

# F

## Figure-Ground Segregation, Computational Neural Models of



Arash Yazdanbakhsh<sup>1</sup> and Ennio Mingolla<sup>2</sup>

<sup>1</sup>Computational Neuroscience and Vision Lab, Department of Psychological and Brain Sciences, Graduate Program for Neuroscience (GPN), Center for Systems Neuroscience (CSN), Center for Research in Sensory Communications and Neural Technology (CReSCNT), Boston University, Boston, MA, USA

<sup>2</sup>Department of Communication Sciences and Disorders, Bouvé College of Health Sciences, Northeastern University, Boston, MA, USA

### Definition

Figure-ground segregation refers to the capacity of a visual system to rapidly and reliably pick out for greater visual analysis, attention, or awareness, or preparation for motor action, a region of the visual field (figure) that is distinct from the combined areas of all the rest of the visual field (ground). The “figure” region is often, but not necessarily, bounded by a single closed visual contour, and the figure region is often said to “own” the boundary between it and any adjacent regions. The figure region is generally experienced as “in front of” (along lines of sight) surfaces of objects that are in the ground region.

Once figure-ground segmentation is achieved, the figure region often delineates a zone of additional attentive visual processing.

### Detailed Description

#### The Importance of Figure-Ground Segregation

Visual function in animals has broadly increased in complexity and competence across eons of evolution, with figure-ground perception being among the later and more intriguing achievements of relatively sophisticated visual species. The importance of figure-ground perception can be seen by considering its advantages to animals that have the capability, as compared to the visual limitations of animals whose visual systems do not support figure-ground perception (Land and Nilsson 2002). For example, the single-celled euglena can use its eyespot and flagellum to orient its locomotion in a light field, but it cannot take account of the edges of individual objects in the sense of figure-ground for purposes of steering. Arrays of visual receptors that sample different directions of the visual field, such as the compound eyes of insects or mammalian retinas, are needed for computing figure-ground relations. Successful figure-ground segregation can facilitate several subsequent visual tasks, such as object recognition or guidance of locomotion toward a goal or around an obstacle.

Some species of animals can perform figure-ground perception only in the presence of visual motion, while others can also do so in static environments. It must be noted that even when the environment outside an animal is perfectly still, the animal itself may be making voluntary movements at a macroscale or involuntary movements (eye tremor or microsaccades) that can help to support figure-ground segregation.

Finally, as we will see in the case of human (or, more generally, primate) figure-ground perception, the space-variant arrangement of receptors on the retina and the associated “cortical magnification” that concentrates the brain’s processing resource to the central region of the visual field are major factors both in achieving and in exploiting figure-ground perception. Voluntary eye movements to an approximate centroid of a relatively homogenous visual region can enhance the visual system’s capacity to delineate an associated figure-ground boundary and thus to perform figure-ground segregation at all. Conversely, as long as the eyes are foveating a “figure” region, a high proportion of cortical tissues is devoted to analysis of that region, courtesy of cortical magnification, thus facilitating object recognition or discrimination between two similar object categories. The ability to deploy foveal vision while performing figure-ground segregation can thus be viewed as a “force multiplier” that enables detailed scrutiny at relatively low metabolic cost (eye movements), rather than requiring a reorientation of an entire head or actual approach of the entire body toward an environmental region.

### Elementary Cues for Figure-Ground Segregation

In visual scenes with little structure, separation of figure from ground may appear as an easy or even trivial task. For example, a small dark spot on an otherwise homogenous light background is often said to “pop out” from its background in the context of a visual search task (e.g., “Find the dark spot,” Treisman and Gelade 1980). Such isolated, salient figural properties “call attention to themselves” and are effective as lures that attract saccades. A consensus of recent decades holds that

regions of higher-than-average (relative to the rest of the scene) activation of certain featural qualities (e.g., of color, contrast, motion, or orientation) in visually topographic “feature maps” make such regions likely targets for attention or of overt eye movements (Wolfe and Horowitz 2017).

An important class of figure-ground phenomena occurs in cluttered visual scenes, where the achievement of figure-ground segregation can amount to the breaking of camouflage. Gibson et al. (1969) demonstrated vivid figure-ground segregation in random dot kinematograms – displays involving some combination of static and moving randomly distributed dots that are analogous in the time/motion domain to random dot stereograms. In random dot stereograms, two eyes receive related, but different, images, typically where a region of dots (the figure) is displaced in one eye’s view relative to the other eye’s view, causing a relative disparity of stimulated retinal locations that is different from the disparities of other corresponding dots in the two eyes (which form the ground).

The importance of motion to figure-ground perception of ecological scenes has been evident since, at least, the seminal paper by Lettvin et al. (1959) “What the frog’s eye tells the frog’s brain,” whose authors noted that frogs of the studied species would starve in the presence of freshly killed flies spread on the ground before them. Nonetheless, humans and certain other species have the capacity to perform figure-ground segregation for many complex static scenes, including pictures – for which endogenous eye movements convey no additional information, as would be the case for views of actual three-dimensional environments.

