Global effects in Figure/Ground segregation by a model with only local interactions

N. Rubin, Center for Neural Science, New York University
M. C. Pugh, Department of Mathematics, University of Toronto
Corresponding author: Nava Rubin, nava@cns.nyu.edu

Abstract

Figure/Ground segregation is a fundamental problem in visual processing. Edges often arise because of occlusion, along the bounding contour of the occluding object. Having detected an edge, the visual system therefore has to decide whether it is there due to occlusion, and if so, which side of it belongs to the front surface ("the Figure"). This is known as the *border ownership* problem. Determining the border ownership of an edge cannot be done locally: it requires integrating information from an image region which contains the entire Figure or large portions of it. Thus, a neuron whose receptive field is small compared to the Figure cannot compute border ownership of an edge of that Figure in isolation. Recently it was reported that the responses of V2 cells to an edge can be strongly modulated by the polarity of border ownership of the edge (Zhou et al. 2000). These effects must therefore be the product of computations done by a network of neurons. One possibility is that such networks rely on direct long-range connections (of the scale of the Figure) and/or feedback from areas with cells of large receptive fields. The model presented here offers an alternative. It is shown that network of small receptive field cells which interact only locally can nevertheless compute Figure/Ground segregation. The long-range effects emerge as a result of iterative propagation of information through a cascade of short-range connections. The model produces results similar to those observed perceptually: it prefers enclosed, convex and/or smaller regions as the Figure. Our results suggest that considerable portions of Figure/Ground segregation can be accomplished by early visual cortex, without a need for feedback from higher areas.

Table of Contents

Section 1: Introduction, pp 2–4
Section 2: The model, pp 4
§2.1: Formulating the problem, pp 6–10
§2.2: Guiding principles, pp 10–11
§2.3: Network and mathematical implemention
§2.3.1: Luminance edges and Figure/Ground transitions, p 12
§2.3.2: Generating the candidate organizations: the cost function $\mathcal{E}$, pp 12–15
§2.3.3: Identifying undesired candidate organizations: the Figural entropy, pp 15–16
Section 3: Simulation results, pp 16–23
Section 4: Discussion, p 23
§4.1: Timing considerations, pp 23–24
§4.2: Region-based processes in physiology and perception, pp 24–25
§4.3: Probability and neural representation, pp 25–26
§4.4: Further extensions to the model, pp 26–27
Section 5: References, pp 27–31
Appendix A: Computing **P**, pp 31–32
Appendix B: Numerical methods and parameters, pp 32–34

1. Introduction

A visual image is a projection of a three-dimensional scene onto a two-dimensional surface. As a result, virtually every image of a real-world scene includes occlusion. When one object occludes another, an edge is formed between them, defined by differences in surface brightness, color or texture. The edge is formed at the boundary of the front surface, and is informative about the shape of that surface. But in the generic case, an edge has no relation to the shape of the occluded surface. In other words, when two surfaces are separated by an edge, only one of those surfaces "owns" the edge – the one which is in front. In order to recover a reliable representation of the shape of surfaces in the 3D world, the visual system must therefore be able to resolve "the border ownership problem" (Nakayama et al. 1995). Determination of border ownership is essential for correct visual segmentation and is intimately related to the process of recovering the true shape of occluded surfaces.

The first clear articulation of the ambiguity inherent in edges is attributed to Edgar Rubin (1921, 1958). Rubin had the insight that the border ownership problem occurs at every edge, in every image, but he cleverly used ambiguous figures to highlight this fact and to study it further. He observed that even when the shape of the border is such that both sides of it form the silhouette of a known object, as in his famous face/vase illusion, only one of the abutting objects can be perceived at any given moment. The perception is bi-stable, alternating from one interpretation to the other over time, but the two objects cannot be perceived simultaneously. Rubin (1921) termed the surface seen in front the "Figure" and the region seen behind the "Ground". By restricting his study to images with two layers of depth, he insured that the latter was indeed the background – the distant-most, shapeless region, which nothing can come behind. In natural scenes, the back region is often a surface with a well-defined shape of its own. But the need to resolve border ownership remains, since the border is still uninformative about the shape of the back surface.

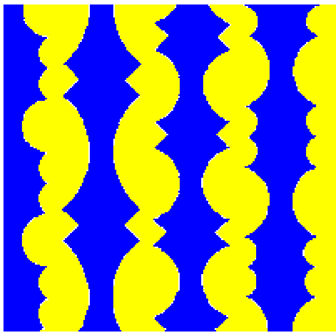

Figure 1: An ambiguous Figure/Ground image demonstrating perceptual bi-stability for unfamiliar shapes. (Adapted from Kanizsa and Gerbino 1976)

Bi-stable Figure/Ground perception can arise also with shapes that do not depict familiar objects. An example is shown in figure 1 (Kanizsa and Gerbino 1976; see also Rubin 1921, figure 2). The perception of this image alternates between two interpretations – yellow regions in front of a blue background, or blue regions in front of a yellow background. Most observers report the former interpretation to be the dominant one (80%; Kanizsa and Gerbino 1976), but with prolonged viewing, both interpretations will be seen. This shows that Figure/Ground alternations are not driven by knowledge of familiar objects.

Rubin's (1921, 1958) observations suggest that the brain has a built-in mechanism which mandates a border to belong only to one surface at a time, and not to both (see also Nakayama et al. 1995; Driver and Baylis 1996; Baylis and Driver 2001; Kourtzi and Kanwisher 2001; Rubin 2001a). This hypothesized mechanism enforces a uni-directional resolution of border ownership, and must operate on every edge, in every image. Ambiguous figures are a useful experimental tool, but what they teach is just as valid for images which give rise to unambiguous Figure/Ground perception.

The neural basis of Figure/Ground (F/G) segregation, i.e., how the brain solves the border ownership problem, remains largely unknown. In particular, it is not yet known at what stage, or "level" of cortical processing border ownership is resolved. (Or, if more than one level of processing is involved, what is the role of each level, and how do they interact to produce the perceptual results observed.) Recently, Zhou et al. (2000; see also Baumann et al. 1997) reported a set of striking findings which bear on the issue. They found that a large fraction of cells in early visual cortex (18% in V1, 59% in V2, 53% in V4) encode information about border ownership. Specifically, those cells exhibited marked differences in firing rates in response to optimally oriented edges, depending on which side of the edge the Figure laid. An illustration of their findings is shown in figure 2. The stimuli shown in panels A and B produce identical stimulation within the cell's "minimum response field" (see Zhou et al. 2000, Methods) and its immediate surround, but the cell consistently responded more to the stimulus in panel A than to that in B. The authors suggested that the difference in responses is due to the inversion of border ownership polarity of the edge: the cell responds when the Figure falls on one side of the edge, but not the other. Consistent with this hypothesis, they found that cells which were insensitive to contrast polarity (3% of the cells recorded in V1, 15% in V2 and 15% in V4) retained their preferred border ownership polarity when the contrast of the images was reversed (panels C and D). Zhou et al. (2000) went on to perform a set of ingenious experiments which provide strong evidence that the modulation of the cell's firing rates was indeed related to border ownership and F/G polarity.

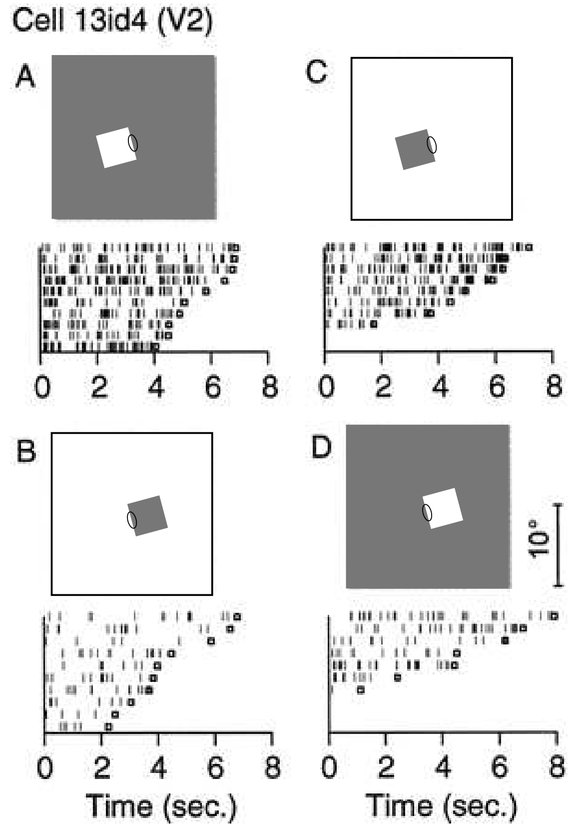

Figure 2: An illustration of the findings of Zhou et al. (2000, adapted from their figure 4). In spite of the identical stimulation within its 'minimum response field' (ellipses), this V2 cell responses strongly only when the Figure is to its left (panel A) and not when it is to the right (panel B). The cell shows preference to this Figural polarity regardless of the contrast polarity of the edge (panels C, D). This suggests that the cell is involved in encoding border ownership.

At first glance, the findings of Zhou et al. (2000) may be taken to mean that F/G segregation is performed in early visual cortical areas. But this interpretation immediately poses new problems. As the authors noted, "... the identification of a region as a Figure requires global image processing (the system needs to evaluate an area of the size of the Figure or more.)" In other words, the resolution of border ownership depends on information from image regions well outside the receptive field of each individual cell. (Zhou et al. found cells that showed sensitivity to F/G manipulations which took place as far as 20° away from their classical receptive field.) The small sizes of receptive fields of cells in V1 and V2 make them seem unsuitable for such global computations. The processing of global image information is generally assumed to be the province of high level visual areas, primarily in inferotemporal cortex, where cells have large receptive fields and complex response characteristics. (We use the terms "low level" and "high level" cortical areas without reference to their function, based solely on their level in the hierarchy of information stream in cortex, as assessed from the distribution of incoming and outgoing connections in specific layers in each area; cf. Felleman and Van Essen 1991). Indeed, authors who previously reported that V1 cells are affected by image manipulations distant from their receptive field (Lamme 1995; Zipser et al. 1996; but see Rossi et al. 2001) proposed that those effects may be the result of feedback from higher level areas (Lamme et al. 1997). Zhou et al. (2000) also seem to favor the interpretation that the border ownership selectivity they found in V1/V2 cells may result from "the presence of top-down signals".

Nevertheless, there are a number of reasons why it may be advantageous for the visual system to achieve Figure/Ground segregation in early visual cortex. A reliable segmentation process needs to be able to identify the location and spatial extent of Figures in the incoming visual image across the entire visual field, and to segment correctly surfaces of any shape, including ones that were never seen before. The retinotopic organization of areas V1 and V2 and the "general purpose" type of processing commonly associated with them (i.e., not designed for specific shapes or objects) fit naturally with these important requirements. Computing F/G segregation in early visual cortex would also free higher-level areas to perform more specialized processing on a small number of segmented surfaces. Another advantage is that areas V1/V2 have immediate access to high spatial resolution information about the image (Lee et al. 1998.) F/G segregation is often sensitive to very subtle manipulations in the image: small changes (restricted to, say, a 1 square-degree area), can change the Figure/Ground assignments of large regions (Gillam 1987; Minguzzi 1987; Rubin 2001b). Therefore, it would be good to perform as much of the computation as possible in the regions which contain cells with receptive fields of the scale of such image manipulations. The purpose of the work presented here was to show that Figure/Ground computations can indeed be achieved by a network whose architecture resembles that of early visual cortex.

2. The Model

We present a model which computes Figure/Ground segregation using "units" which resemble early cortical neurons (V1/V2). The model units have small receptive fields and each unit is connected only to its nearest neighbors. Aside from these two attributes, the model does not attempt to be faithful to physiological details. The computations are done by a set of equations that treat each unit as a simple element characterized solely by its level of activity at each moment. At present, no attempt is made to formulate how those equations would be implemented in a biologically realistic model that takes into account the full complexity of factors such as cells' membrane potentials and synaptic transmission. Nevertheless, we argue that the model has significant implications for research on the neural basis of F/G segregation. It demonstrates that it is possible to observe global effects in a network of local elements. Moreover, the model shows biases similar to those of human observers in its F/G assignments: a tendency to judge enclosed and/or smaller surfaces as the Figure, and a bias to ascribe ownership to the convex side of a border. The fact that these effects can be obtained in a model which uses locally connected units with small receptive fields suggests that the possibility that F/G segregation is achieved in early cortex warrants further consideration.

The general observation that global effects may arise in a system of locally-interacting elements is well known in the physical sciences (e.g., thermodynamic phase transitions, magnetization), and it has been used previously in vision models (e.g., Marr and Poggio 1976; Hildreth 1984; Sha'ashua and Ullman 1988). The intuitive explanation is that as the system evolves in time, information may be mediated across long distances by a cascade of local interactions. How such interactions give rise to the specific global F/G effects observed, however, has not been addressed so far.

The approach we take is different from that prevalent in current vision studies. Most models focus on edges: the bounding contours of surfaces. We term this approach *contour-based*. In contour-based models, units whose receptive fields fall on homogeneous regions in the image normally do not participate in the processing and representation of the scene (Ullman 1976; Sha'ashua and Ullman 1988; Heitger and von der Heydt 1993; Nitzberg et al. 1993; Iverson and Zucker 1995; Williams 1996; Williams and Jacobs 1997; Yen and Finkel 1998). The goals of segmentation are thus defined in terms of contours: identify continuous contours, group together parts of the same contour which are separated in the image because of occlusion etc. Computations take place primarily along contours – e.g., testing for colinearity or relatability (cf. Kellman and Shipley, 1991; Elder and Zucker 1993), using "association fields" to identify contours embedded in a noisy background (Field et al. 1993; Yen and Finkel 1998), or constructing illusory and occluded contours (Heitger et al. 1992; Heitger and von der Heydt 1993; cf. von der Heydt et al. 1984; Peterhans and von der Heydt 1989; Sugita 1999; Bakin et al. 2000). In the contour-based approach, resolution of border ownership may not even be part of the computational task: contours may be detected and represented independently of the surfaces they bound (but see Finkel and Sajda 1992; Weiss 1997).

