Form Perception

Stephen Grossberg[1]

Department of Cognitive and Neural Systems
Center for Adaptive Systems
and
Center of Excellence for Learning in Education, Science, and Technology
Boston University
677 Beacon Street, Boston, MA 02215, USA

Running title: Form Perception

Key Words: Object perception, Perceptual grouping, Surface perception, Figure-ground perception

December, 2007

Technical Report CAS/CNS-TR-07-020

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

Form Perception

Stephen Grossberg
Department of Cognitive and Neural Systems
Center of Excellence for Learning in Education, Science and Technology
Boston University
677 Beacon Street, Boston, MA 02215
steve@bu.edu
http://www.cns.bu.edu/Profiles/Grossberg

Synonyms
Object perception
Perceptual grouping
Surface perception
Figure-ground perception

Definition
Form perception refers to our ability to visually perceive objects in the world in response to the patterns of light that they caste on our retinas.

Characteristics
When an observer gazes steadily at a stationary object, form perception is facilitated by miniature eye movements, such as tremors, drifts, and microsaccades, that cause visually-responsive cells to respond more vigorously to the object whenever the eyes move [1]. Motion of an observer relative to the object also refreshes cell responses. Even when cells respond vigorously, form perception represents a major challenge for the brain because our retinas have a large blind spot and retinal veins that impede light from reaching photodetectors (Figure 1). How does the brain compensate for these holes in the visual world?
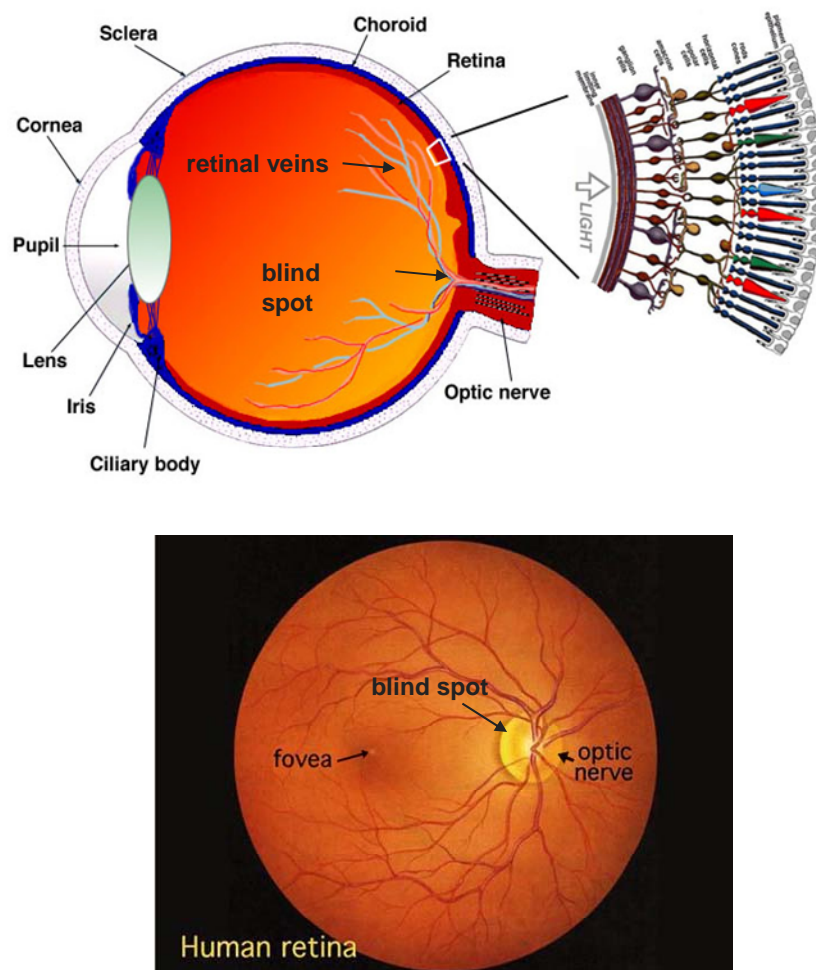
Figure 1. A side view of an eye shows the blind spot and retinal veins. A top view of the photo-sensitive retina shows how big the blind spot and veins are relative to the fovea, which has the highest resolution for form vision.