### The Scope of the Present Article

This article focuses on computational models of human or primate figure-ground segregation. We indicate whether models are intended to address moving or static scenes, or both, and whether they rely on binocular vision. Reasons for this focus include (1) that primate visual systems have the greatest complexity and competency among known biological systems and (2) that many machine vision systems are

designed with the intent of matching if not exceeding human performance in important visual tasks. While many computer vision algorithms are designed without reference to animal examples, it makes sense to study natural visual systems for inspiration in those domains for which machine performance continues to lag, as is the case with figure-ground segregation in cluttered scenes.

### What Is Border-Ownership and How Is It Related to Figure-Ground Perception?

A “figure” object and its borders occlude the objects and their borders of the background. Figure 1a shows the famous painting by Leonardo da Vinci: Mona Lisa occludes the trees, river, bushes, and other objects in the background.

The foreground borders in (Fig. 1a) obviously “belong” to the object (i.e., Mona Lisa) in front, which is the figure. The shape of the borders is often informative about the identity of the object (a human) that is depicted, which is said to “own” the border – thus figure-ground perception can be an important step on the way to object recognition. Note however that Evans and Treisman (2005) have shown that people can perform animal detection in natural scenes with great success even though they are unable to localize the animal, so figure-ground perception cannot be said to be a required step in object recognition. Rather, the *localization* of a figure’s borders and proper assignment of the ownership of associated borders are logical requirements of figure-ground segregation.

### Border-Ownership Is Independent from the Polarity of Contrast Across Borders

A clever artist can devise displays in which the ownership of a border is ambiguous, leading to bistable percepts. See Fig. 1b, the Rubin (doctoral dissertation 1915) faces/vase. While consciously experiencing the faces against a black background, the borders separating the white and black areas are the face borders. Note that flipping the contrast of Fig. 1b and making the faces black and the vase white (Fig. 1c) do not change the relationship between the figure-ground separation and border-ownership, i.e., when the

faces are seen against the uniform background, the borders between the black and white areas belong to the faces and the same for the vase.

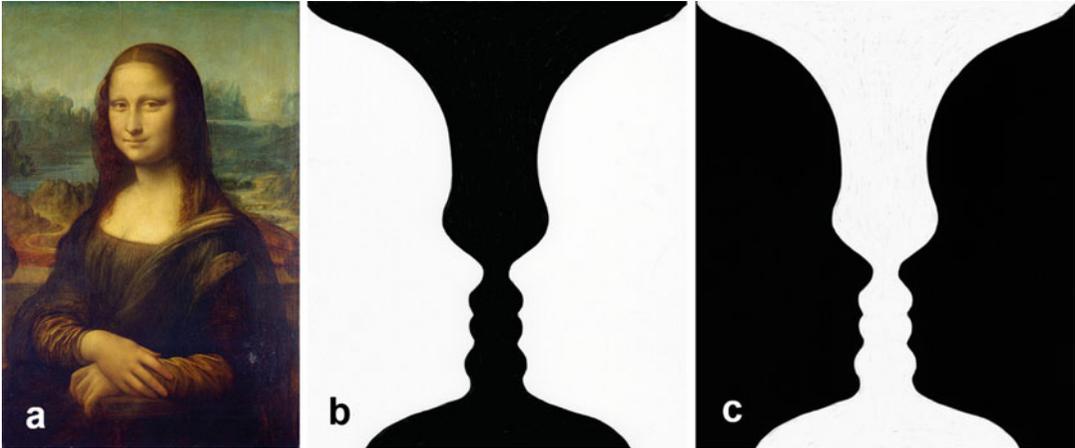
A good deal of attention in introductory textbooks is devoted to bistable displays (Fig. 1b–c) when discussing figure-ground perception. Although such displays help to make an important point about the relationship between figure-ground relations and border-ownership, they are a potentially misleading starting point relative to figure-ground perception in natural scenes. In bistable displays, volitional manipulation by the viewer of semantic categories such as face or vase can “drive” a resulting percept. The bistability of labeling border-ownership is not typical of most normal perceptual settings, which are closer to the depiction of Fig. 1a. Note also that finding boundaries is trivial in binary-valued displays with solid, connected regions, such as Fig. 1b and c, as opposed to conditions of, for example, animal camouflage (Fig. 2).

### Psychophysical Antecedents of Recent Computational Neuroscience of Figure-Ground Perception

The second half of the twentieth century saw a great deal of human psychophysics devoted to the study of texture segregation, visual segmentation and grouping, and visual search.

Linking these studies – many of which sought to find fundamental visual “elements” of perception – was the goal of characterizing how, out of the “blooming, buzzing, confusion” of time-varying visual scenes, our visual systems are able to focus on relatively coherent objects of interest (figures). The texture segmentation and grouping studies characterized conditions under which perceptually continuously connected boundaries between regions could form, even in the absence of continuous differences in luminance or contrast along those boundaries (such as is evident in Fig. 2). Thus, the question of “ownership” of boundaries was often not raised.

Visual search studies, pioneered by Anne Treisman in the 1980s (Treisman and Gelade 1980) and developed since by many, most notably Jeremy Wolfe (Wolfe and Horowitz 2017),



**Figure-Ground Segregation, Computational Neural Models of, Fig. 1** Examples of 2D images helping formulate the concept of figure-ground segregation and border ownership, irrespective of some visual attributes such as edge contrast. (a) Mona Lisa by Leonardo da Vinci.