The approach presented here is *region-based*. Like contour-based models, it recognizes the important role of edges as a source of information about surface boundaries in the scene. But the goal of region-based models is to identify the surfaces bounded by edges – not the edges per se. An edge is treated as a potential part of the bounding contour of a region and the model attempts to find this hypothesized region. As a result, assignment of border ownership is an integral part of the computational task in region-based models. To achieve this goal, the flow of information in region-based models is not confined to be along contours. Instead, all units contribute to the computation – including those whose receptive fields fall within homogeneous parts of the image. Possibly due to the physiological findings of the abundance of neural responses to edges (e.g., Hubel and Wiesel 1968), there has been a longstanding focus on contour-based computations for vision, with relatively small attention to region-based processes (but see Mumford et al. 1987; Paradiso and Nakayama 1991; Kimia et al. 1995; Grossberg 1997). Recently, however, several region-based segmentation models in computer vision have shown promising results in delineating what are called "salient" regions in the image (roughly, the computer vision equivalent of what we term Figural surfaces; cf. Zhu et al. 1995; Shi and Malik 1997; Geiger et al. 1998; Sharon et al. 2000; see also Finkel and Sajda 1992; Grossberg 1997). In this paper, we apply the region-based approach to resolving the border ownership problem and relate it to human perception.

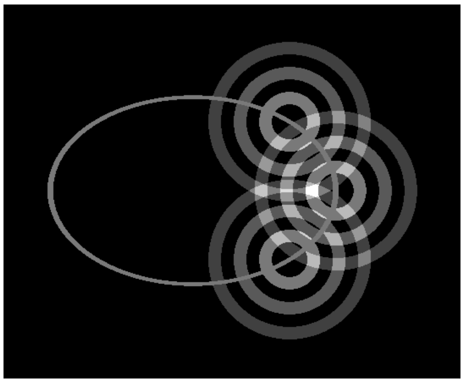

Figure 3: How region-based models give rise to global Figure/Ground effects. Signals about the probability that a point belongs to the Figure are launched from locations near the edge and propagated all around. The cumulative effects of those signals is different on the two sides of the contour. This asymmetry leads the model to prefer the "inside" as the Figure. (The decreasing contrast of the circles denotes attenuation of signals away from their source.)

Allowing signals to propagate also within homogenous regions provides a natural way for global F/G effects to arise. An intuitive illustration of this is shown in figure 3. Consider three local segments along the bounding contour of an ellipse. The goal of a region-based model is to find out if these segments belong to a contour that bounds a region. Because each unit has only local information, this is a non-trivial computation. But by propagating signals between neighbors, the collection of units can actually "find out" that one side is more likely to be the Figure. Suppose that a cascade of signals is launched from each location, with the level of activation of each unit denoting a possibility (or probability) that the unit belongs to the Figure. Because there is no prior knowledge which side of each edge segment may be the Figure, those signals would have to be launched in a completely isotropic way – as indicated by the concentric circles. Nevertheless, inspecting figure 3 reveals that the signals on the inside of the ellipse reinforce each other more strongly: the maximum (summed) activity is higher than for signals on the outside, and points of equal activity lie deeper (further from the edge) on the inside. Note, that for this imbalance to emerge, signals had to propagate also between units that lie on homogeneous regions of the image. Region-based models can therefore exhibit a preference for the enclosed region to be the Figure without needing to implement any special bias in the individual units. (The simple case of the ellipse confounds the properties of closure and convexity; those will be disentangled in section 3.)

Before moving on, there are two qualification we need to make. One, the model presented here is designed for situations where the foreground and background surfaces are homogeneous, i.e., they are not textured and do not contain other internal edges. We are assuming that computations such as identifying regions of uniform texture are handled by other modules (in the human or artificial visual system). The output of these modules can then be used as input to our model. In the tradition established by E. Rubin (1921), who used untextured images in his studies, we restricted ourselves to images of homogeneous surfaces (like that shown in figure 1) and thus could concentrate on the computations most relevant to F/G segregation and border ownership. For real-world applications, this model would therefore need to be combined with models that take care of the work of the other modules mentioned, e.g., texture segregation. The other simplification we took, again following Rubin (1921), is that we restricted ourselves to images with two layers of depth (which also allows us to refer to the back surface as Ground). But extending the model to handle more layers of depth does not require conceptual changes (see Discussion).

2.1. **Formulating the problem.** The model starts by representing the input image by a set of units which signal the luminance and/or color at a small, restricted part of the image. In physiologically-inspired terms, the input is represented using a set of small receptive-field units that tile the image retinotopically. The output is also represented by a set of retinotopically organized small receptive field units. In principle, the density of output units and the extent and location of their receptive fields may be different from those of the input units; however for simplicity we choose them here to be the same. The activity level of each output unit represents the probability that the local region of the image corresponding to the unit's receptive field belongs to a Figural surface. Using the index k for the model units, we denote the activity of output unit k by $P(k)$, where $0 \leq P(k) \leq 1$. The interpretation of an output unit k having a value $P(k) = 1$ or $P(k) = 0$ is straightforward: the former case indicates certainty that the corresponding location is part of a Figural surface, whereas the latter indicates certainty that that location is part of the background. Intermediate values of $P(k)$ indicate varying amounts of uncertainty about whether the location belongs to the Figure or the background. (See section 4 for a discussion of possible neural implementations of intermediate P values.)

Next, we need to define the desired relationship, or transformation between the input and the output in the model. Broadly speaking, we want the output to correspond to what is observed in perception. Thus, to know the desired output for a particular input image, we need to ask observers what they perceive as Figure in that image; the desired output of the model would then be the one with $P(k) = 1$ at units that fall within the (perceptually) Figural region, and $P(k) = 0$ elsewhere. The computations performed by the model should produced the desired output (or a result close to it), i.e. behave like human perception.

To consider a concrete example of an input and its possible outputs, refer to figure 4. It presents an overview of the model, as applied to one simple image. At this point, we focus only on select parts of figure 4. The top row shows the array of 100x100 input units representing the image. We use color to denote different image regions, to prevent confusion with illustrations of subsequent model stages, where grayscale will be used to denote the values of $P(k)$. (Recall that in principle, these two regions could be differentiated by other attributes, e.g. luminance or texture.) Although the input units can signal different values of color or luminance, this does not solve the border ownership problem.

The top panel of figure 4, is perceived as a yellow ellipse (the F) in front of a blue background. The desired output would therefore be $P(k) = 1$ for all units k inside the ellipse, and $P(k) = 0$ for those outside. Denoting $P(k) = 1$ with white, $P(k) = 0$ with black, and using a grayscale for intermediate values, the bottom panel of figure 4 shows the output produced by the model. Clearly, it is a good approximation of the desired output.

En route to finding this output, many other possible outputs are evaluated. Two of them are shown in panels I and J of figure 4. Panel I corresponds to a percept of an elliptical "hole" in a blue foreground, looking onto a yellow background. Panel J shows a possible (but undesired) output with little relation to the input image: transitions between F and G occur in places where there are no edges in the image (note: we use the term "edge" to mean luminance-, color- or texture-defined edges in the input image.) Furthermore, many of the output units have intermediate values of $P(k)$ (gray), indicating an output with many "undecided" units. The model evaluates the possible outputs by computing a certain value, or "cost" for each of them. The success of the model is that the output with the lowest cost approximates

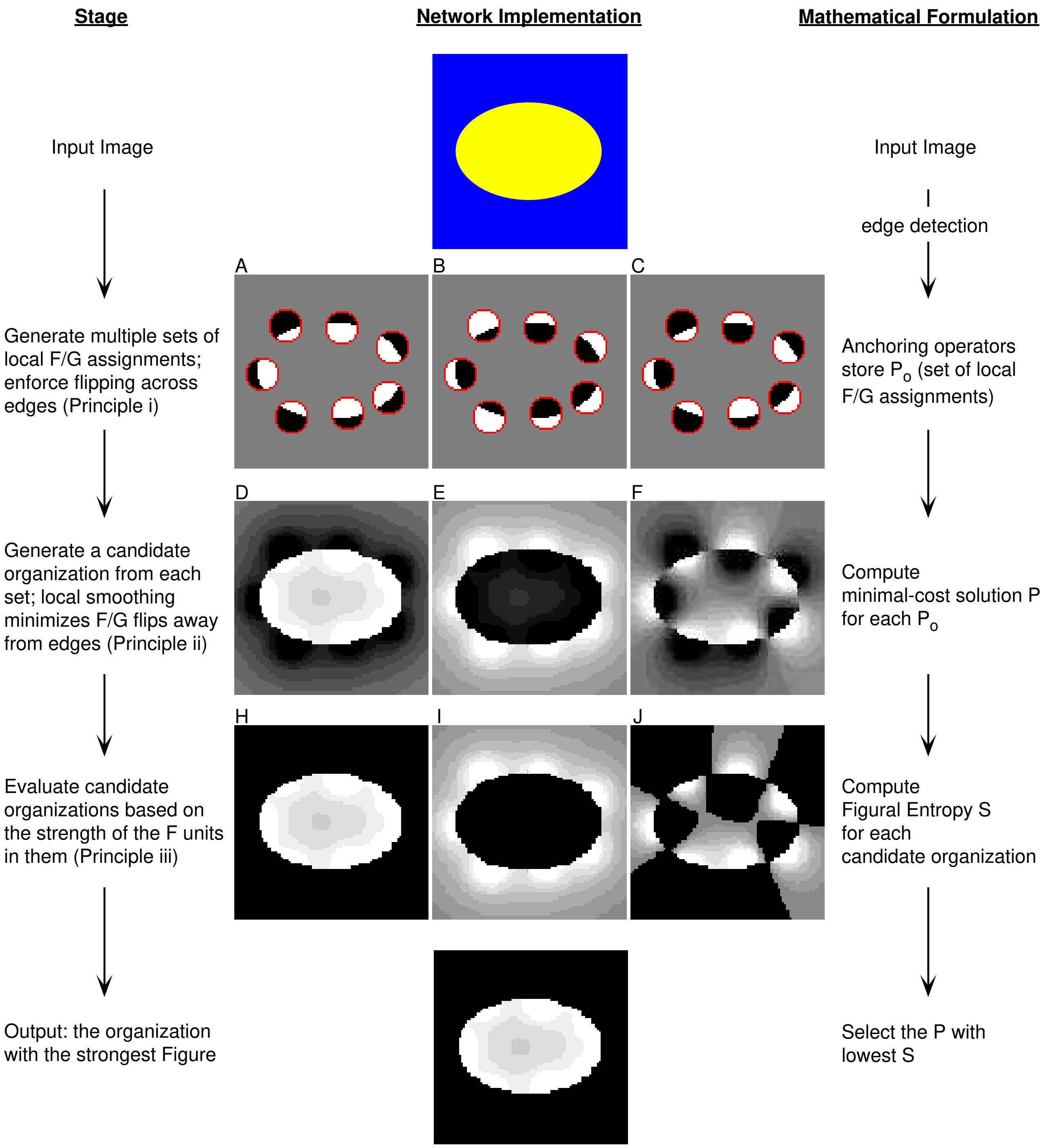

Figure 4: An overview of the model. The left and right columns describe the stages guided by Principles i-iii and their mathematical formulation, respectively. The central column shows the network implementation. *Top panel*, the input image. *Bottom panel*, the output. Grayscale indicates $P(k)$, the probability that unit k is Figure, with $P(k) = 1$ in white, $P(k) = 0$ in black, and gray for intermediate values.

well the F/G organization observed perceptually, while undesired organizations (such as 4I and J) lead to higher costs.

In a sense, the idea that the observed perceptual organization is the one which minimizes (or maximizes) a cost function dates back to the Gestalt Psychologists, who theorized that the brain selects visual interpretations which maximize "perceptual goodness" (or Prägnanz; cf. Koffka 1935.) Here, however, we are faced with the challenge of devising a formula that measures the "goodness" of Figure/Ground organizations quantitatively.

In addition to handling input images that give rise to unambiguous F/G interpretations (like that in figure 4), we would like the model to be able to predict when a certain image may lead to perceptual bi-stability. Thus, for an input image like that in figure 1, we expect to find two outputs with low cost – one corresponding to the yellow regions being F, the other to the blue regions being F. As will be seen in section 3, the model will even predict that the cost of the former will be slightly lower than that of the latter, corresponding to its observed perceptual dominance. Nevertheless, both costs will be much lower than that of other, 'nonsense' organizations, which are not observed perceptually.

2.2. **Guiding principles.** In this section, we outline the main stages of computations in the model. We reserve details of the mathematical and network implementation for the next section; here we focus on the ideas the model is based on, and how they promote outputs that are in agreement with human perception. Figure 4 will be used to give an overview of the stages of the model. For readers who are not interested in the mathematical implementation, reading this section should be enough to understand how the model works (i.e., it is possible to skip section 2.3 and go directly to the simulations).

The model is based on the following three principles:

(i) F/G boundaries are likely to be present along luminance/color gradients (edges).
(ii) F/G boundaries are unlikely where edges are absent.
(iii) Among all possible organizations satisfying (1) and (2), prefer the one(s) where the Figural units ($P(k) > 0.5$) have the strongest F assignment – i.e., where they are closest to 1.