Boundary and surface processes facilitate this goal. Boundary processing includes perceptual grouping, boundary completion, and figure-ground separation. Surface processing includes compensation for variable illumination, also called "discounting the illuminant", and surface filling-in using the surviving illuminant-discounted signals. These processes are carried out in different processing streams within the visual cortex (Figure 2). Both streams go through the Lateral Geniculate Nucleus (LGN), which transmits signals from the retina to the visual cortex. Two streams (in blue) compute boundaries and surfaces. The third stream is sensitive to visual motion. The boundary stream goes from retina through the LGN parvo stage (named for its "parvocellular" cell type) to the cortical stages V1 interblob, V2 interstripe, V4, and on to inferotemporal cortex. The surface stream goes from retina through LGN parvo to V1 blob, V2 thin stripe, V4, and inferotemporal cortex.
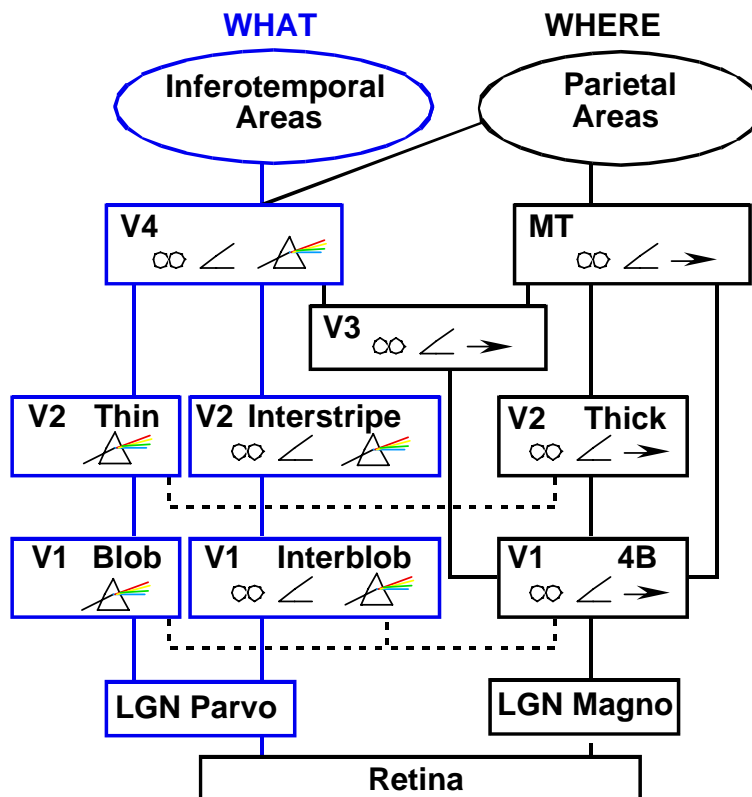
Figure 2. Three processing streams in visual cortex. Boundaries are computed through the interblobs, and surfaces through the blobs, at multiple stages until area V4. Cortical cells have different selectivities to image properties. For example, rainbow = tuned and/or opponent wavelength selectivity (incidence at least 40%), angle symbol = orientation selectivity (incidence at least 20%), spectacles = binocular disparity selectivity and/or strong binocular interactions (V2; incidence at least 20%), and right-pointing arrow = direction of motion selectivity (incidence at least 20%). [Adapted with permission from "Concurrent processing streams in monkey visual cortex," by E. A. DeYoe and D. C. van Essen, 1988, Trends in Neurosciences, II, p. 223. Copyright 1988 by Elsevier Science.]