(Adapted from [https://en.wikipedia.org/wiki/Mona\\_Lisa](https://en.wikipedia.org/wiki/Mona_Lisa) under the public domain), (b) Rubin faces/vase. (Taken from [https://commons.wikimedia.org/wiki/File:Multi\\_stability.svg](https://commons.wikimedia.org/wiki/File:Multi_stability.svg) and is released under the public domain by Alan De Smet), and (c) its contrast inverted version



**Figure-Ground Segregation, Computational Neural Models of, Fig. 2** Figure-ground segregation during camouflage, when the figure and background have similar textures and their average luminance are close to each other. Such scenes can pose another level of challenge to computational models of figure-ground segregation, as the

figure borders cannot be readily detected by the lining up of the units detecting edge contrast, as tiny edge contrasts with different orientations are almost uniformly distributed. (Adapted from pixabay.com; photo is released under CC0 Creative Commons)

focused on describing conditions under which a region of a scene would “pop out” and thus be treated as a “figure” for subsequent processing, with the rest of the display acting as “ground.” When conditions for pop-out were not in play in

a given display, search could be “guided” by top-down expectations (Wolfe and Horowitz 2017), or a number of candidate regions (Grossberg et al. 1994) might be evaluated in order of likelihood of containing the object being searched for. In scenes

containing an object of search, the finding of that object, whether quickly or slowly, inevitably is accompanied by the phenomenal experience of the visible regions of that object as “figure.” A series of psychophysical studies led in turn to related paradigms involving brain imaging or electrophysiological recordings in primates, which have advanced our understanding of figure-ground perception tremendously.

Finally, an important thing to note about figure-ground perception is that, unlike the examples discussed to this point, in daily life it does not generally occur in a flat plane, such as in a picture. When we look at a spoon or cup that we are about to reach for, the surface of the object is extended and often curved in three dimensions. Tyler and Kontsevich (1995) introduced the important concept of an “attentional shroud” – a fuzzy manifold in perceptual space that envelopes the surfaces of attended objects. The shroud delineates both the limits of focal attention and also the tendency of attention to involuntarily “spread” along the visible surfaces of an attended figure, rather than into the rest of the scene (the ground). Indeed, the delineation of attentional resources for visual processing after the determination of figure-ground segregation can be viewed as an important functional reason for engaging in figure-ground segregation in the first place.

### Brain Areas Involved in Figure-Ground Perception

Although figure-ground segregation is fundamental to visual perception, how the visual system performs it is not well understood. Coding of border-ownership starts early in primate brain visual stream. A direct link between visual figure-ground perception and the responses of certain single neurons has, however, been established in the early visual system.

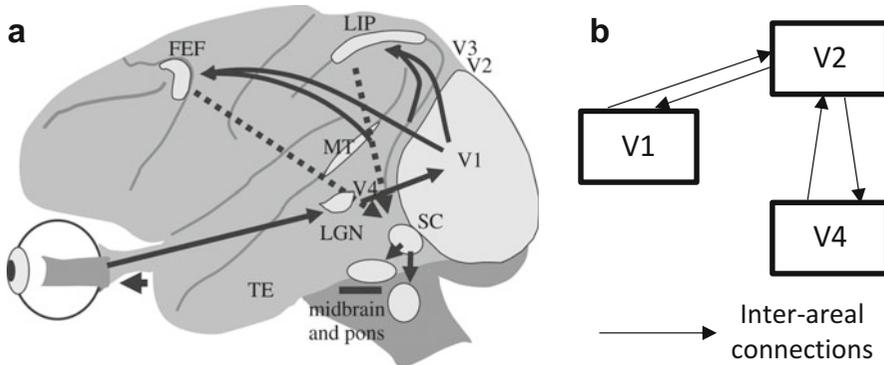
These cell responses may require the simultaneous activation of parts of visual areas V1, V2, and V4 (Fig. 3) acting as a functional network (Layton et al. 2012), whose neurons are monosynaptically connected to each other across visual areas (Fig. 3b). Such connections are termed *inter-areal*. On the other hand, neural connections within each area (i.e., V1, V2, etc.) are called

*intra-areal*. Zhou et al. (2000) and Ko and von der Heydt (2018) have found that about half of sampled cells from primate visual areas V2 and V4 preferentially respond to borders whose ownership is related to figure-ground relations (Fig. 4). These cells are referred to as border or B-cells and have “side-of-figure” selectivity that indicates a neurophysiological correlate of percepts observed in Fig. 1b–c, namely, that a border is owned by either the region to one side (for example vase) or another of that border (face), but not both.