These three principles may seem innocuous, but we will see that in concert they provide enough constraints to allow the model to converge at the right solutions. Furthermore, these principles can be implemented in a physiologically plausible way, on a network of small receptive field units with local connections between them.

At the first stage, the model generates sets of F/G assignments over the entire image. At this point, each set is only required to satisfy Principle (i): that the F/G assignments should flip across the edges. The image is sampled with a collection of local operators whose "receptive field" extends several input units, to make it possible to detect edges within them, but still much smaller than the scale of the Figure (typically 8-15 units in diameter; see section 2.3.1 and Appendix B). If an edge (defined as gradient > criterion) falls within the receptive field of an operator, the units within its receptive fields are labeled *border units*. The operator then assigns all its border units on one side of the edge as F, and those on the other side as G. Which side gets F is decided at random at this point. The rest of the units in the model (i.e., the non-border units) are labeled as undecided (0.5). We term the local operators *anchoring operators* to reflect the fact that they will exert influence on the model units to remain at their assigned values. The extent of this influence will be large for border units (i.e., those assigned 1 or 0), and small for non-border units.

Since the model has no "global" knowledge (e.g., units with receptive fields of the scale of the global Figure), F/G assignments are generated independently within each anchoring operator. As a result, there will be many possible sets of F/G assignments. Each set corresponds to a different combination of the two possible F/G polarities within each anchoring operator. Panels A-C in Figure 4 show three examples of such sets. (The anchoring operators that fall on edges are illustrated by the red circles.) The two sets of most interest to us are those in panels A and B. Panel A corresponds to the interpretation observed perceptually, of an elliptical Figure, and panel B to the reversed F/G interpretation – an elliptically-shaped hole in a 'front' surface. Explaining why A is preferred to B is a major goal of the model. But there is also something in common to these two sets: they are globally consistent. Because of the local nature of the F/G assignments, the vast majority of the sets of initial F/G assignments will not share this property, i.e., they will be globally *in*consistent. Panel C shows one such example. Following the circumference of the ellipse, the polarity of F/G values changes erratically. This is unavoidable given the independent assignments at each operator – there is no way to a priori guarantee global consistency of the anchoring F/G assignments. Among all possible anchoring F/G assignments, there are many more cases like panel C: given a anchoring operators, there are $2^a - 2$ globally inconsistent anchoring F/G assignments and only 2 consistent ones. Another major task of the model is therefore to identify and exclude the globally inconsistent sets. What will make the model select the globally consistent anchoring F/G assignments, albeit rare, is that they lead to organizations that satisfy also Principles (ii) and (iii).

At the next stage, the model takes each set of local F/G assignments and generates from it a *candidate organization*: the best F/G organization that can be obtained for the given set. What makes it "best" is that it is the organization which follows Principle (ii) most closely: minimize F/G boundaries in image regions that do not contain edges. Formally, this is achieved by means of a *cost function* $\mathcal{E}$ which penalizes for F/G transition away from edges. The candidate organization is the one for which $\mathcal{E}$ is minimized for the given set of F/G assignments. (A detailed description of the mathematical formulation, as well as an explanation of how the minimization process is implemented in the network, is given in section 2.3.2.) Consider panels D-F of figure 4, where the candidate organizations obtained for the F/G assignments in panels A-C, respectively are shown. When the set of F/G assignments is globally consistent, as in panels A and B, the minimization of $\mathcal{E}$ leads to a candidate organizations with the majority of units near 1 or 0 (panels D and E, respectively). In contrast, the candidate organization in panel F, which was generated from the globally inconsistent set C, contains many units near 0.5 (gray). The reason for this is that in order to minimize $\mathcal{E}$ (and thereby adhere to Principle (ii)), the model smoothly interpolated between the conflicting values near the edges, so that sharp F/G transitions would not occur away from the edges. The abundance of units near 0.5 in globally-inconsistent candidate organizations will be used in the next, final stage of the model to prune them, leaving us only with the globally consistent organization(s).

The final stage of the model implements Principle (iii): it evaluates each of the candidate organizations based on the strength of its Figure and selects the strongest one(s). The 'Figure' in a candidate organization is defined as the collection of units k that have $P(k) > 0.5$. Referring back to figure 4, panels H, I and J show the Figures of candidate organizations D, E and F, respectively (the units with $P(k) < 0.5$ were all set to black). To evaluate the strength of each Figure, the model computes a function called *entropy*, denoted $\mathcal{S}$, which is monotonically decreasing the more units in the Figure are near 1. (For details see section 2.3.3.) It then chooses as its output the candidate organization with lowest $\mathcal{S}$. For the case of the ellipse in figure 4, the entropies of the Figures in panels H, I and J are 0.34, 0.74 and 0.77 respectively. The output is therefore that shown in panel H (reproduced in the bottom row), which is consistent with perception. Note, that the model pruned the undesired organizations although no explicit computation to detect global inconsistencies was built in. An intuitive explanation for why the candidate organization in panel F leads to high Figural entropy was already given: globally inconsistent sets of F/G assignments (e.g., panel C) lead to candidate organizations with many units with $0.5 < P(k) \ll 1$, which in turn lead to high entropy. With regard to the high Figural entropy of the globally consistent candidate organization in panel E, some intuition was given in figure 3, and this issue will be revisited in section 3.

The large difference in values between the entropy of the Figure in panel H and that of panel I predicts that the interpretation of the image as an elliptical Figure will be much stronger than that of an elliptically-shaped hole. Indeed, the entropy of the "hole" interpretation is not much lower than that of the "nonsense" interpretation J. This is in very good agreement with what is observed perceptually. For other images, however, there may be more than one candidate organization with entropy considerably lower than the rest. In such cases, the model's prediction is that perception will alternate between these organizations. Some examples of such bi-stable images will be discussed in section 3. (Note: the "hole" interpretation can be promoted perceptually by introducing stereoscopic cues which disambiguate the depth relationships of image regions; here we do not consider such additional factors in the resolution of border ownership.)

To summarize, once multiple sets of F/G assignments are generated in accordance with Principle (i), the model proceeds in two stages. In the first stage it generates a candidate organization from each set, by minimizing a cost function $\mathcal{E}$. The structure of $\mathcal{E}$ implements Principle (ii), be penalizing F/G transitions away from edges. As a consequence, global inconsistencies in the F/G assignments result in candidate organizations with many 'undecided' units, while globally consistent sets result in candidate organization with few undecided units. In the second stage, the model chooses the candidate organization with the lowest entropy $\mathcal{S}$, which leads to the elimination of candidate organizations with many undecided units, in accordance with Principle (iii).

The present model allows for an efficient minimization of the cost function $\mathcal{E}$ as a dynamical process in the network, but it does not offer a similarly efficient implementation for finding the best candidate organization(s) with lowest $\mathcal{S}$. Here, this second stage requires a search through the set of candidate organizations, a process which is not biologically plausible. An efficient convergence to the low entropy organization(s) can be achieved if one introduces local interactions which favor consistent polarity of F/G assignments between neighboring anchoring operators. In effect, this turns the present "two stage" version of the model (first minimize the cost $\mathcal{E}$, then compute the entropy $\mathcal{S}$) into a single stage of minimizing a cost function which incorporates both $\mathcal{E}$ and $\mathcal{S}$ (Pugh and Rubin, in preparation). But since the mathematical formulation of the combined $\mathcal{E} - \mathcal{S}$ model is much more complicated, we limit the discussion here to the two-stage model, which allows us to present the main ideas and achievements of such models in a simpler form.

2.3. **Network and mathematical implementation.**

2.3.1. *Luminance edges and Figure/Ground transitions.* The first stage in the model was described in detail in section 2.2. Below we give a brief summary and introduce further notation and details. After a pre-processing stage of edge detection, the model samples the image with a dense set of anchoring operators. F/G assignments are then generated locally within each anchoring operator. We denote the anchored values P_0. The values of P_0 reflect Principle (i): that F/G transitions are likely near edges. If an edge falls within an anchoring operator, it assigns values $P_0(k) = 1$ (F) or $P_0(k) = 0$ (G) to its border units, such that all units on one side of the edge are assigned one value, and those on the other side are assigned the opposite value. Otherwise, the units within the anchoring operator are non-border units and they are assigned $P_0(k) = 0.5$. The independent generation of P_0 values within each anchoring operator leads to multiple sets of F/G assignments. The next stage will be to generate a candidate organization from each set.

The spatial distribution of anchoring operators in the model is dense, i.e., an operator could be centered at any input unit location. But not all anchoring operators are activated when processing a given image. First, to avoid conflicting anchoring assignments of a single border unit, operators with overlapping receptive fields are not allowed to be simultaneously active. In addition, the overall density of activated operators is a parameter in the model. In the simulations presented here, we choose which anchoring operators will be activated at random (subject to the constraints listed above). We expect that for natural scenes, anchoring operators may be activated around the stronger edge segments in the image.

2.3.2. *Generating the candidate organizations: the cost function $\mathcal{E}$.* The next step is to compute the candidate organization that each set of F/G assignments gives rise to. This step aims to satisfy Principle (ii): F/G boundaries are unlikely where no luminance borders are present. Thus, we seek a candidate organization which agrees with the anchoring F/G values assigned to the border units, while minimizing the amount of F/G reversals at non-border units.

We denote the collection of values $\{P_0(k)\}$, which is defined on the entire intermediate layer (both border and non-border units), by $\mathbf{P}_0$. We will denote the resulting candidate organization by $\mathbf{P}$: the collection of values $\{P(k)\}$ that specify the probability that the corresponding location in the image belongs to the Figure (for the given $\mathbf{P}_0$). Finding a network configuration that best satisfies a set of constraints – in our case, adherence to the F/G assignments (Principle i) while minimizing F/G boundaries away from edges (Principle ii) – is done by expressing them in a cost function that the network minimizes. (How network dynamics can perform this minimization will be shown later; cf figure 5 and Appendix A.) We use the following cost function $\mathcal{E}$:

$$\begin{aligned} \mathcal{E}(\mathbf{Q}) \;=\; & \sum_{k=1}^{n^2} \sum_{j \in N_k} \mu_{kj} \left[Q(k) - Q(j)\right]^2 \\ & + \sum_{k \in A} \left[Q(k) - P_0(k)\right]^2 \\ & + \sum_{k \in A^c} \nu \left[Q(k) - P_0(k)\right]^2 \end{aligned} \tag{1}$$

$\mathcal{E}$ was written here as a function of $\mathbf{Q}$; like $\mathbf{P}$, it denotes a collection of F/G values of all the units. However, unlike $\mathbf{P}$, $\mathbf{Q}$ does not necessarily minimize $\mathcal{E}$. For example, it could be a random collection of values – in this case, the value of $\mathcal{E}$ would simply be very large and far from the minimum. Furthermore, the quadratic dependence of $\mathcal{E}$ on $\mathbf{Q}$ means that there is only one $\mathbf{P}$ that minimizes it, and

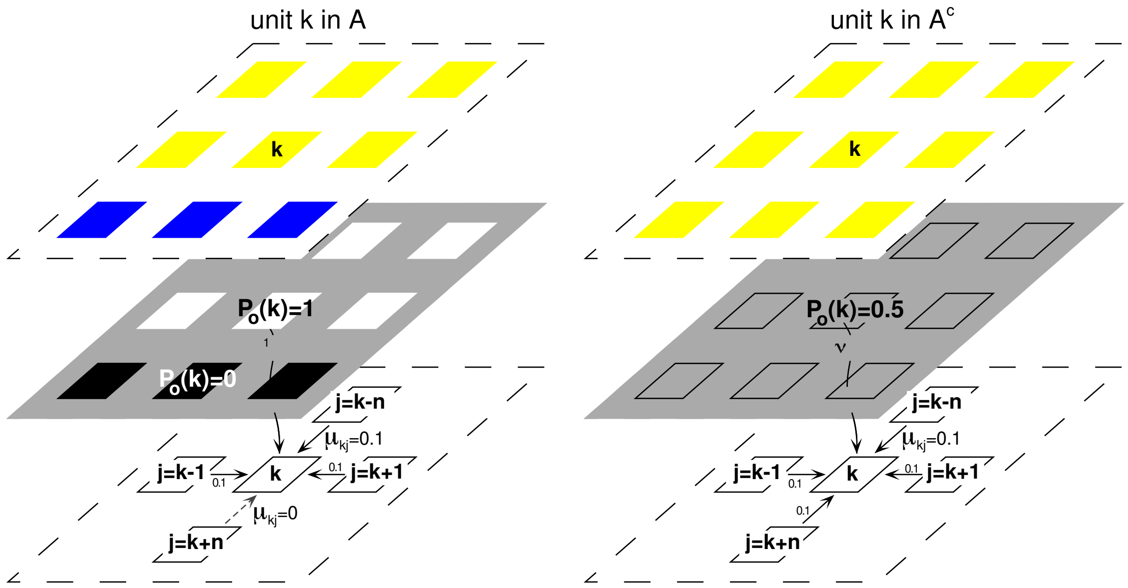

Figure 5: Network implementation of equation (1). *Left*, The connections to a border unit (i.e., one that is near an edge and is covered by an anchoring operator). *Right*, The connections to a non-border unit. In both panels, the top layer shows the image, the intermediate layer shows the values of P_0 and the bottom layer shows the network which computes P. Appendix A shows that this network will settle in a state which minimizes equation (1).

this is the candidate organization that arises from $\mathbf{P}_0$. The other notations are best understood by consulting figure 5, which shows two illustrations of a small piece of the network. Because the network is based on local connections, we focus on a single unit and its nearest neighbors. For a unit with index k, its nearest neighbors to the 'west' and 'east' have indices $k-1$ and $k+1$, respectively. Since the network is $m \times n$, the nearest neighbors to the 'north' and 'south' are $k-n$ and $k+n$.