Why are there several cortical processing streams? Theoretical and experimental evidence suggests that they compute *complementary* properties [2], or properties akin to a lock fitting its key, or puzzle pieces fitting together. Each stream exhibits complementary strengths and weaknesses. Inter-stream interactions overcome these complementary deficiencies and generate percepts of conscious form.
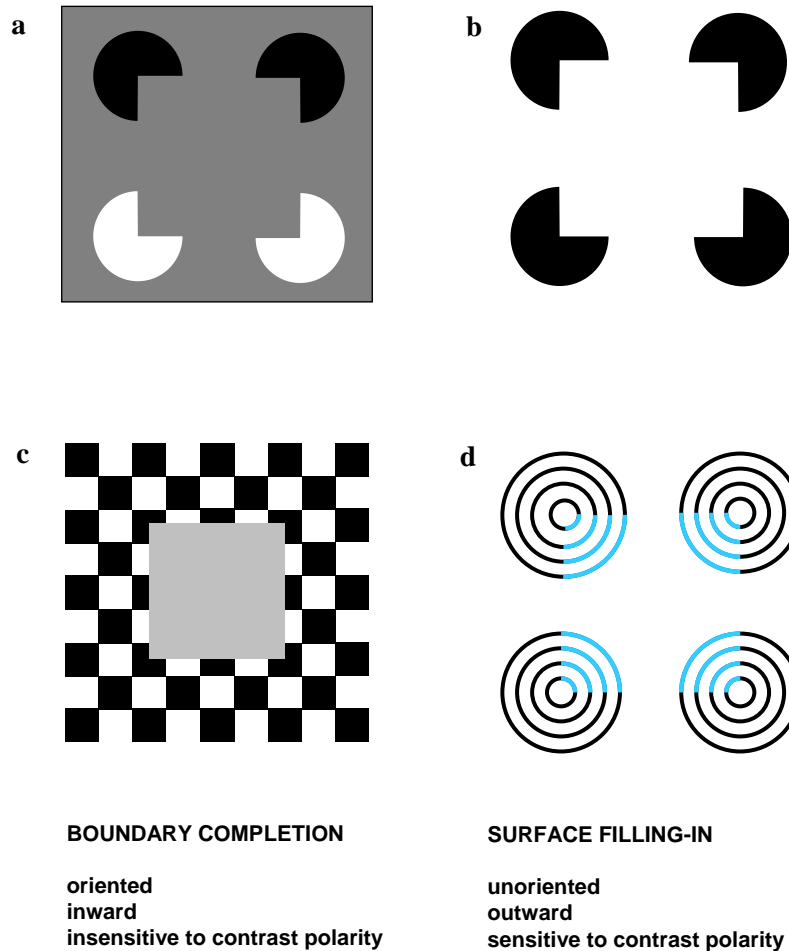
BOUNDARY COMPLETION

oriented
inward
insensitive to contrast polarity

SURFACE FILLING-IN

unoriented
outward
sensitive to contrast polarity

Figure 3. (a) opposite-contrast Kanizsa square shows that both opposite-contrast polarity and same-contrast polarity collinear edges can group together, and that both sorts of groupings are part of the same boundary completion process. Because two pac men are darker than the background gray, and the other two are lighter than the background gray, they induce lightening and darkening effects that cancel out within the Kanizsa square, thereby creating an invisible, or amodal, square percept that is recognized but not seen. (b) same-contrast Kanizsa square is visible because all four black pac men induce brightness signals within the square that create a brighter square after surface filling-in. (c) pooling of opposite contrast at every position along the square borders illustrates how the brain can build an object boundary around a textured background and thus why "all boundaries are invisible." (d) neon color spreading vividly illustrates the computationally complementary properties of boundary completion and surface filling-in that are summarized at the bottom of the figure.