### Temporal Dynamics of Border-Ownership Neural Response

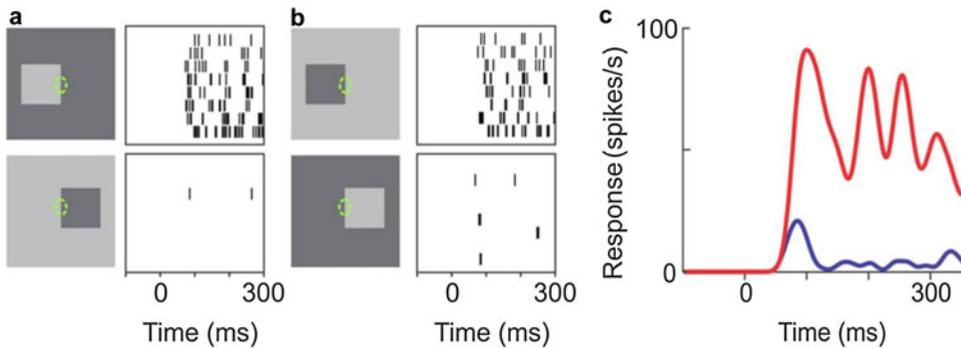
The border-ownership-related neural response can emerge within 25 msec after the B-neuron’s response onset. In V2 and V4, the differences were found to be nearly independent of figure size up to the limit set by the size of the display (21°) that Zhou et al. (2000) used (Fig. 4). Because such a response depends on the processing of an image region that is at least as large as the figure and thus requires the transmission of information over some distance in the cortex, *these short delays constrain the functional network underlying the B-cell response and the biological plausibility of neural models*.

The short delay of the emergence of border-ownership signals constrains the contribution of inter-areal and intra-areal connections in a biologically plausible neural model that can perform figure-ground segregation (Layton et al. 2012). This is because inter-areal connections are myelinated and transfer neural signals much faster than the intra-areal ones, and therefore, the small delays observed in border-ownership signals demand fast inter-areal connections. Researchers have proposed that B cells access global information either intra-areally, i.e., by lateral connections within a single visual cortical area, such as V2 (Zhaoping 2005), or inter-areally, i.e., where cells with larger receptive fields (e.g., V2) communicate contextual information about the scene via feedback projections to visual areas whose cells have smaller receptive fields (e.g., V1; see the diagram in Fig. 3b). Intra-areal and inter-areal axonal conduction velocities have been estimated



**Figure-Ground Segregation, Computational Neural Models of, Fig. 3** Brain areas at some distance can be axonally connected, such as inter-areal connections between visual areas V1, V2, and V4. Inter-areal connections via myelinated axons are much faster than within area (intra-areal, unmyelinated) connections. As a result, during a stream of visual inputs, multiple brain areas can be activated within a short period of time, forming

a multiple-area network for processing the input. (a) Lateral view of the left hemisphere of a monkey brain, roughly showing the locations of V1, V2, and V4. (Adapted from Wurtz (2015) under the Creative Commons Attribution 4.0 International license). (b) A simple schema showing inter-areal and mutual connections between areas V1, V2, and V4, highlighting the fast inter-areal connections that can generate a functional multiple-area visual network



**Figure-Ground Segregation, Computational Neural Models of, Fig. 4** The von der Heydt group discovered the presence of border-ownership neurons. The neuron receptive field is shown by small circle/ellipse which is over the figure edge. Local contrast (light/dark) is the same across each column and over the classic receptive field of

the neuron, yet, the neuron shows preference for figure to the left (1st row in a and b) than to the right (2nd row in a and b). Such cells, therefore, encode border-ownership, and in this particular case, left-side ownership. (Adapted from von der Heydt (2015) under CC BY License)

to be 0.3 and 3.5 m/s in early visual areas, respectively (Bullier 2001). Hence, inter-areal connections can be an order of magnitude faster than intra-areal connections for propagating information across the visual field. B-cell responses to 3° squares did not differ in latency compared to those to an 8° square, which is consistent with the use of inter-areal connections, but not intra-areal connections, to propagate contextual figure-ground information. Although a variable amount of time is required to propagate information about a figure

within a single cortical area, transmitting the information to another area with large receptive field cells could afford a roughly fixed delay irrespective of the figure size in the visual field (Layton et al. 2012). Hence, *it appears that connections within a single cortical area alone could not plausibly account for the fast global scene integration that is observed in B-cell border-ownership responses* (Layton et al. 2012), but see Zhaoping (2005), who argues otherwise.

The immediately preceding discussion is an example of an emerging paradigm shift in understanding of cortical visual processes. Most physiological studies related to cell response to luminance-based and hue-based edges of the primate visual system in the past half-century have followed the path established by Hubel and Wiesel (1962) and largely focused on the function of individual areas or subpopulations of cells *within a visual area*. Topics of such studies include how different cells in the primary visual cortex respond to luminance base edges, to the light-dark polarity of contrast, and to variations of edge orientation. It instead appears that the cortex can solve complex problems *with networks that span multiple areas*. The visual system may thus rapidly recruit an assembly of cortical areas to determine border-ownership in figure-ground segregation, a single emergent function. Neuroanatomical evidence indicates that early visual areas such as LGN, V1, V2, and V4 are massively interconnected with numerous feedforward and feedback connections (Sincich and Horton 2005, see Fig. 3b). Feedforward connections are believed to quickly propagate sensory visual information to cortical areas further up the visual hierarchy to serve a rich perception of the visual scene. Feedback projections are often said to play a modulatory role with respect to bottom-up sensory visual signals by increasing the gain of neuronal responses in attended regions and performing contextual integration. To date, few studies have hypothesized that feedback projections play a crucial as opposed to supplementary roles for the functions of early visual cortices. It appears that the simultaneous activation of multiple areas early in the visual system not only performs modular functions that are later combined but also that such activation can collectively solve problems that individual cortical areas cannot solve alone (Layton et al. 2012).