The first term of equation (1) deals with the connections between unit k and its nearest neighbors, which are denoted as a set by N_k. Using j as a generic index for a neighboring unit, its input to unit k is weighted by μ_{kj}. The values of these weights are shown in the bottom layers in figure 5. They depend on whether unit k is near an edge (left panel) or not (right panel). For units near an edge, the incoming weights from neighbors on the other side on the edge (in the figure, it is unit $k+n$) are set to zero. The weights from units on the same side of the edge have value μ, which is a free parameter (μ should be less than 1; in the simulations presented later, we took $\mu = 0.1$). For units not near an edge (right panel), the incoming weights equal μ for all neighbors. (Note, that this means that μ_{kj}, the incoming connection to unit k, is equal to μ_{jk}, the outgoing connection from k to j. This symmetry will be used later.)

The second term in equation (1) deals with how border units are affected by their anchored F/G assignments (here, A stands for "anchored" and denotes the set of border units.) Consider again the unit k on the left panel of figure 5. This unit is not only near an edge, but is also in A (recall that not all units near edges will fall within an anchoring operator; for illustrative purposes, the figure depicts a case where unit k as well as its neighbors are all in A.) The F/G assignments are shown on the intermediate layer. Unit k and three of its neighbors were assigned $P_0(k)$ of 1 (white), while the fourth neighbor was assigned 0 (black). But unit k will only be affected by its own anchored value, $P_0(k)$. This is denoted by the arrow from the intermediate layer unit k to the bottom layer. This connection weight is 1 (i.e., it is not weighted by a free parameter.)

Finally, the third term in equation (1) deals with how non-border units are affected by their unbiased F/G assignment of 0.5 (A^c denotes the complement set of A). This is illustrated for unit k on the right panel of figure 5. The connection from the intermediate layer to unit k in the bottom layer is weighted by a very small number, which we denote as ν. It is another free parameter (e.g., $\nu = 0.0002$ for the results presented in section 3).

To understand how the structure of the cost function $\mathcal{E}$ affects the solution $\mathbf{P}$, consider the three choices of $\mathbf{P}_0$ in panels A-C of figure 4 and their resulting minimizers $\mathbf{P}$ shown in panels D-F. Despite their obvious differences, the three minimizers share some common characteristics. One, they are smooth except where there are edges in the image; two, they show fidelity to the anchoring values near the edges; and three, they contain some units with intermediate values of $P(k)$ (i.e., neither 1 nor 0). These commonalities directly relate to the three terms in $\mathcal{E}$. Below, we discuss each term in turn, and then discuss their combined effect.

Smoothness. A notable thing about all three candidate organizations in panels D-F of figure 4 is the absence of sharp transitions away from the edges. Along the edges, there are jumps between Figure and Ground (white and black, respectively), as dictated by Principle (i). In contrast, away from the edges, the gray levels vary smoothly. This is an implementation of Principle (ii), and mathematically, it is a result of the first term in $\mathcal{E}$. This term increases with every pair of neighboring units which have different values of $P(k)$; the larger the difference, the larger the additional cost. The exception to that is when the neighbors are separated by a luminance border, in which case differences between the values of $P(k)$ and $P(j)$ are not penalized, since μ_{kj} was set to 0 for such pairs (to allow for F/G changes across luminance borders). This is why sharp F/G transitions occur also in portions of the edges where anchoring units were not activated (cf. figure 4D-E.)

The smoothness away from the edges reflects the fact that, for a fixed set of anchored values along the edges, the function $\mathbf{P}$ which minimizes the first term in $\mathcal{E}$ is given by interpolating between the edge values. We therefore call the first term *the smoothness term.* It has the effect of smoothly propagating, or spreading the F/G values from border units to non-border units, i.e., from edges onto entire regions. Note, that although the smoothness term explicitly couples only neighboring units (and therefore preserves the locality of the model), a given unit can have an effect on the value at much more distant units. This is because chains of pair-wise terms $[P(k) - P(k+1)], [P(k+1) - P(k+2)], \ldots, [P(k+n-1) - P(k+n)]$ effectively couple units separated by n other units (as long as they are not separated by an edge). Intuitively, we can think of this as contributions to the cost by a cascade of pairs of neighboring units with different values of P. Thus, although the model admits only local *connections* ($\mu_{kj} \neq 0$ only if units j and k are nearest-neighbors), the smoothness term enables long-range *interactions* (global effects) in the model.

Fidelity to the anchoring values. There is a clear relation between each candidate organization and the set of anchoring values that led to it. Specifically, the minimizers $\mathbf{P}$ are close to $\mathbf{P}_0$ at the border units. This is a result of the second term of $\mathcal{E}$. This term gets a positive contribution whenever $P(k)$ at a border unit k deviates from the anchored value $P_0(k)$ there. Thus, $\mathcal{E}$ penalizes for deviations from anchored values, although it allows such deviations, in principle. The minimizer $\mathbf{P}$ will therefore tend to remain faithful to the anchored values at the border units. This allows the model to evaluate different possible $\mathbf{P}_0$'s through their effect on $\mathbf{P}$.

The effect of $\mathbf{P}_0$ *at non-border units.* The third term in $\mathcal{E}$ penalizes deviation of $P(k)$ from $P_0(k)$ at non-border units. However, the cost for such deviations is much smaller than at border units: they are weighted by the small value of ν. The effect of this term is best understood by inspecting panels D or E in figure 4 – for simplicity we pick one of them, D. It shows the solution which minimizes $\mathcal{E}$ for the $\mathbf{P}_0$ shown in panel A. There, the anchoring values assigned at the edges were globally consistent. If $\mathcal{E}$ consisted only of the first two terms (i.e., if ν were set to zero), the minimizing solution would be very simple: $P(k) = 1$ for all units k inside the ellipse, and $P(k) = 0$ for all units outside of it. This solution leads to nulling of both the first and the second terms in $\mathcal{E}$ and since both terms are quadratic (i.e., always positive), a function which nulls them is necessarily the one which minimizes them.

Thus, with only the first two terms in $\mathcal{E}$, the spread of F/G values away from the border units would be potentially unlimited: all units, no matter how far away from the edge, may end up having "hard" F/G values $\mathbf{P} = 1$ or $\mathbf{P} = 0$. This might seem like a good thing, at first glance, but in fact there are several reasons why this is undesirable. Considering physiological plausibility, it is more reasonable to assume that there would be some loss of signal as it is transmitted via a chain of short-range connections. From a computational point of view, attenuating the F/G signal as one moves away from the edge can be advantageous: for example, in resolving conflicting F/G signals which arrive at a single location from different directions, it would allow the unit to go with the stronger signal which would then imply the closer edge. Finally, as we shall see in the next section, unlimited spread of the F/G signal leads to predictions which are inconsistent with perception, specifically about the effects of closure, convexity, and size on F/G perception.

The interplay between the terms. While isolating the terms affords some understanding of the cost function, in the full problem they play off of one another. If the cost function consisted only of the second and third terms, then it would be minimized by $\mathbf{P}_0$ - i.e., the output would be identical to the input (as these two terms control the 'faithfulness' of the solution of equation (1) to the anchored values at border units and to the $\mathbf{P} = 0.5$ values elsewhere). If, on the other hand, the cost function consisted only of the first term, then the penalty it puts on deviations from smoothness would drive the system to a trivial solution of constant P values within regions that do not contain a luminance border, with no regard to the values set by the anchoring operators. The requirement that the solution minimizes the sum of all three terms creates an interplay between these conflicting demands. As a result, the minimizer $\mathbf{P}$ will be as smooth as possible (first term) while remaining faithful to the initial biases (second and third terms). The free parameters μ and ν affect the relative weight of these demands. The second term has no parameter and is of order 1. The first term scales like μ: by taking μ small we are valuing fidelity to the anchored values at the border units over smoothness. By taking ν very small, we value a modicum of fidelity to the unbiased values ($P_0(k) = 0.5$) but this is the weakest demand of the three.

2.3.3. *Identifying undesired candidate organizations: the Figural entropy.* The final step of the model is to evaluate the "global goodness" of each of these organizations and thus select the best output F/G organization(s). This stage implements Principle (iii): prefer the organization(s) where the Figure is the strongest. The 'Figure' in a candidate organization is defined as the collection of Figural units (units k with $P(k) > 0.5$). Its strength is quantified by the Figural entropy:

(2)

$$\mathcal{S}(\mathbf{P}) = -\frac{1}{N_{\text{fig}}} \sum_{k, P(k) > .5} 2P(k) \log_2(P(k))$$

where N_{fig} is the number of Figural units. The name "entropy" comes from the similarity of equation (2) to the formula for entropy in statistical mechanics (Reichl 1998) and information theory (Cover, 1991). However, we make no claim for a physical or information-theoretic basis for $\mathcal{S}$. The selection of the function $-2P \log_2(P)$ is in fact somewhat arbitrary; any monotonically decreasing function defined over $[0.5, 1]$ would do for our purposes. Our choice has the advantage that it offers a convenient scale to intuitively grasp the "goodness" of the Figure. The value of $\mathcal{S}$ is confined between zero and one: if all Figural units have $P(k) = 1$, $\mathcal{S}$ will be 0; as more and more Figural units depart from 1 and approach 0.5, $\mathcal{S}$ approaches its maximal value of 1.

Using the Figural entropy as a measure, the model ranks the candidate organizations $\mathbf{P}$ of a given image, creating a hierarchy of F/G organizations. In some cases, there will be one candidate organization whose entropy value is significantly lower than that of all other organizations. As seen earlier (section 2.2), this is the case for the ellipse in figure 4. The Figural entropy of the (potential) output in panel H is 0.34, while those of panels I and J are 0.74 and 0.77, respectively. The output of the model is therefore that in panel H (redrawn on the bottom panel), in agreement with perception. In other cases, there may be more than one candidate organization with low entropy. The model predicts that these situations will result in perceptual bi-stability, as will be discussed in the next section.

Before proceeding to the simulations, another note about $\mathcal{S}$ needs to be made. The strength, or "goodness" of a candidate organization is determined only by the values of the Figural units ($P(k) > 0.5$). This is motivated by the difference in perceptual quality of Figure and Ground (Rubin 1921, 1958). However, mathematically it is possible to draw information from the values of the Ground units as well. Consider the two globally consistent sets of F/G assignments, $\mathbf{P}_0$, in panels 4A and B. They are symmetric about 0.5 with respect to each other: one set can be obtained from the other by the transformation $(x \rightarrow (1 - x))$, i.e., from $1 - P_0$. The structure of the cost function $\mathcal{E}$ preserves the symmetry with respect to this transformation. Therefore, the solution to the $\mathbf{P}_0$ in panel B can be obtained by reflecting the solution to the $\mathbf{P}_0$ in panel A about 0.5. In other words, the values of $P(k)$ shown in panel E equal $(1 - P(k))$ of those shown in panel D. This relation is not limited to the pair of globally consistent solutions: each set $\mathbf{P}_0$ has a complementary set obtained by reflection. But this fact is particularly useful for the globally consistent ones, because it allows one to evaluate the Figural entropies of both of them from a single solution. Specifically, applying equation (2) to the $(1 - P)$ values of the *background* units of one solution gives the same value as one would get from computing the Figural entropy of the complementary solution. We shall use this extensively when discussing further results in the next section.

3. Simulation results

This section presents results from a set of images designed to study how the model behaves under conditions which are known to affect Figure/Ground perception. The bulk of this section can be understood even if section 2.3 was skipped, although we make occasional reference to it.

The first image was already introduced in the previous section: it is the ellipse in figure 4. We have already seen that the model output corresponded to that observed perceptually: the organization with the lowest entropy in figure 4 was that of an ellipse in front (panel D). figure 3 gave some intuition as to why this organization was preferred over that of panel E (an elliptic hole): the Figure units are, on average, stronger (closer to 1) when signals propagate "inwards" than "outwards". Next, we wish to disentangle the effect of the two properties that contributed to the preference of panel D: the ellipse being the enclosed region and the convexity of its bounding contour.

To study the effect of convexity in isolation, we ran the model on images that contained regions with either convex or concave sides, where no region was enclosed in another. To save space, figure 6A presents only the result of the simulation for one of the two globally consistent solutions. (The original image can be inferred from it easily.) It is known that observers tend to perceive convex regions as the Figure (Kanizsa and Gerbino 1976; see also Liu et al. 1999).

This tendency is also shown by the model: the entropy of the solution in figure 6A, i.e., when the convex regions are Figure, was 0.48 ± 0.03 (see Appendix B for how confidence intervals were computed). In contrast, the entropy for the other globally-consistent solution, i.e. when the concave regions are presumed Figure, was 0.71 ± 0.03. (Globally inconsistent solutions produced even higher entropies and will not be discussed further.)