What are these complementary interactions? Figures 3a and 3b illustrates three pairs of complementary properties using visual illusions. By viewing Figure 3b, our brains complete a Kanizsa square boundary even though the image contains only four black pac-man figures on a white background. In this way, form boundaries are completed over the blind spot (Figure 1). Thus many form percepts that appear to be "real" are "visual illusions". What we call a visual illusion is often just an *unfamiliar* combination of boundaries and surfaces.

Visual boundaries and surfaces obey complementary computational rules in the following sense. In response to the images in Figures 3a and 3b, boundaries form *inwardly* in an *oriented* manner between cooperating pairs of (almost) like-oriented and (almost) colinear inducers, such as the edges of the four pac man, or pie shaped, inducers.

All Boundaries Are Invisible
The square boundary in Figure 3a can be recognized even though there is no visible brightness or color difference on either side of the boundary. It is perceptually invisible, or "amodal". Figure 4a illustrates another invisible boundary that can be consciously recognized. FACADE theory [3] predicts that *all boundaries are invisible* within the interblob cortical stream (Figure 2).
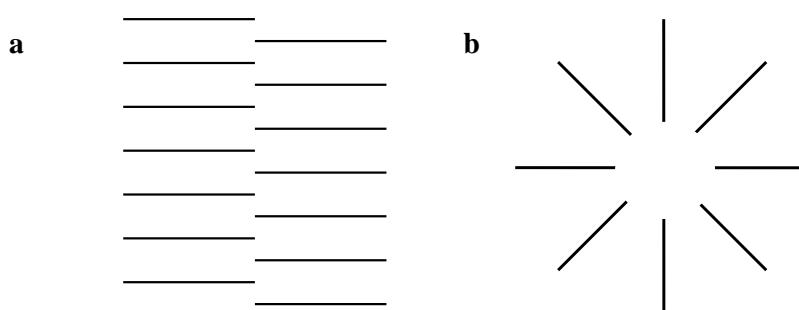
a   b

Figure 4. (a) The vertical boundary that is induced by the offset horizontal gating is invisible, or amodal. It is an invisible percept, and thus cannot be seen, but nonetheless can be consciously recognized. (b) The circular disk in the Ehrenstein illusion can be both seen and recognized because the circular boundary traps different levels of filled-in brightness inside and outside its contour.

Why are all boundaries invisible? The vertical boundaries in Figure 3a form between black and white inducers that possess opposite contrast polarity (black-to-gray or white-to-gray) with respect to the gray background. The same is true of the boundary around the gray square in Figure 3c. To build a boundary around the entire square form, despite these contrast reversals, the boundary system pools, or adds, signals from pairs of *simple cells* that are sensitive to the same orientation and position, but to opposite contrast polarities. Pooling occurs at the *complex cells* in the V1 interblob area.

By pooling light/dark and dark/light contrasts at each position, boundaries become *insensitive* to contrast polarity, so that "all boundaries are invisible". These pooled signals activate the *inward* and *oriented* boundary completion process that forms the illusory square in the V2 interstripe area [4]. Figure 3 summarizes these three properties of boundary completion.

Boundary completion helps to perceive continuous geometrical forms, discontinuous texture-defined forms, and to separate figures from their backgrounds (Figure 5). Boundaries arise as coherent patterns of excitatory and inhibitory signals across cortical feedback circuits. Classical

geometrical ideas such as points and lines are replaced by nonlinear neural networks that do on-line decision-making to select and complete the statistically most favored boundary groupings of a scene, while suppressing noise and incorrect groupings.
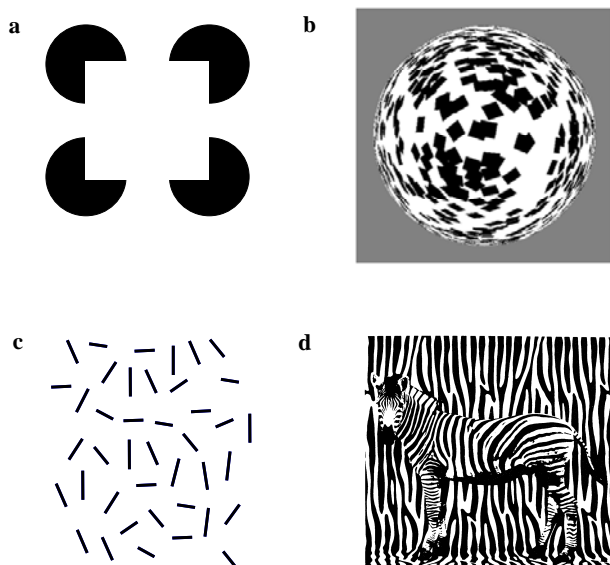