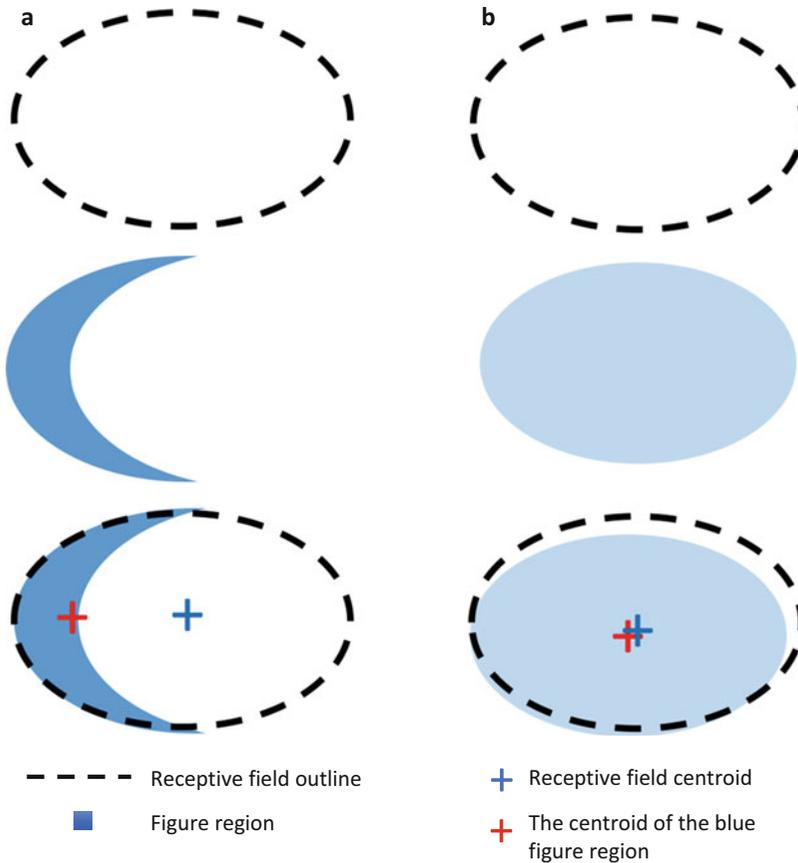
### **A Teardrop Model of Figure-Ground Segregation and an Edge-Integration Model**

Layton et al. (2014) presented a model of figure-ground segregation in which inter-areal feedback plays a crucial role in disambiguating a figure's interior and exterior. The model processes are

based on primate receptive field scatter (jitter in receptive field center locations) and variation in receptive field sizes to generate a code for inside versus outside of a visible figure. Feedforward projections from V4 signal the curvature of object boundaries (curved contour cells), and feedback projections from visual temporal areas group neurons with different receptive field sizes and receptive field center locations in a teardrop geometry. Neurons sensitive to convex contours, which respond more when centered on a figure, balance feedforward information and feedback from higher visual areas. The model produces maximum activity along the medial axis of figures irrespective of the concavities and convexities of figure contour. The model feedback mechanism balanced with feedforward signals is crucial for figure-ground segregation and localization (Fig. 5).

Another computational approach in solving the figure-ground problem for static scenes is given by Grossberg (1993, 1994). This neural model involves feedforward and feedback interactions between cortical units that represent boundaries and surfaces at multiple spatial scales. The assignment of border-ownership drives interactions where signals for boundaries that are nearer to the viewer suppress signals for further-away boundaries in a way that supports featural visibility (lightness) of the figure region while allowing aligned segments of far boundaries to complete and support amodal completion of partially occluded surfaces. Note that, while differing in important ways from the Layton et al. (2012, 2014) models, the Grossberg (1993, 1994) approach introduced the idea of cooperation and competition among representations in multiple cortical areas as a key requirement of figure-ground segregation.

A contrasting approach to modeling figure-ground segregation is taken by Kogo et al. (2010), who view figure-ground perception as a nonlinear differentiation/integration process that simultaneously accounts for border-ownership, lightness, and depth perception. A crucial step in this model is called "Integration for Surface Completion," where estimates of depth in a featurally homogeneous region's interior are



**Figure-Ground Segregation, Computational Neural Models of, Fig. 5** Localizing a C-shaped (crescent moon shaped) figure is difficult from the perspective of a single neuron's convex receptive field (shown as dotted black outlines in the top and bottom rows of the figure). The middle row shows two stimulus shapes – a concave C shape in column (a) and an ellipse in column (b). Note that the shape in (a) is darker and has a greater contrast with the white background, leading to higher activation per unit area of a neuron's receptive field. Thus the degree of

activation of the two neurons in the third row can be approximately the same. Localization is a different matter, however. The figure centroid is shown by two red cross symbols, +, whereas the centers of the two neurons' receptive fields are indicated by blue crosses. Localization is poor in case (a). The teardrop model of Layton et al. (2014) offers an inter-areal network to resolve figure localization in cases like in (a). Note that the areas chosen for investigation in the study described in Fig. 4 are also "C-shaped," where the C is a block shape with right-angle corners

"filled-in" through values that are computationally propagated from signals from the region's edges. This integration step differs with the Layton et al. 2014 approach, which depends on estimates of depth from multiple receptive fields in a higher visual area whose centroids are distributed ("jittered") across the region's area. Integration is a computation that would seem to "naturally" occur within a visual area, and further investigation of comparisons between classes of models employing integration versus inter-areal

feedback is needed to better understand figure-ground segregation.