This difference between the two globally consistent solutions is a direct result of convexity. As one moves away from an edge into the Figural, convex region in figure 6A, the values of $P(k)$ fall off from 1 (turn from white to gray), and then start to rise again as one approaches the edge on the other side of the region. Similarly, the $P(k)$ values for units on the concave, Ground side increase from 0 (turn from black to gray), and then decrease back. However, the fall off from 1 on the convex side is slower than the rise

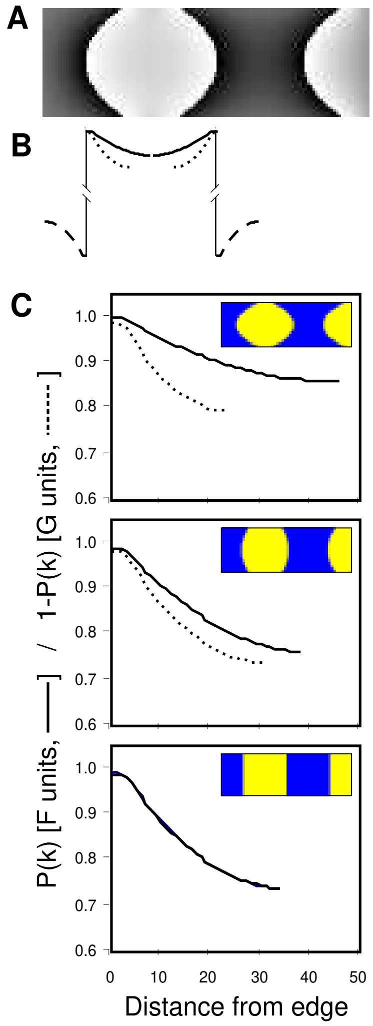

Figure 6: The effect of convexity on Figure/Ground organization. *A*, The candidate organization $\mathbf{P}$ for a globally consistent set of F/G assignments where the convex regions are be Figure. *B*, A cross section showing $P(k)$ as a function of the distance of Figural unit k from the edge (solid curve). The dashed curve shows $P(k)$ for G (background) units, and the dotted curve shows a reflected version of $1-P(k)$ for the same units, allowing the faster decay of the G units to be compared directly with the slower decay of the F units. *C*, Plots of the decay of $P(k)$ and $1-P(k)$ for F and G units, respectively, as a function of distance from the edge. Three levels of convexity are shown, decreasing from top.

from 0 on the concave side. To illustrate this difference more clearly, panel B shows a graph of $P(k)$ for units which lie along a central horizontal cross section. (For simplicity, only one cycle of the solution is shown.) The solid curve denotes the value of $P(k)$ for F units, and the dashed curve for G units. To enable direct comparison, the values of $(1 - P(k))$ for the G units have been "reflected" about the edge and redrawn on the Figural (convex) side with a dotted line. In section 2.3.3 it was shown that the entropy of the other globally-consistent solution can be derived from the $(1 - P(k))$ values of the G units in this solution. The faster decay of the G units therefore explains the higher entropy for the other solution, where the concave regions are considered Figure.

The effect of convexity depends on the curvature of the edge: the more curved the edge, the more mutual reinforcement there is between the F units on the convex side of the edge (see figure 3). As a result, the difference in the rates of decay on the convex and concave sides increases with curvature. This is shown in figure 6C. The values of $P(k)$ for the F units and of $(1 - P(k))$ for the G units are plotted for three levels of curvature (the lowest level is zero, i.e., straight edges, for which the two curves coincide.) Consequently, the values of the Figural entropies for the two solutions (convex regions are F versus concave regions are F) become closer and closer as the curvature decreases, and coincide when the edges are straight. This predicts that the lower curvature images will give rise to more perceptual bi-stability.

Note that the effect of convexity is in conflict with another factor: the distance of a unit from the edge. A unit at the center of a convex region of our image was always further away from the edges than a unit at the center of a concave region (cf. figure 6A). When all other factors are equal, then the greater the distance from the edges, the more decay a unit suffers. Nevertheless, the effect of convexity more than compensates for this attenuation.

Another form of conflict between two factors – convexity and closure – is when the circumference of an enclosed region contains portions which are concave (e.g., a kidney bean shape.) Near those portions, F units on the "in" side of the curve will decay faster than G units on the "out". But since the region is enclosed, there are always more portions of its circumference which are convex (with respect to its inside), with the net result being that the candidate organization where the enclosed region is F prevails (data not shown).

Next, we discuss the model's behavior with perceptually ambiguous images. At first sight, one might expect that the model will signal ambiguity by settling into a state where many units have $P(k) = 0.5$. However, this would not be consistent with perception: ambiguous images lead to multi-stability, where there are several (typically two) distinct configurations, each with well-defined Figure and Ground regions. Upon prolonged viewing, which configuration is experienced alternates – but at any given moment, there is a definite sense which regions are Figure. Therefore, for ambiguous images, the model should generate several candidate organizations with low entropy (i.e., *few* units near 0.5). The more comparable these entropies are, the more balanced the competing percepts would be. Conversely, significant differences between the (low) entropies means that one of the percepts would be more dominant.

To test these predictions, we ran the model on the perceptually bistable image shown in figure 1 (Kanizsa and Gerbino 1976; Kanizsa 1979). The Figural entropy of the globally consistent organization with the yellow regions as Figure is 0.42 ± 0.02, and for the blue regions as Figure it is 0.45 ± 0.02. The Figural entropies for globally inconsistent configurations, in contrast, were much higher, averaging around 0.70. These results are in

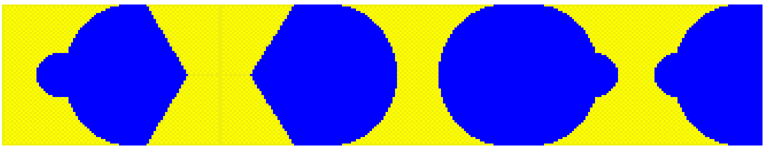

Figure 7: Like figure 1, this ambiguous image also leads to perceptual bi-stability, but the greater degree of convexity here leads to a stronger dominance of the blue regions as Figure (Adapted from Kanizsa 1979.)

good agreement with the perception of this image. The close values of the two globally consistent organizations predict that both would be perceived. At the same time, the slightly lower entropy of the "yellow is Figure" interpretation predicts a dominance of this percept.

Kanizsa and Gerbino (1976) attributed the advantage of the yellow regions in figure 1 as Figure to convexity. Note, however, that these regions are not convex in the strict, mathematical sense of the word. It is possible, in principle, to quantify intermediate degrees of convexity (although we will not do this here). A shape which is convex in the usual sense (e.g., the ellipse or the light regions in figure 6) would get a score of (say) 1, while the yellow regions in figure 1 would get a lower score, but yet higher than the blue regions. It is interesting that both the model and human perception show sensitivity to such intermediate levels of convexity. Kanizsa (1979) was evidently aware of this nuance. In discussing figure 7, he noted that the convexity there was "more accentuated" (than in figure 1), and that the dominance of the convex regions was, in turn, stronger (pg. 109). The results the model gives for this image reflect this, too: the entropy for the blue regions as Figure is 0.48 ± 0.03, while that of the yellow regions as Figure is 0.57 ± 0.03. This difference is larger than that obtained for figure 1, but smaller that that obtained for the ellipse (see also Pao, Geiger and Rubin 1999).

Kanizsa and Gerbino (1976) used figures 1 and 7 to demonstrate an additional point, which is that convexity wins over symmetry: in both images, the less-convex regions are symmetric, but nevertheless perception is dominated by the non-symmetric, but more-convex regions. Our model suggests that the visual system's greater sensitivity to one global property (convexity) over another (symmetry) may be related to what can be computed by networks of locally connected units.

Next, we consider the effect of size on Figure/Ground organization. Graham (1929) found that smaller regions tend to be perceived as Figure. To isolate the effect of size in our model, we used images comprised of parallel strips of alternating widths. The strips differed only in size, i.e., no differences could arise from closure or convexity effects. We created a set of images with varying relative strip sizes. For each image we computed the Figural entropy for the two globally consistent solutions. The results are shown in figure 8. For all images except when the strips were of equal width, the Figural entropy was significantly lower when the narrower strips were F. This effect grows as the difference in size increases.

What is the cause for the effect of size? As in the case of convexity, the difference in entropies results from differences in decay from the $1, 0$ anchored values near the edges. However, here the difference does not come from the *rate* of decay: as was shown in figure 6c, for a straight edge the rate of decay is equal on the F and G sides. Instead, the effect of size arises because the decay extends further for the wider strips, resulting in a higher proportion of units with $P(k)$ far from 1 or 0. As discussed in section 2.3.2, the decay occurs because the non-border units have an anchoring value of 0.5.

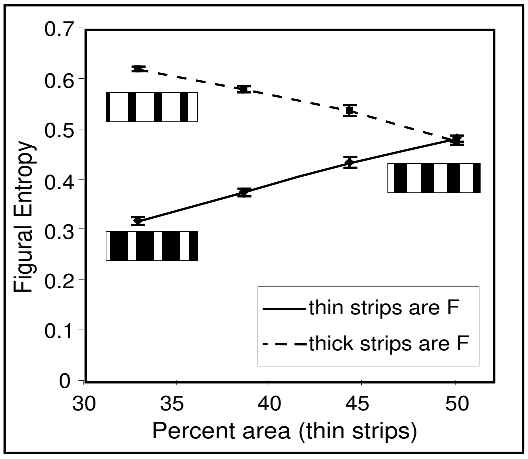

Figure 8: Effect of relative size on F/G organization. The entropy of the two globally consistent solution is shown for images with different ratios of strip widths. Similar to perception, the model shows a preference for the thinner strips to be Figure.

The third term in equation (1), which penalizes for deviations from this anchored value, is weighted by the small parameter ν. The rate of decay therefore depends on the value of ν. Large values will lead to fast decay; as ν is taken to be smaller and smaller, there will be more and more propagation of the anchored values $(1, 0)$ from the border units into the depth of the regions. Therefore, the graphs shown in figure 8b will change with a different choice of ν. Nevertheless, as long as ν is non-zero, there will be an effect of size. The perceptual effect of size therefore supports the notion that non-border units should have some small "inertia" towards an undecided F/G state. The specific way of implementing this may take different forms. In our model, ν was set for this image and then its value was chosen according to a specific scaling law for the other images; see Appendix B). More experimental data are needed in order to constrain ν or, more generally, explore the implementation of F/G decay away from edges (as well as its potential dependence on scale.)

The images we have considered so far differed in several ways, but there is also an important property which they all shared. In all cases, the contours that bound the regions perceived as Figure coincided with the luminance (and/or color) edges in the image. In terms of the model, this meant that these images yielded solutions that were always in perfect agreement with both Principles (i) and (ii): there were F/G boundaries along all luminance edges, and there were no F/G boundaries elsewhere. (In the case of ambiguous images, the polarity of the F/G boundaries could reverse, perceptually as well as in the model, but the boundaries always coincided with the luminance edges in the image.) There are, however, images that give rise to percepts which deviate from one or both Principles (i) and (ii) in places. The first class of such images we consider are those that contain figures outlined by open contours. Outlined figures can give rise to a sense of an enclosed region even when this region is not bound by a luminance edge all around. A simple example is shown in figure 9A. Observers report perceiving an enclosed, near-elliptical Figural region. Although it may be difficult to introspect what happens (perceptually) in the vicinity of the gap, formally there must be a transition between F for units inside and G for those outside. This is at odds with Principle (ii), which stated that F/G transitions are unlikely where no luminance edges are present. But the model is nevertheless capable of handling this situation. The reason is that the implementation of Principle (ii) via minimization of $\mathcal{E}$ allowed for deviations from it: F/G transitions away from edges are penalized, but they are not banned.

To run the model on this image, we treated the outlining contour the same way we have treated luminance edges so far. Panels B and C show two globally consistent candidate organizations with the Figure units on the inside versus outside, respectively. Panel D shows a globally inconsistent organization where the polarity of the F/G assignments

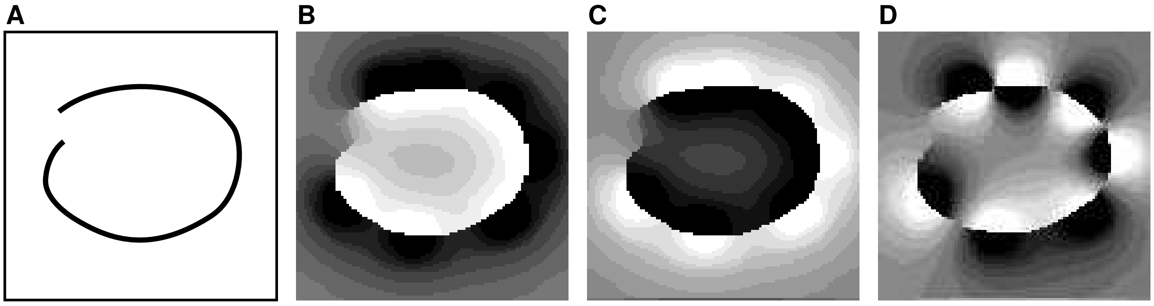

Figure 9: How the region-based model interprets figures outlined by open contours. *A:*, Perceptually, the image gives rise to a sense of an enclosed Figural region even though it is not bound by a luminance edge all around. *B, C:* Globally consistent candidate organizations where the F units were set on the inside and outside, respectively. *D:* A globally inconsistent organization.

reverse between different anchoring operators. The qualitative resemblance to the three candidate organizations shown for the ellipse (panels 4D-F) is evident. The entropies of the three organizations in 9B-D are 0.46, 0.71 and 0.73, respectively. The model thus produces a nearly-elliptical Figural region (panel B), in accordance with perception. Upon closer inspection, traces of the gap in the outlining contour can be seen in the solution. Across the two sides of the outline, there are sharp F/G transitions also in the portions where no anchoring operators were activated. (This is because the coefficients μ_{kj} in the cost function, eq. (1), equal zero when units k and j are separated by an edge.) In contrast, between the two ends of the outline, the transition between F ($P(k) > 0.5$) and G ($P(k) < 0.5$) is more gradual. This is because, in the absence of a luminance contour there, the coefficients μ_{kj} are non-zero, enforcing a smooth transition between F and G.