Figure 5. Examples of perceptual grouping and boundary completion: (a) an illusory square emerges from the four pac men to be seen as the well-known Kanizsa square. (b) a discrete texture in a 2D picture can generate a percept of a continuous 3D surface by differentially activating a multiple-scale plexus of form-sensitive boundaries that is called a *boundary web*. (c) colinear lines can pop-out from a field of randomly oriented lines by being linked by an emergent boundary. (d) an emergent boundary can separate the figure of the zebra from its background. T-junctions between the zebra and the background can help to push the background behind the zebra figure.

Only Some Surfaces Can Be Seen

If boundaries are invisible, then how do we see anything? An early stage of surface processing compensates for variable illumination, or *discounts the illuminant*. Otherwise, illuminant variations could seriously distort all percepts. Discounting the illuminant attenuates color and brightness signals except near regions of sufficiently rapid surface change, such as edges or texture gradients. "Feature contours" are selected at such positions and thus are relatively uncontaminated by illumination gradients.

Neural models have proposed how these feature contour signals trigger, at later processing stages, a filling-in process that completes surface representations [5]. Filling-in can allocate brightness and color to particular depths on a 3D surface via a process called *3D surface capture* (Figure 6).
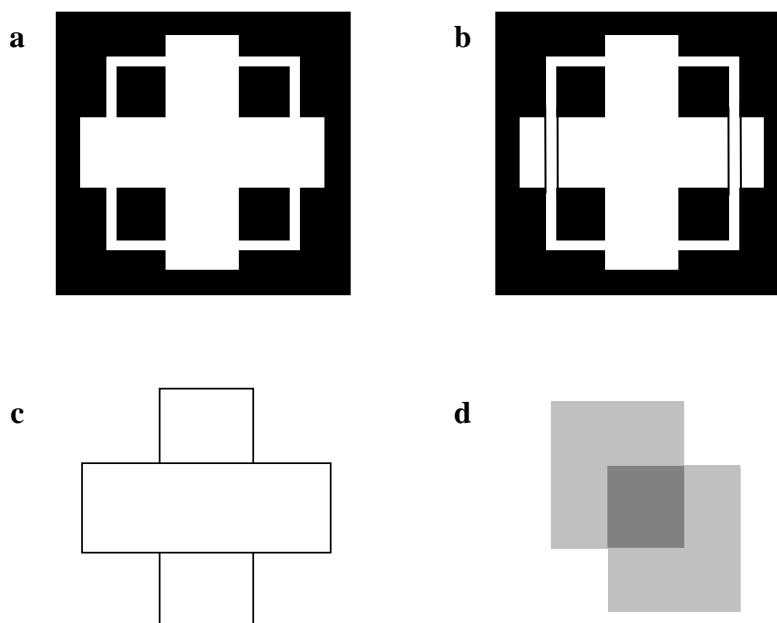
Figure 6. (a) A Kanizsa stratification figure. The percept usually looks like a white cross in front of a partially occluded white square, but can switch to look like the square in front of a partially occluded cross. The white region that is shared by the square and the cross is "captured" by the surface form that appears to be in front. (b) Bayesian theorists claim that what we see is likelihood statistics. However, this figure causes the highly unlikely percept of a cross both in front of and behind the square. (c) This image of three abutting rectangles is perceived as a horizontal bar that partially occludes a vertical bar just behind it in depth. The occluded vertical surface is invisible [3]. (d) This juxtaposition of three figures generates a bistable percept of a transparent square in front of a background square. What determines when a surface form appears opaque vs. transparent? stable or bistable [3]?