### Neural Responses to Texture-Defined Static Figures

Lamme (1995) is an important monkey single-unit recording study involving displays in which a figure-ground segregation is easily perceived (by humans) through differences in the orientation of texture elements. Although the receptive field of certain neurons in early visual areas of monkey

lies entirely within the “figure” texture region, the neurons exhibit an enhanced firing rate compared to when the monkeys were presented a display containing a uniform texture of the neuron’s preferred orientation throughout the display.

The interior enhancement effect persists when the edges of the square are 8–10° and the modulation occurs after an 80–100 ms latency from the onset of the stimulus, which suggests *feedback from neurons with larger RFs* may be involved. A temporal analysis indicates that neural activity related to the edges of the figure emerges first, following a short latency, then interior enhancement occurs in the “late component” of the response (Roelfsema et al. 2002). A neuron that shows interior enhancement continues to fire at an elevated rate when the RF is centered at different positions within the texture-defined figure, and the firing rate drops when the RF is centered on the background. Neurons in V2 demonstrate a greater degree of interior enhancement compared to those in V1, and the magnitude of interior enhancement response is greatest in V4.

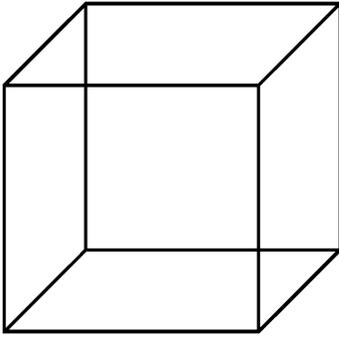
### Motion-Based Figure-Ground Segregation

Fishes, frogs, moths, and snakes demonstrate adaptations in the appearance of their body to hide from predators. Through camouflage, the markings on the prey’s body cause the animal to be grouped with, rather than stand out from the surrounding. This strategy is effective as long as the animal *does not move*, because many predators and humans cannot easily detect stationary objects that resemble their surroundings in texture, color, and luminance. However, when a previously invisible animal breaks camouflage by sudden motion, humans rapidly perceive a figure at a different depth from the surroundings, even if the texture is statistically identical. When a figure moves in front of a similarly textured background, it is said to produce kinetic occlusion (Gibson et al. 1969; Kaplan 1969). No reliable luminance contrast exists between the figure and background; there is only the relative motion between the texture patterns separated by a kinetically defined edge (kinetic edge). How do humans perceive the figure at a different depth than the background (figure-ground segregation),

despite the figure and background possessing patterns with statistically identical luminance? Obviously, this is a demanding scene processing task for the primate visual system, which goes beyond simple contrast processing and is likely to involve brain areas involved in motion processing (the middle temporal area, MT, among others) as well as form processing (including V4 and probably other visual areas) to extract the moving form from visual motion cues scattered within the camouflaged scene. In this regard Layton and Yazdanbakhsh (2015) have suggested an interconnected network model of areas V1-V2-V4-MT, in which MT and V4, both monosynaptically connected to V2, process the motion and form signals, respectively. Another possible approach to motion-based figure-ground segregation is described by Barnes and Mingolla (2013), whereby motion onset and offset signals generated by accretion or deletion of texture elements via occlusion act as key gating signals for figure-ground segregation. Note that it is quite possible that the primate visual system has evolved multiple mechanisms, with partially overlapping competencies, to address the important needs of figure-ground segregation.

### Need for Feedback Among Brain Areas

As noted in “the section [Temporal Dynamics of Border Ownership Neural Response](#)”, the primate visual system can rapidly recruit several cortical areas in the service of determining figure-ground relations. One way to investigate the mechanisms underlying the context-sensitive aspects of perceptual organization is the use of images that create bistable perception in which the perceptual interpretations of the image keep alternating over time, while the physical input image is kept constant. One of the most famous images to induce perceptual bistability involving figure-ground organization is Rubin’s “face-or-vase” image (Fig. 1b–c): The perceptual interpretation keeps alternating between “two faces” on the sides and a “vase” in the center, while the competing area is perceived as a part of the background. Because the perceptual switch in this case is specifically linked to the reversal of the figure-ground organization, we can call this phenomenon bistable figure-



**Figure-Ground Segregation, Computational Neural Models of, Fig. 6** The Necker cube is a classic example of bistable figures. Sometimes the lower square appears in front and sometimes in the back. The switch is instantaneous, showing our visual brain exhausts possible 3D alternatives which result in the same 2D projection

ground organization. Rubin presented his dissertation in 1915, with one of the key points that when two neighboring fields share a border, one will be seen as figure against the other as the background and that the percept of a figure is dominated by one of the fields against the other sharing that border. Pitts et al. (2011) used a Rubin faces/vase display to investigate the dynamics of figure-ground processing via EEG with humans and found that reorganization (faces to vase or back) could occur on a timescale of 200 msec or less, which allows for signals to propagate both in feedforward and feedback directions across visual areas.