Region-based models thus offer a natural way to interpret how fragmentary edge information can give rise to complete, enclosed Figural regions. Importantly, this can happen even in the absence of "good continuation" of the contour fragments – as in the case for figure 9a. This is in contrast with contour based models, which do not offer a natural way to complete regions when good continuation is absent. The advantage of the region-based approach follows from its stated goal, to find Figural regions, and the resulting distinction it makes between luminance edges and F/G transitions.

Another class of images that give rise to percepts where edge information and F/G transitions do not coincide – i.e., when Principles (i) and/or (ii) are violated – is those that give rise to globally inconsistent F/G organizations. An example is shown in figure 10A. Observers report that there are two "lobe-shaped" Figural regions, a blue one on the left and a yellow one on the right. Following the single luminance edge that bounds these two regions, it is evident that the percept just described involves a reversal of the F/G polarity at some point. This suggests that global consistency is not, in general, an absolute perceptual constraint, and that other factors may override it. This important observation was already made by Rubin (1921, cf. figure 2).

To study the behavior of the model for this image, we compare several different candidate organizations. We first consider a globally consistent organization, shown in panel B. Here, the yellow region was taken as Figure, leading to the right-hand lobe being Figure and the left lobe as part of the background. This organization (which, as already mentioned, does not correspond to perception) yields a Figural entropy of 0.71 ± 0.02. Next, we flipped the F/G assignments of the anchoring operators around the left lobe, to

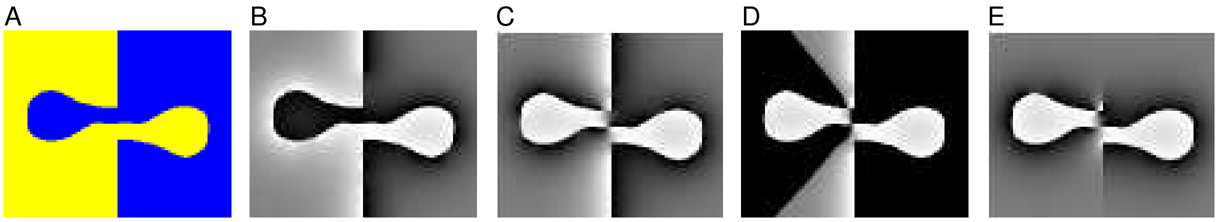

Figure 10: *A:* This image gives rise to a globally inconsistent Figure/Ground organization. Observers report perceiving two lobe-shaped Figural regions, a blue one on the left and a yellow one on the right, but this necessitates a reversal of the F/G polarity along the continuous contour that bounds both. *B:* A globally consistent candidate organization. *C:* When the polarity of the F/G assignments around the left lobe is reversed, the organization is not longer globally consistent, but nevertheless yields lower entropy. *D:* The output of the model based on the candidate organization in *C* yields, in addition to the two Figural lobes, regions which are not observed perceptually. *E:* The organization observed perceptually is achieved when anchoring operators are suppressed from enforcing F/G flips along the vertical edges that separate the blue and yellow sides.

make the resulting organization better resemble the reported percept. This is shown in panel C. The Figural entropy is lower: 0.56 ± 0.02. This organization, however, still differs somewhat from that observed perceptually. This can be seen more easily in panel D, where the background units ($P(k) < 0.5$) were all set to black, making the F/G transitions clearly visible. In addition to the two lobes, this organization contains pie-shaped Figural regions which are not observed perceptually. Moreover, parts of the F/G transitions around these regions occur where there are no luminance edges in the image. As discussed previously, such "spurious" transitions are an inevitable result of F/G assignments which are globally inconsistent. (Flipping the F/G assignments of the units around the left lobe forced a transition between the G units outside it and the F units along the left sides of the vertical edges.) The fact that the organization in panel C nevertheless yielded a lower entropy than that in panel B reflects the model's strong preference for enclosed, smaller regions to be the Figure. The mutual reinforcement of Figure units inside the lobes more than compensated for the penalty incurred for the 'gray' ($P(k) \simeq 0.5$) units along the spurious F/G transition.

Enforcing F/G transitions along all edges, including the vertical ones, was done in order to adhere to Principle (i), which stated that F/G transitions are likely along luminance edges. The reported percept, however, suggests that for the image in panel A, this principle may need to be reconsidered. Reporting the two lobes as the Figures in the image implies, among other things, that the vertical luminance edges are not perceived as F/G transitions. We therefore tested what happens if we suppressed the anchoring operators along the vertical edges. The result of minimizing equation (1) for this modified set of F/G assignments is shown in panel E. The entropy of this organization is 0.34 ± 0.04. Note, however, that in spite of its obvious advantage, this organization will not be obtained by running the version of the model described here. The present version automatically enforces F/G flips across all luminance edges (i.e., it can only give organizations like those in panels B and C). Panel E was obtained by manually suppressing select luminance edges from being loci of F/G transitions, as guided by our knowledge of the percept. In subsection 2.2, we mentioned that there should be mechanisms that allow the suppression of certain luminance edges from being considered as F/G transitions. There, we gave as an example the case of texture. The image in figure 10A serves as a reminder that, more generally, the visual system also has the means to admit the possibility of sharp luminance variations on a single surface (due to a change of surface property, e.g., paint color, or to illumination, e.g., shadows), or the existence of two abutting surfaces. The present model focuses on how to assign F/G polarity to edges, which is a major component in segmenting the image. But clearly, the ultimate interpretation of any image must involve many other factors, and possibly more than one iteration, until a self-consistent, ecologically valid interpretation of the scene is reached.

4. **Discussion**

We presented a model for Figure/Ground (F/G) segregation. The starting point of the model is the need to determine which of the two sides of an edge is in front (the Figure), and which side is in the background. This "border ownership" problem is particularly severe when the units have only local information (small receptive fields), since its resolution requires integrating information from an image region approximately of the size of the Figure. Zhou et al. (2000) reported that early visual cortical cells (V1/V2/V4) show sensitivity to the polarity of border ownership induced by Figures much larger than their "minimum response fields". We therefore asked whether those effects could be accounted for by computations performed in those early areas, or whether receiving global information via feedback from higher areas was necessary. We found that Figure/Ground segregation can indeed be computed by a model network of small receptive-field units which interact only locally (nearest-neighbor connections). Importantly, the model shows sensitivity to global image properties in a way similar to human perception: it prefers enclosed, smaller or "more convex" regions as the Figure. This suggests that the sensitivity to Figure/Ground polarity found by Zhou et al. (2000) for early cells (V1, V2 and V4) may be computed in those regions, i.e., it may not necessitate feedback from higher areas.

4.1. **Timing considerations.** What allows the model to show global effects despite having only local connections is that signals can propagate through the network iteratively. The iterations lead to long-range propagation of information as the system evolves with time. This raises a natural question: can such a model be fast enough to account for the rapid F/G segregation effects observed experimentally? Zhou et al. (2000) concluded from their results (figure 20) that "the cortical processing that leads to border-ownership discrimination requires no more than ~ 25 msec." This poses a challenge to iteration-based models such as ours, but does not exclude them a-priori. The number of iterations needed for global effects to emerge is no more than half the number of units that span the most distant points in the Figure. Here, we restricted the model units to be of a single scale, to keep things simple. But in reality computations like those described here would need to be performed at multiple scales, including somewhat coarse ones. Cross-talk between the different scales could then allow the system to identify the appropriate scale for analysis of a given Figure – that which converges rapidly to the most stable solution. In such a more realistic system, a relatively small number of iterations may suffice for a wide range of Figural surfaces. For modestly sized Figures, say those bound within 5°, as few as 2-3 iterations can provide an adequate solution even with 1° receptive field units. For larger Figural regions, the computations would be carried out by units with larger receptive fields, which are known to exist in early cortex (at least 3° parafoveally in V2, cf. Foster et al. 1985; see also Van Essen et al. 1984; Maunsell and Newsome 1987; Sceniak et al. 2001.) Given estimates of the extent of lateral connections, and how rapidly those signals may spread (T'so et al. 1986; Grinvald et al. 1994; Das and Gilbert 1995; see also Movshon and Newsome 1996; Girard et al.

2001; Hupe et al. 2001 for between-area iterations times), Figural regions as large as 30° may still be resolved within time frames of the order of 25 msec. Nevertheless, even with multiple-scale computations, iteration-based models predict a trend of slower convergence for computations on larger regions (cf. Paradiso and Nakayama 1991.) Further behavioral and physiological experiments are needed to test this prediction, that the time required for border-ownership resolution should grow with Figure size.

4.2. **Region-based processes in physiology and perception.** An important property of the present model is that it is region-based. The iterative propagation of signals gives rise to global Figure/Ground effects because these signals are allowed to propagate into homogenous (edge-free) regions in the image (recall figure 3). Although region-based computations have not been emphasized in experimental or theoretical studies in vision, there is evidence for the existence of such processes. Cells whose receptive fields fall within homogeneous regions can show differential responses if the apparent brightness of the region they represent is affected by changes to the far surround (Rossi et al. 1996; MacEvoy et al. 1998; Rossi and Paradiso 1999.) Several other studies (Lamme 1995; Zipser et al. 1996; Lee et al. 1998; Lamme et al. 1999) showed that cells within homogeneous, or homogeneously textured regions exhibit heightened activity if those regions belong to a Figure (but see Rossi et al. 2001.) The increased activity had a latency of 30-40 ms relative to the onset of cells' responses. Interestingly, some of the later studies reported that the increased activity was greater near the edge than away from it, towards the center of the Figure (Lee et al. 1998; Lamme et al. 1999; Rossi et al. 2001) – similar to what happens in our model.

Psychophysical studies have also focused on the role of contour-based processes for segmentation, especially for perceptual completion of occluded and illusory surfaces (Kellman and Shipley 1991; Ringach and Shapley 1996; Rubin 2001b; but see Mumford et al. 1987). Nevertheless, there are perceptual phenomena that are more naturally explained by positing region-based processes. As shown in this paper, the preference for enclosed, convex regions to be perceived as Figure is one such example. More examples can be found in the domain of illusory contours. Consider panels A and B in figure 11: both show a group of lines which terminate along a circular arc. But observers report quite different percepts in the two cases: the terminators in panel A induce an illusory contour (IC) which bounds an occluding white surface, while such an IC is not reported for panel B. Models of how ICs are induced by line terminators have emphasized computations along the contour defined by the terminators (Heitger et al. 1992). But such a contour-based approach cannot account for the perceptual difference in panels A and B, since the shape of the contour along the lines' terminators is identical in the two cases (panel C). In contrast, if signals propagate not only along the contour, but also into the region posited as an occluder, there would be more mutual enhancement in A (see panel D), similarly to how a preference for convexity arises in our region-based model. Other examples for IC phenomena that are more naturally explained by a region-based approach can be found in Gillam (1987; figures 30.8, 30.9). Indeed, the emergence of illusory contours is intimately related to the border ownership problem (Nakayama et al. 1995), and therefore it is quite likely that region-based processes play a part in IC-related computations as well.

From a computational point of view, region-based models have several attractive features, which explains their recent popularity in computer vision (e.g., Zhu et al. 1995; Shi and

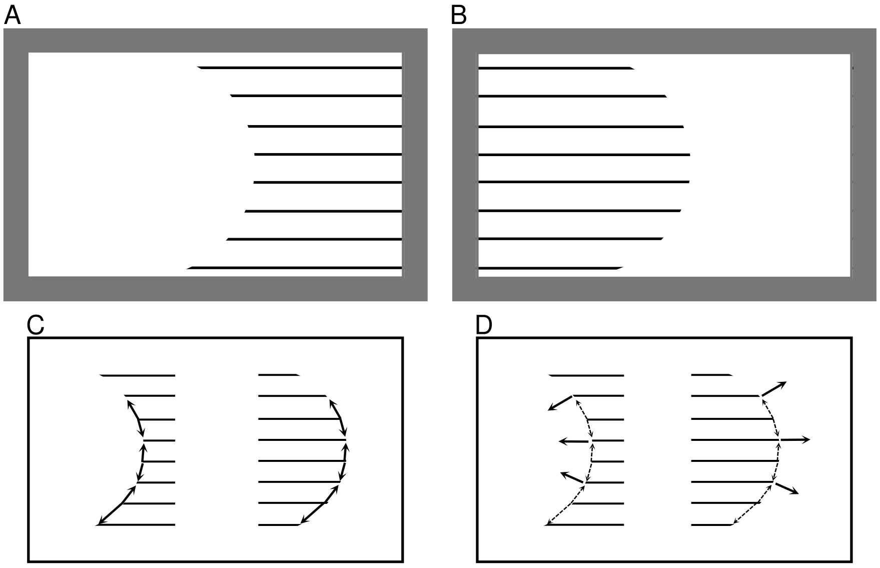

Figure 11: Region-based effects in the formation of illusory contours (ICs). The terminators in panel A give rise to an IC which bounds an occluding white surface on the left. In contrast, no IC is observed in panel B. The difference in the perception of the two images cannot be explained by a purely contour-based approach, as the contours traced by the terminators are identical in the two cases (panel C). Positing region-based computations, on the other hand, provides a natural explanation (panel D.)

Malik 1997; Geiger et al. 1998; Sharon et al. 2000). In general, in such models the output of the computation is a collection of pixels, grouped together and labeled as a 'region'. Such regions are naturally bounded by closed contours, and therefore region-based models can be less sensitive to image noise that interrupts parts of a luminance edge. (Recall, for example, figure 9, which readily yielded a Figural region in our model, but would pose a problem for contour based models.) Furthermore, the bounding contours of the regions produced as output are globally consistent by definition, thus eliminating the need to test for self-consistency of traced contours as valid surface boundaries (Williams and Hanson 1996; Williams 1997.)