Neon color spreading [6] illustrates filling-in. In Figure 3d, boundaries of the black circular segments cause small breaks, or *end gaps*, in the boundaries of the blue circular segments. Blue color spreads through these gaps *outwardly* in an *unoriented* way until it hits a boundary or attenuates due to its spread. Filling-in can lead to visible percepts because it is *sensitive* to contrast polarity. These three properties of surface filling-in (outward, unoriented, sensitive to contrast polarity) are complementary to those of boundary completion (Figure 3).
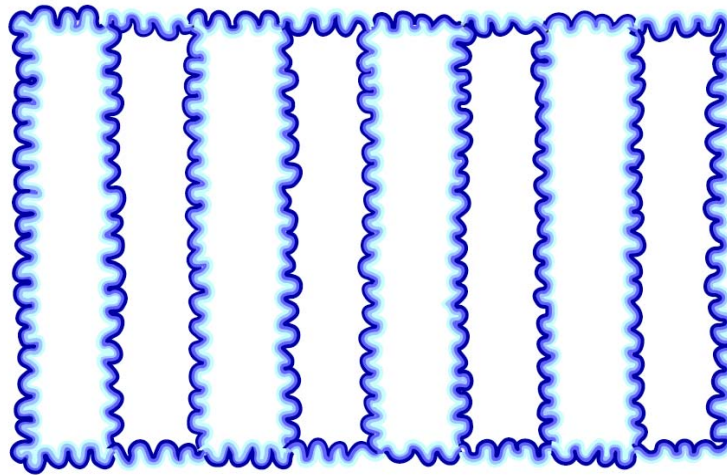
Figure 7. An example of the watercolor illusion due to Baingio Pinna. The spatial juxtaposition of the more-contrastive thin blue curves abutting less-contrastive light blue curves enables the more contrastive boundaries to inhibit the less contrastive boundaries and thus release surface filling-in of the lighter blue color.

Another striking example of filling-in is the watercolor effect. In Figure 7, the interior of the light blue regions is really white! Neon color spreading and the watercolor effect have both been explained using boundary completion and surface filling-in [7].

The mechanisms that fill in surface representations across the blind spot, can also do so after discounting the illuminant. They also clarify how we see continuous blue surfaces, such as the sky, despite the fact that the spatial distribution of blue cones on the retina is very sparse ([8]).

3D Form Vision and Figure-Ground Separation

How do two eyes work together to generate percepts of 3D form, including percepts of occluding and occluded forms in depth (Figure 5). Multiple problems need to be solved [3], including:

3D Surface Capture and Filling-in. How do multiple depth-selective boundary representations interact with multiple depth-selective surface filling-in domains to selectively *capture* brightness and color signals at prescribed depths (Figure 6)?
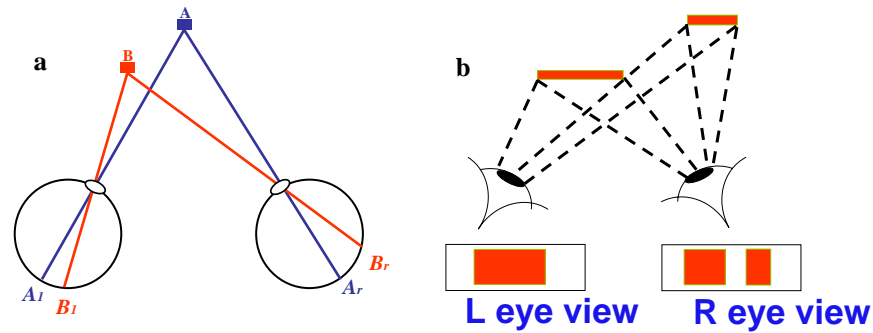
Figure 8. (a) When the two eyes foveate an object such as A, an object such as B activates the two retinas at displaced positions B_l and B_r, leading to a binocular disparity that helps us see objects in depth. (b) Sometimes only one eye can see part of an object in depth when it is occluded by another object, thereby eliminating some disparity dues to depth. This sort of daVinci stereopsis is nonetheless solved by the brain [3].