Another example of bistable images is the Necker cube (Fig. 6), in which the front side of the cube swings between the top and bottom squares. Similar to Rubin vase, such a perceptual bistability can well be temporally modulated by attention, i.e., with a bit of practice, one can “will” the top or bottom square closer and vice versa at a given time. Similarly, in Rubin faces/vase case, one can make the face or vase the figure by attention at a time. Such demonstrations further support the claim that figure-ground perception is a complex process that involves both top-down and bottom-up interactions in the primate brain. The degree of facility of such voluntary attentional switches has been exploited as potential markers for progression of Parkinson’s disease (Diaz et al. 2015a, b).

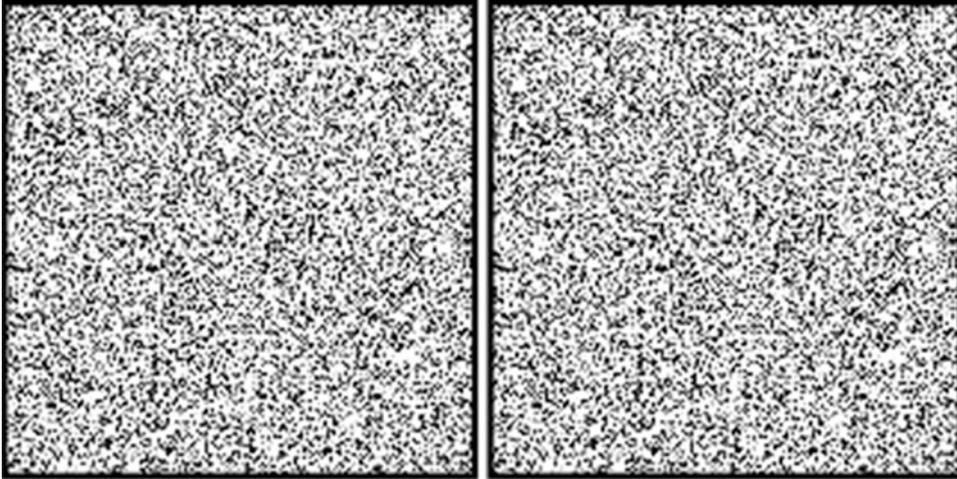
### “2D” Versus “3D” Displays: Figure-Ground Segregation and Stereopsis

Depth and separating a figure from background are tightly linked. A figure is generally closer in depth than the background. This brings up the question whether a figure can be purely defined and segregated from a ground area by binocular depth, with no reliance on contrast differences along an edge. Béla Julesz in 1959 generated pairs of images with random dots for which, if were shown to each eye separately, a portion of dots (e.g., forming a square) was seen at a different depth from the rest (Julesz 1971). Note that figure-ground segregation is thereby achieved, with the resulting stereoscopic border being owned by the nearer figure region. Closing each of the eyes (or just looking at each half pair instead) vanishes the percept of the circle as a figure against its background (Fig. 7). This is an example of pure stereopsis involvement (and nothing more) in forming a figure against its background. Here again, similar to the animal in camouflage case, there is no simple contrast between the figure and its background, and the figure and its borders are depth based. Are there single-cell data available for such figure-ground segregation to be used for a neural modeling approach?

Von der Heydt et al. (2000) investigated the representation of stereoscopic surfaces and edges in monkey visual cortex and whether such neural responses carry information of the figure borders and surfaces. Two key points are discovered in their work: first, if a figure like a square is generated by random dot stereogram in which there is no first-order (contrast-based) edge, the V1 cells respond to the stereoscopically defined figure *surface* rather than to its edges; second, V2 cells respond to the stereoscopically defined edges of the figure rather than to its surface.

### Conclusion

For most naturally occurring scenes, figure-ground perception is a quick, automatic, seemingly effortless, and successful process, with certain naturally occurring static images or humanly engineered camouflage being notable exceptions.



**Figure-Ground Segregation, Computational Neural Models of, Fig. 7** Random Dot Stereogram. Looking at each half-pair doesn't result in seeing a particular pattern, yet, if the left and right eyes see only the left and right half-pairs, respectively, a circle in the middle appears farther. This example shows that while there is no visual border in

each half-pair, additional cues can create a pattern with clear borders. Additional cues can be stereopsis (here), motion, or large scale grouping of figure elements. (Figure is adapted from [https://commons.wikimedia.org/wiki/File:Mh\\_stereogramm\\_randomdot.png](https://commons.wikimedia.org/wiki/File:Mh_stereogramm_randomdot.png) under Creative Commons Attribution-Share Alike 3.0 Unported)

Even in cases of successful camouflage, however, our visual experience is often one of a figure-ground segregation – just not the segregation that camouflage is designed to hide. Indeed, it is difficult to engineer a visual display that does not support figure-ground organization, i.e., a *ganzfeld* – a completely homogeneous light field – or very shallow gradients of luminance throughout the visual field. In contrast, if existing neural models capable of figure-ground segregation were to be challenged with a wide gamut of natural scenes, all of today's models would suffer multiple and systematic failures, because existing neural models have been designed for a very small subset of such scenes. The diversity of the cortical networks and associated computational strategies that of our visual system uses to perform figure-ground segregation is partially revealed by variety of visual illusions, but the known class of figure-ground illusions is a particularly small and non-representative subset of natural scenes. Given that our visual system performs figure-ground segregation with its all-neural machinery capable of doing many other visual processes, understanding the underlying neural processing underlying figure-ground segregation can be indeed a difficult undertaking. It is only in recent years that

“border-ownership” has been considered a “feature” to the human visual system and hence worth of study through single-unit neural recording or brain imaging. Much remains to be done.

