptive fields, Annual Meeting of Society for Neuroscience, Abstract number 211.10, 2000.
[31] Z. Li, A saliency map in primary visual cortex, TRENDS Cogn. Sci. 6 (1) (2002) 9–16.
[32] C.Y. Li, W. Li, Extensive integration field beyond the classical receptive field of cat's striate cortical neurons—classification and tuning properties, Vision Res. 34 (18) (1994) 2337–2355.
[33] A. Popple, Z. Li, Testing a V1 model—perceptual biases and saliency effects from pre-attentive segmentation, Presented at the First Annual Meeting of the Vision Science Society, Sarasota, Florida, USA, May 2001, J. Vision, 1(3), 148a, http://journalofvision.org/1/3/148, DOI 10.1167/1.3.148.
[34] K.S. Rockland, J.S. Lund, Intrinsic laminar lattice connections in primate visual cortex, J. Comp. Neurol. 216 (1983) 303–318.
[35] A.F. Rossi, R. Desimone, L. Ungerleider, Contextual modulation in primary visual cortex of macaques, J. Neurosci. 21 (3) (2001) 1698–1709.
[36] M.P. Sceniak, D.L. Ringach, M.J. Hawken, R. Shapley, Contrast's effect on spatial summation by macaque V1 neurons, Nat. Neurosci. 2 (8) (1999) 733–739.
[37] A.M. Sillito, K.L. Grieve, H.E. Jones, J. Cudeiro, J. Davis, Visual cortical mechanisms detecting focal orientation discontinuities, Nature 378 (6556) (1995) 492–496.
[38] E.L. White, Cortical Circuits, Birkhauser, 1989.
[39] S. Wolfson, M.S. Landy, Discrimination of orientation-defined texture edges, Vision Res. 35 (20) (1995) 2863–2877.
[40] K. Zipser, V.A. Lamme, P.H. Schiller, Contextual modulation in primary visual cortex, J. Neurosci. 16 (22) (1996) 7376–7389.
[41] C.D. Gilbert, T.N. Wiesel, Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex, J. Neurosci. 9 (7) (1989) 2432–2442.