Binocular Fusion, Grouping, and daVinci Stereopsis. How are the depth-selective boundaries formed that control surface capture? Our two eyes experience slightly different views of the world that lead to relative displacements, or disparities, on their retinas of observed scenes (Figure 8). These disparate retinal images are binocularly matched and fused at disparity-sensitive simple cells in V1. Simple cell outputs combine at complex cells whose outputs to V2 activate grouping cells that form depth-selective boundaries, which capture feature contour signals at the corresponding depth-selective surface filling-in domains (Figure 9).
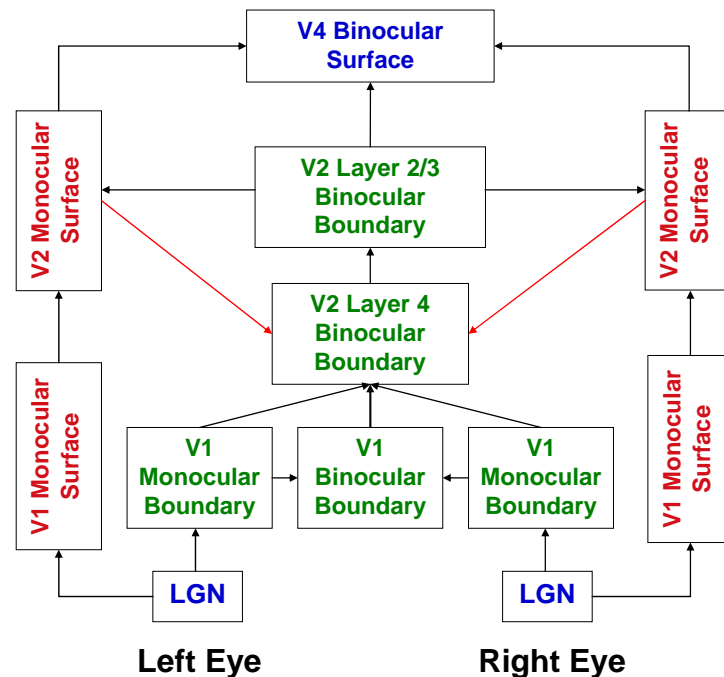

Figure 9. Macrocircuit diagram of some processing stages whereby the boundary stream (green) and the surface stream (red) interact to form a visible percept of binocular form.

Simple cells form the V1 monocular boundaries. Both simple and complex cells form V1 binocular boundaries. V2 layer 2/3 cells carry out long-range boundary completion, as in Figure 5. Their signals contain filling-in within the V2 monocular surfaces. The V2 monocular surfaces combine to form V4 binocular surfaces, constrained by boundary signals from V2 layer 2/3. These binocular surfaces are predicted to give rise to visible percepts of surface form.

Multiple Scales into Multiple Boundary Depths. When a single eye views an object in depth, the retinal image may be due to either a large object far away or to a small object nearby. How is this ambiguity overcome to activate the correct disparity-sensitive cells? The brain uses multiple receptive field sizes, or scales, that compute a "size-disparity correlation" between retinal size and binocular disparity. Each scale can fuse multiple disparities, although larger scales can fuse a wider range of disparities [9]. Multiple boundary scales react differently to different regions of texture or shading, leading to a multiple-depth form-sensitive "boundary web" of small boundary compartments. Each depth's boundary web captures different feature contours, whose filling-in can lead to a curved surface percept (Figure 5b).
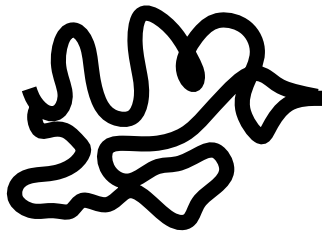