4.3. **Probability and neural representation.** A key feature of the model is the use of intermediate values to represent probabilities that image locations are Figure or background. It is therefore natural to wonder about the neural plausibility of such a probabilistic scheme. While in the present model we do not commit to any specific neural implementation, a few possibilities can be suggested. One, there may be a single population of neurons whose role is to signal whether the location they correspond to is part of the Figure ($P(k) = 1$), in which case they fire at a maximal rate, or in the background ($P(k) = 0$), in which case they are quiescent. In this case, intermediate F/G values would mean that the neurons are firing less than their maximal rate. Another possibility is to have two populations of cells: at each location, there would be a corresponding "F neuron" which fires maximally if the location is part of the Figure and a "G neuron" which fires if the location is part of the background. The F and G neurons would be mutually inhibitory since, ideally, only one of them fires, corresponding to $P(k)$ values of 1 and 0. The less desirable case in which both neurons fire (at intermediate rates) would correspond to an intermediate value of $P(k)$. This latter scheme is similar to that favored by Zhou et al. (2000) to account for their results (see their figure 28).

Interestingly, this scheme can offer an alternative to the "temporal binding" hypothesis for how to encode that different neurons belong to the same surface (cf. Finkel and Sajda 1992; Wang and Terman 1997; Roskies 1999 and references therein). Instead of using correlated firing as a code, the fact that a collection of F neurons are active simultaneously may itself signal that they belong to the same surface, even for non-contiguous populations. (At its simplest form, however, this mechanism only allows the signaling of one Figural surface at any given moment.)

The probabilistic nature of the activity of units in the model has an appealing advantage: it offers a natural way to incorporate additional cues in a distributed way. In the present version, the setting of the F/G assignments by anchoring operators was unbiased, i.e., there were equal chances for F to be on either side of an edge. But there may be other sources of information in the image (or image statistics) about which side is more likely to be in front. Such information is easily incorporated into the model by biasing the probability in favor of a certain side to be F. The most obvious example is stereoscopic depth, which can disambiguate the relative depth of (all or parts of) an edge. There are also monocular (pictorial) cues for depth stratification, such as T-junctions (Rubin 2001b) and concave cusps (Stevens and Brookes 1988), which could easily be incorporated in a probabilistic scheme. Global biases about depth relationships can be implemented also. For example: the lower parts of images tend to be perceived as nearer than the upper parts (presumably reflecting the learned statistics of the ground plane.) This effect can override the preference for convex regions to be Figure: when a region is divided horizontally by a curved edge, the bottom part is often seen as Figure even if it lies on the concave side of the edge. This bias can be implemented in the model by having $P_0(k)$ change as a function of the y value (height) of the unit in the image, i.e., by introducing priors about the probability of lower and upper parts in the image to be Figure. In this regard, the model presented here is related to "Bayesian networks" and "belief nets" which have been used for image analysis in several domains (Weiss 1997; Freeman and Viola 1998).

4.4. **Further extensions to the model.** The model presented here was kept as simple as possible in order to isolate and highlight the main ideas it implements. To make it a realistic model of how F/G segregation may be implemented in the brain, however, it needs to me extended in many directions. Some of these extensions have already been mentioned. Below, we revisit the most important one and then list further directions to extend the model which were not discussed before.

An important modification to the model was mentioned in section 2.2. The present, two-stage version requires a search through the set of candidate organizations to find the low-entropy ones. This was done in order to present the principles of the model in their simplest form, but it is not biologically plausible. A one-stage model which minimizes a cost function that incorporates both $\mathcal{E}$ and $\mathcal{S}$ can be achieved by introducing local interactions which favor consistent polarity of F/G assignments between neighboring anchoring operators (Pugh and Rubin, in preparation).

Another limitation of the present version of the model is that it does not represent more than two layers of depth at a time (the Figure and the Ground), and this obviously needs to be addressed in order to handle real-world occlusion situations. (This does not necessarily imply a need to invoke numerous layers of F/G units: an alternative approach may be to combine the model with an attentional/control module, so that detailed segmentation is performed on only a piece of the image at a time. In support of this idea, there is evidence that human observers are able to perform only rather crude segmentation on "busy" images that contain many surfaces; cf. Gurnsey et al. 1996) A related problem is that the present model does not have a mechanism to allow the background (or, more generally, occluded surfaces) to continue behind the Figure. This is necessary for correct recovery of the shape of surfaces, including linking disjoint region into a unitary surface. Such "perceptual completion" processes clearly take place in human vision (E. Rubin 1921, 1958; Nakayama et al. 1995; Driver and Baylis 1996), and neural correlates of it have recently been reported (Baylis and Driver 2001; Kourtzi and Kanwisher 2001; Rubin 2001a). Future extensions of the model should therefore incorporate completion mechanisms.

The region-based model we presented here makes minimal contact with contour-based computations: the only way in which contours were used was to determine where F/G transitions are likely to occur (Principle i). However, in reality it is likely that there is intensive cross-talk between contour-based and region-based computations. One example of how such cross-talk may significantly improve the performance of the model is by making use of local contour-orientation information. In the present version, signals propagate isotropically in all directions (cf. figure 3). Preliminary results suggest that if stronger signals are sent in the direction orthogonal to the local edge orientation, the mutual enhancement on the convex/enclosed side of the edge is more pronounced (Pao 2001). Such "directional diffusion" therefore offers a way to significantly enhance F/G resolution, and possibly also accelerate convergence.

Finally, the present model focused entirely on how Figure/Ground resolution may be achieved in a stimulus-driven way, relying solely on the geometrical properties of shapes in the image. But it is known that high-level factors such as object knowledge and attention can affect F/G resolution (Peterson and Gibson 1994; Driver and Baylis 1998). Any ultimate model of segmentation would undoubtedly need to incorporate "top-down" information alongside "bottom-up" computations like those described here. In addition to enabling high-level effects to emerge, feedback from higher cortical areas may also facilitate the computations in complex situations, e.g. for very large Figures or when the image is very cluttered (cf. Kienker et al. 1986).

Acknowledgments: We thank Davi Geiger, Jean-Michel Hupe, J. Anthony Movshon, Ken Nakayama, Robert Shapley, Eitan Sharon, Shimon Ullman and Yair Weiss for helpful discussions and comments on the manuscript. NR is supported by NSF grant IBN-9720305 and an Alfred P. Sloan Fellowship. MP is supported by NSF grants DMS97-21430 and DMS99-71392 and an Alfred P. Sloan Fellowship.

References

Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. J Neurosci 20:8188-8198.

Baumann R, van der Zwan R, Peterhans E (1997) Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. European Journal of Neuroscience 9:1290-1303.

Baylis GC, Driver J (2001) Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. Nat Neurosci 4:937-942.

Cover TM (1991) Elements of information theory. New York: John Wiley & Sons, Inc.,.

Das A, Gilbert CD (1995) Long-range horizontal connections and their role in cortical reorganization revealed by optical recording of cat primary visual cortex. Nature 375:780-784.

Dongarra JJ, Bunch JR, Moler CB, Stewart GW (1979) LINPACK Users' Guide. Philadelphia: SIAM.

Driver J, Baylis GC (1996) Edge-assignment and figure-ground segmentation in short-term visual matching. Cognit Psychol 31:248306.

Driver J, Baylis GC (1998) Attention and visual object segmentation. In: The Attentive Brain (Parasuraman R, ed), pp 299325. Cambridge, MA: Mit Press.

Elder J, Zucker SW (1993) Contour Closure and the Perception of Shape. Vision Research 33:981-991.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for a local association field. Vision Research 33:173-193.

Finkel LH, Sajda P (1992) Object discrimination based on depth-from-occlusion. Neural Computation 4:901-921.

Foster KH, Gaska JP, Nagler M, Pollen DA (1985) Spatial and temporal frequency selectivity of neurones in visual cortical areas V1 and V2 of the macaque monkey. J Physiol 365:331363.

Freeman WT, Viola PA (1998) Bayesian model of surface perception. Neural Information Processing Systems 10:787-793.

Geiger D, Pao K, Rubin N (1998) Salient and multiple illusory surfaces. Proc. IEEE Conf. Comp. Vis. Patt. Rec. June '98, Santa Barbara:118-124.

Gillam B (1987) Perceptual grouping and subjective contours. In: The Perception of Illusory Contours (Petry S, Meyer GE, eds), pp 268-273. New York: Springer-Verlag.

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 85:1328-1331.

Graham CH (1929) Area, color and brightness difference in a reversible configuration. J. General Psychology 2:470-481.

Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by realtime optical imaging of macaque monkey primary visual cortex. J Neurosci 14:2545-2568.

Grossberg S (1997) Cortical dynamics of threedimensional figure-ground perception of twodimensional pictures. Psychol Rev 104:618-658.

Gurnsey R, Poirier F, Gascon E (1996) There is no evidence that Kanizsa-type subjective contours can be detected in parallel. Perception 25:861-874.

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 32:963-981.

Heitger F, von der Heydt R (1993) A computational model of neural contour processing: figure-ground segregation and illusory contours. Proc. Int. Conf. Comp. Vis. 4:32-40.

Hildreth EC (1984) The measurement of visual motion. Cambridge, Mass: MIT Press.

Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215-243.

Hupe JM, James AC, Girard P, Bullier J (2001) Response modulations by static texture surround in area V1 of the macaque monkey do not depend on feedback connections from V2. J Neurophysiol 85:146-163.

Iverson L, Zucker SW (1995) Logical/linear Operators for Image Curves. IEEE Trans. Pattern Analysis and Machine Intelligence 17:982-996.

Kanizsa G (1979) Organization in Vision. New York: Praeger.

Kanizsa G, Gerbino W (1976) Convexity and symmetry in figure-ground organization. In: Vision and Artifact (Henle M, ed), pp 2532. New York: Springer Co.

Kellman P, Shipley T (1991) A theory of visual interpolation in object perception. Cognitive Psychology 23:141-221.

Kienker PK, Sejnowski TJ, Hinton GE, Schumacher LE (1986) Separating figure from ground with a parallel network. Perception 15:197-216.

Kimia B, Tannenbaum A, Zucker SW (1995) Shapes, Shocks, and Deformations. Intl J. Computer Vision 15:189-224.

Koffka K (1935) Principles of Gestalt Psychology. New York: Harcourt Brace & Co.

Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. Science 293:1506-1509.

Lamme VA (1995) The neurophysiology of figure-ground segregation in primary visual cortex. J Neurosci 15:1605-1615.

Lamme VA, Rodriguez-Rodriguez V, Spekreijse H (1999) Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. Cereb Cortex 9:406-413.

Lamme VAF, Zipser K, Spekreijse H (1997) Figure-ground signals in V1 depend on extrastriate feedback. Invest. Ophthal. & Vis. Sci. (Suppl.) 38:4490.

Lee TS, Mumford D, Romero R, Lamme VA (1998) The role of the primary visual cortex in higher level vision. Vision Res 38:2429-2454.

Liu Z, Jacobs DW, Basri R (1999) The role of convexity in perceptual completion: beyond good continuation. Vision Res 39:4244-4257.

MacEvoy SP, Kim W, Paradiso MA (1998) Integration of surface information in primary visual cortex. Nat Neurosci 1:616-620.

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

Maunsell JH, Newsome WT (1987) Visual processing in monkey extrastriate cortex. Annu Rev Neurosci 10:363-401.

Minguzzi GF (1987) Anomalous Figures and the Tendency to Continuation. In: The Perception of Illusory Contours (Petry S, Meyer G, eds), pp 71-75. New York: Springer-Verlag.

Movshon JA, Newsome WT (1996) Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J Neurosci 16:7733-7741.

Mumford D, Kosslyn SM, Hillger LA, Herrnstein RJ (1987) Discriminating figure from ground: the role of edge detection and region growing. Proc Natl Acad Sci U S A 84:7354-7358.

Nakayama K, He ZJ, Shimojo S (1995) Visual surface representation: a critical link between lower-level and higher-level vision. In: Visual Cognition (Kosslyn SM, Osherson DN, eds), pp 1-70. Cambridge, MA: MIT press.

Nitzberg M, Mumford D, Shiota T (1993) Filtering, segmentation, and depth. Berlin/New York: Springer-Verlag.

Pao H, Geiger D, Rubin N (1999) Measuring convexity for Figure/Ground separation. Proc. 7th IEEE Intl. Conf. Comp. Vision :948-955.

Pao HK (2001) A Continuous Model for Salient Shape Selection and Representation. In: Courant Institute for Mathematical Sciences: New York Univerity.

Paradiso MA, Nakayama K (1991) Brightness perception and filling-in. Vision Res 31:1221-1236.

Peterhans E, von der Heydt R (1989) Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. J Neurosci 9:1749-1763.

Peterson MA, Gibson BS (1994) Must figure-ground organization precede object recognition? An assumption in peril. Psychological Science 5:253-259.

Reichl LE (1998) A modern course in statistical physics, 2nd Edition. New York: John Wiley.

Ringach DL, Shapley R (1996) Spatial and temporal properties of illusory contours and amodal boundary completion. Vision Research 36:30373050.

Roskies AL (1999) The binding problem. Neuron 24:7-9.

Rossi AF, Desimone R, Ungerleider LG (2001) Contextual modulation in primary visual cortex of macaques. J Neurosci 21:1698-1709.

Rossi AF, Paradiso MA (1999) Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. J Neurosci 19:6145-6156.

Rossi AF, Rittenhouse CD, Paradiso MA (1996) The representation of brightness in primary visual cortex. Science 273:1104-1107.