Figure 10. Both Albert Bregman and Gaetano Kanizsa discussed how an occluding form can "own" a shared boundary with an occluded form. Such border ownership influences how the occluded form may be recognized. The black occluder facilitates recognition of the partially occluded B shapes by owning the shared boundaries. This allows the remaining B boundaries to be completed at a farther depth, thereby facilitating their recognition there.

Recognizing Objects vs. Seeing Their Unoccluded Parts. In most scenes, some objects partially occlude others. How do we know which features belong to which objects? Figure 10 clarifies some issues [10]. In the lower left, gray B shapes can be recognized despite partial occlusion by the black snakelike occluder. This happens because the boundaries that are shared by the occluder and the gray shapes are assigned by the brain to the black occluder by a process of

"border ownership", leading to seeing the black occluder as closer. With the shared boundaries removed from the gray shapes, the B boundaries can be completed behind the positions of the black occluder, just like boundaries complete in response to Figures 3 and 6. In the lower right, the occluder is removed and the B shapes are harder to recognize.

From Boundary-Surface Complementarity to Consistency. Given that boundary and surface computations are complementary, how do we see a single percept? How does *complementarity* become *consistency*? Consistency can be realized by feedback that occurs between the boundary and surface streams (Figure 9) Remarkably, this feedback seems also to initiate the process of separating forms from each other during figure-ground perception [3].

References

1. Marinez-Conde S, Macknik SL, Hubel DH (2004) The role of fixational eye movements in visual perception. Nature Reviews Neuroscience 5: 229-240.

2. Grossberg S (2000) The complementary brain: Unifying brain dynamics and modularity. Trends in Cognitive Sciences 4 :233-246.

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

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

5. Pessoa L, De Weerd P (Eds) Filling-in: From perceptual completion to cortical reorganization. New York: Oxford University Press.

6. Redies C, Spillmann L (1981) The neon color effect in the Ehrenstein illusion. Perception 10: 667-681.

7. Pinna B, Grossberg S (2005) The watercolor illusion and neon color spreading: A unified analysis of new cases and neural mechanisms. Journal of the Optical Society of America A, 22, 2207-2221.

8. Boynton RM, Eskew RT Jr, Olson CX (1985) Blue cones contribute to border distinctness. Vision Research 25: 1349-1352.

9. Julesz B, Schumer RA (1981) Early visual perception. Annual Review of Psychology 32: 1013-1022.

10. Nakayama K, Shimojo S, Silverman GH (1989) Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. Perception 18: 55-68.

Glossary Entries

Texture segregation
A visual texture refers to the size, shape, and arrangement of different elements of a scene or picture. Texture segregation refers to the brain's ability to group into distinct regions and objects different parts of a scene that have different texture statistics. Boundary and surface processes in the brain can use texture, as well as edge, shading, depth, and color, information to generate percepts of visual form. Figures 5b-d in the essay on Form Perception provide examples of texture segregation.

Texture processing
Texture processing refers to the brain's ability to respond to the local statistics of textured regions in a scene or picture. Texture processing includes texture segregation, and can generate recognizable percepts of object form.

Shape processing
Shape refers to a distinctive combination of boundary and surface properties of a visual form. Shape processing refers to the brain's ability to combine many types of locally ambiguous visual signals, such as edges, texture, shading, depth, and color, to generate an emergent representation of an object's shape.

Visual memory
Visual memory refers to the ability to recall pictures, scenes, words, or other information that is presented visually. Visual memories can last for a short amount of time, as in visual iconic memories or afterimages, or for many years, as in the type of learned long-term memory that enables us to visually imagine and recognize events that have occurred years or even decades ago.