Rubin E (1921) Visuell wahrgenommene Figuren. Copenhagen: Gyldendals.

Rubin E (1958) Figure and Ground. In: Readings in Perception (Beardslee DC, Wertheimer M, eds), pp 194-203. Princeton, NJ: D. Van Nostrand Company, Inc.

Rubin N (2001a) Figure and ground in the brain. Nat Neurosci 4:857-858.

Rubin N (2001b) The role of junctions in surface completion and contour matching. Perception 30:339366.

Sceniak MP, Hawken MJ, Shapley R (2001) Visual spatial characterization of macaque v1 neurons. J Neurophysiol 85:1873-1887.

Sha'ashua A, Ullman S (1988) Structural saliency: the detection of globally salient structures using a locally connected network. Proc. of 2nd Int. Conf. Comp. Vis. Clearwater, FL.:321327.

Sharon E, Brandt A, Basri R (2000) Fast Multiscale Image Segmentation. In: Conference on Computer Vision and Pattern Recognition, pp 70-77. South Carolina: IEEE.

Shi J, Malik J (1997) Normalized cuts and image segmentation. Proc. of IEEE Conf. on Comp. Vis. and Patt. Rec. Puerto Rico:731-737.

Stevens KA, Brookes A (1988) The concave cusp as a determiner of figure-ground. Perception 17:35-42.

Sugita Y (1999) Grouping of image fragments in primary visual cortex. Nature 401:269-272.

Ts'o DY, Gilbert CD, Wiesel TN (1986) Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. J Neurosci 6:1160-1170.

Ullman S (1976) Filling-in the gaps: the shape of subjective contours and a model for their generation. Biological Cybernetics 25:1-6.

Van Essen DC, Newsome WT, Maunsell JH (1984) The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. Vision Res 24:429-448.

von der Heydt R, Peterhans E, Baumgartner G (1984) Illusory contours and cortical neuron responses. Science 224:1260-1262.

Wang D, Terman D (1997) Image segmentation based on oscillatory correlation. Neural Computation 9:805-836.

Weiss Y (1997) Interpreting images by propagating Bayesian beliefs. In: Neural Information Processing Systems (Mozer MC, Jordan MI, Petsche T, eds), pp 908-915.

Williams LR (1997) Topological reconstruction of a smooth manifold-solid from its occluding contours. Int. J. Comp. Vis. 23:93-108.

Williams LR, Hanson AR (1996) Perceptual completion of occluded surfaces. Computer Vision and Image Understanding 64:1-20.

Williams LR, Jacobs DW (1997) Stochastic completion fields: a neural model of illusory contour shape and salience. Neural Computation 9:837-858.

Yen SC, Finkel LH (1998) Extraction of Perceptually Salient Contours by Striate Cortical Networks. Vision Research 38:719-741.

Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. J Neurosci 20:6594-6611.

Zhu SC, Lee TS, Yuille A (1995) Region competition: unifying snakes, region growing and MDL for image segmentation. Proc. Fifth Int. Conf. in Comp. Vis., :416-425.

Zipser K, Lamme VA, Schiller PH (1996) Contextual modulation in primary visual cortex. J Neurosci 16:7376-7389.

Appendix A. **Computing P**

The Algebraic approach

To find the function $\mathbf{P}$ which minimizes $\mathcal{E}$ we use the fact that the derivative of any well-behaved function is zero at points where the function is minimal (or maximal). This is true also for functions of more than one variable, such as $\mathcal{E}(Q(k))$. Therefore, we take the derivatives of $\mathcal{E}$ as a function of $Q(k)$ for all units k and demand that these derivatives equal zero:

$$\frac{\partial \mathcal{E}}{\partial \mathbf{Q}} = 0. \tag{A1}$$

The derivatives of $\mathcal{E}$ are given by

$$\begin{aligned} \frac{\partial E}{\partial Q(k)} &= 4 \sum_{j \in N_k} \mu_{kj} \left[Q(k) - Q(j) \right] \\ &\quad + 2\nu_k \left[Q(k) - P_0(k) \right]. \end{aligned} \tag{A2}$$

The first term comes from differentiating the first term in equation (1), using the symmetry $\mu_{kj} = \mu_{jk}$. N_k denotes the set of nearest neighbors of unit k, as before. The second term unifies the derivatives of the second and third terms of equation (1); the coefficient ν_k equals 1 if unit k is a border unit and ν otherwise. Referring back to figure 5, note that the coefficients of the first term in equation A2 correspond to the connection weights between unit k and its neighbors j in the bottom layer, and the coefficient in the second term corresponds to the connection from the unit storing $P_0(k)$ in the intermediate later to unit k. Thus, the local neighborhood of unit k in the network computes the value of $\frac{\partial \mathcal{E}}{\partial Q(k)}$, i.e., the effect of a small change in $Q(k)$ on the cost function $\mathcal{E}$.

Equation (A2) is linear in $\mathbf{Q}$, since it is the derivative of $\mathcal{E}$ which is quadratic in $\mathbf{Q}$. Therefore, equation (A1) can be solved as a set of N linear equations, determining the minimizer $\mathbf{P}$. This is the *algebraic* approach. In the generic case, such a system of N equations with N unknowns would take $O(N^3)$ operations to solve. But since each unit only interacts with its nearest neighbors, this makes the linear system in equation (A2) sparse. This allows the efficient methods developed specifically for sparse systems (Dongarra et al. 1979) to be applied. These methods find $\mathbf{P}$ quickly with $O(N)$ computations.

The dynamical-systems approach

The algebraic approach is useful for finding $\mathbf{P}$ numerically, but it has a significant disadvantage in terms of physiological plausibility. We would not want to suppose that a neural system uses sparse linear algebra (at least, not explicitly). Therefore, it is important to show that the computation of $\mathbf{P}$ has an alternative formulation, one that lends itself readily to network implementation. Eq. (A2) gives the value of the gradient of $\mathcal{E}$ for any $\mathbf{Q}$. Instead of equating the gradient to zero, as in the algebraic approach, we use it to define a *dynamical* system. In this case, $\mathbf{Q}$ evolves from one moment to the next in such a way as to decrease $\mathcal{E}$ the fastest. Specifically, for each unit k

$$Q(k, t + \Delta t) = Q(k, t) - \Delta t \cdot \frac{\partial \mathcal{E}}{\partial Q(k)} \tag{A3}$$

with $\frac{\partial \mathcal{E}}{\partial Q(k)}$ given by equation (A2). Since the network in figure 5 computes these derivatives (locally) for every k, equation A3 therefore shows that it will converge towards the state $\mathbf{P}$ which minimizes the cost function $\mathcal{E}$. In theory, one needs to let the system evolve infinitely to obtain this asymptotic value. However, for a desired degree of precision, $\mathbf{P}$ will be approximated in a finite number of time steps.

In the continuum limit ($N \to \infty$ with a fixed image size and $\Delta t \to 0$), eq. (A3) yields a partial differential equation (PDE) which is also used to describe heat flow in the presence of external heat sources. This is useful because it makes it possible to draw insights about the time evolution of the model from known properties of heat flow.

Appendix B. Numerical methods and parameters

The model was simulated with Matlab (The MathWorks, Inc., version 5.3.1.29215a, 1999.) The input image, represented by an $m \times n$ matrix $\mathbf{J}$ with values of 0 and 1, was edge-detected with the Canny operator. Anchoring operators were then placed on the image with a specified high density, disallowing overlap of two or more operators. Each anchoring operators was then checked to see if an edge fell within it, and if so, all units within it were labeled border units. The border units were given F/G assignments P_0 according to the following rule: for the two globally consistent organizations, each unit was given the value of the image $\mathbf{J}$ at the same location, or of its contrast-reversed image $1 - \mathbf{J}$, respectively. For all other organizations, for each anchoring operator the border units on one side of the edge received a value of 1 and those on the other side received 0, with a random decision which side receives which value. Finally, all non-border units receive a value of $\mathbf{P}_0 = 0.5$.

Next, we find $\mathbf{P}$ which minimizes the cost function $\mathcal{E}$ (eq. 1). As discussed in subsection 2.3.2, this corresponds to solving the linear system given by equations (A1-A2).

Treating the unknown $\mathbf{P}$ as a vector of length $N(= m \times n)$, those equations can be rewritten as $\mathbf{A}\vec{P} = \vec{c}$, where $\mathbf{A}$ is a symmetric $N \times N$ matrix and $\vec{c}$ is a vector of length N. The matrix $\mathbf{A}$ is sparse, with at most five nonzero entries in each row. Using the notation k_{west}, k_{east}, k_{north} and k_{south} for the four nearest neighbors of unit k (their indices are $k-1$, $k+1$, $k-n$ and $k+n$, respectively; see figure 5), the k-th row in the linear system is:

$$\begin{aligned} &\mathbf{M}_{west}(k)\,[P(k_{west}) - P(k)] \\ &\quad + \mathbf{M}_{east}(k)\,[P(k_{east}) - P(k)] \\ &\quad + \mathbf{M}_{north}(k)\,[P(k_{north}) - P(k)] \\ &\quad + \mathbf{M}_{south}(k)\,[P(k_{south}) - P(k)] \\ &\quad - V(k)P(k) = -V(k)P_0(k). \end{aligned} \tag{B1}$$

The diffusion matrices are defined by:

$$\begin{aligned} \mathbf{M}_{west}(k) &= \mu\, J(k_{west})J(k) \\ &\quad + \mu\,(1 - J(k_{west})(1 - J(k)) \end{aligned} \tag{B2}$$

	image size	lengthscale	ν	extension
Fig.1	200×200	30 pixels	0.000556	27 pixels
Fig.6	70×210	70 pixels	0.0002	10 pixels
Fig.7	67×331	50 pixels	0.0002	36 pixels
Fig.8	210×210	30 pixels	0.000556	27 pixels
Fig.10	180×220	50 pixels	0.0002	36 pixels
Images used for illustrative purposes				
Fig.4	100×100	50 pixels	0.0004	36 pixels
Fig.9	86×80	50 pixels	0.0004	36 pixels

Table 1:

with $\mathbf{M}_{east}$, $\mathbf{M}_{north}$ and $\mathbf{M}_{south}$ defined analogously. They implement the requirement that the value of μ_{kj}, the connection between neighboring units k and j, should be zero if they are separated by an edge, and μ otherwise (first term in eq. 1). $V(k)$ is 1 for border units and ν for non-border units, implementing the second and third terms in eq. (1).

Equation (B1) holds for each unit that is not on the boundary of the network (i.e., $n < k < n(m-1)$ and $((k-1) \bmod n) \neq 0, n-1$). For units on the boundary, one has to choose what boundary conditions to apply, thus determining the remaining $2n + 2(m-2)$ equations. We used Dirichlet boundary conditions: $\mathbf{P}(k) = 0.5$ for all k on the boundary. As a result, the P values go to 0.5 as one approaches the boundary. To prevent the boundary conditions from affecting the output of the model, we solved for $\mathbf{P}$ on an 'extended' version of the image which contained an additional band of units around the boundary, and then 'cropped' it back before computing the Figural entropy. (To obtain the extended version of an image we simply defined the target image as a cropped version of a larger image.) The width of the band was determined empirically to be such that the effect of the boundary dissipated faster than the band width. (The effect of the boundary was considered eliminated if the change in the entropy produced by extending the band further was in the fourth significant digit or higher.) All the results shown in paper are for the cropped images. Table 1 lists the widths of the band used for each image.

Finally, the entropy was computed according to eq. (2). When confidence intervals are given, they were computed by generating ten sets of anchoring operators, producing P_0 for each according to the case at hand (globally consistent with polarity of $\mathbf{J}$ or $1 - \mathbf{J}$, or globally inconsistent), and computing $\mathbf{P}$ and the resulting entropy for each. The values given for the entropy represent mean $\pm$ 2std.

Free parameters. Their values were chosen as follows:
The anchoring operator were always of diameter 7 units, except for figures 4 and 9 where they were made larger (15 units) for illustrative purposes.
The density of the anchoring operators was 0.88 ± 0.05 operators per 100 units (the variability arises from the need to eliminate overlapping operators.) This value holds for all the results shown, except for figures 4 and 9 where they were placed manually.
The parameter μ was set to 0.1 for all the simulations shown here.
The choice of the parameter ν, which controls the rate of decay of $P(k)$ away from the edges, is related to the scale of the image (as measured by number of units, not physical size.) Consider the image in figure 8. It is 210×210 pixels and was processed by a network of that many units. If the same image was doubled in resolution and processed by a network of 420×420 units without changing ν, the entropy value for the best candidate organization (as well as all other organizations) would be different. The reason is that the new Figural regions extend twice as many units and therefore would suffer

comparatively more decay than before, leading to higher Figural entropy. But this higher value is an artifact of the different scales of the images, i.e., it is not truly representative of a stronger Figural organization in one of them. To preserve Figural entropy when scaling an image, ν should be scaled by $(l/l_n)^2$, where l and l_n are the old and new lengthscales, respectively. (Strictly speaking, the diameter of the anchoring operators needs to be scaled, too, but we did not do this as we found it had a negligible effect on the results.) Generally, the appropriate lengthscale of an image is determined not by its overall size but rather by the size of the Figural regions. The decay rate should be scaled with respect to the Figure if one is to make meaningful comparisons between the performance of the model on different images. We fixed the value of ν at 0.0002 for figure 6 and then scaled it for other images by their lengthscales as given in Table 1. (As mentioned in the Discussion, ultimately the computations would need to be done at multiple scales, with cross-talk between the different scales to choose the best organization. For the present purpose of explaining the principles of the model, however, we pre-selected the relevant scale for each image manually.)