One of the characteristics of visual neurons is their receptive fields, which are derived from the rich connectivity between neurons in different visual areas. We have not reached a sufficiently detailed understanding of the spatial and temporal characterization of neuronal networks that perform the sophisticated task of figure-ground segregation on the order of a few tens of milliseconds. Progress in neuroanatomy and connectomics increasingly is revealing the geometry of available connections in brain networks, but a functional understanding of physiology remains constrained by several factors. One is the inherent spatial and temporal limits of each recording modality, be it single or multiunit recording, EEG, MEG, fMRI, or other modality. Another factor is that the structures of current computational neural models are far simpler than their biological counterparts, as is a virtual requirement of any progressive attempts at model-building. A satisfactory level of understanding of figure-ground segregation awaits further integrated cycles of experimentation and model development.

Before closing, we note that this article has not touched on two topics that a reader might wonder about. The first is deep learning or deep convolutional networks. Recent successes of this class of modeling architectures in object recognition are impressive in their accuracy, outperforming humans in many tasks for which a sufficient database of accurately labeled exemplars has been compiled to train the artificial neural network. Such networks have also enjoyed increasing successes in boundary segmentation and semantic segmentation, in the latter case assigning each pixel in an image the name of a category (e.g., “grass,” “road,” “sky,” etc.) that most likely is its source in an imaged scene.

Accurately finding all boundaries in a scene or semantically labeling all pixels in an image is, however, almost the exact opposite of figure-ground segregation. Animals have not evolved with the luxury of virtually limitless computational power or billions of computing cycles to before interacting with objects in their world. As noted at the start of this article, the significance of human figure-ground segregation is intimately tied to the machinery of primate vision – trying to balance the fine resolution of the fovea with a wide field of view, creating the tradeoffs of eye movements and cortical magnification. Our default mode of visual scanning is to focus on one object at a time for a few hundred milliseconds, taking in “at a glance,” or perhaps with a handful of saccades, the important characteristics of each “figure” that we scrutinize in turn. It is thus not reasonable to expect computational approaches that treat large images as translationally uniform arrays of pixels to be informative about human figure-ground segregation. With this said, any machine vision algorithm that is designed to specifically segment an image into two regions, with one completely surrounded by another, can be said to be addressing the figure-ground problem in some way. A particular relevant example is that of Sarti et al. (2000) which uses a “seed” location (analogous to a point of foveation) to drive completion of illusory boundaries through an algorithm that uses partial differential equations and level-set methods. Such approaches, however, can be better characterized

as machine algorithms “inspired by” primate neuroscience than as computational models “of” primate visual function, which is the focus of the present article.

The other topic is neural synchrony, which has been proposed as a coding strategy underlying perceptual binding between disparate regions of the visual field (Wolf Singer 1999). Dong et al. 2008 have reported that synchrony depends on the border-ownership selectivity of the neuron pairs being recorded. Neural synchrony has become a bit of a “field unto itself,” with controversies about methods and meaning of results that are not at all related to figure-ground perception. Moreover, even if it were conclusively shown that, during figure-ground perception, neurons coding a figure fire synchronously while those coding the ground fire in anti-phase with the neurons coding the figure but in synchrony with each other, a key question would remain: How did each member of each population get recruited to fire when it did, as opposed to some other time? In other words, what functional receptive field properties underlie the “sorting” of parts of the visual field into figure and ground? This article has focused on the latter questions.

## Cross-References

- ▶ [Attentional Pop-Out](#)
- ▶ [Attentional Shroud](#)
- ▶ [Border-Ownership](#)

## References

- Barnes T, Mingolla E (2013) A neural model of visual figure-ground segregation from kinetic occlusion. *Neural Netw* 37:141–164. <https://doi.org/10.1016/j.neunet.2012.09.011>. Epub 2012 Oct 6.
- Bullier J (2001) Integrated model of visual processing. *Brain Res Rev* 36(2–3):96–107
- Diaz-Santos M, Cao B, Mauro SA, Yazdanbakhsh A, Neargarder S, Cronin-Golomb A (2015a) Effect of visual cues on the resolution of perceptual ambiguity in Parkinson’s disease and normal aging. *J Int Neuropsychol Soc* 21:1–10
- Diaz-Santos M, Cao B, Yazdanbakhsh A, Norton DJ, Neargarder S, Cronin-Golomb A (2015b) Perceptual,

- cognitive, and personality rigidity in Parkinson's disease. *Neuropsychologia* 69:183–193
- Dong Y, Mihalas S, Qiu F, von der Heydt R, Niebur E (2008) Synchrony and the binding problem in macaque visual cortex. *J Vis* 8(7):30.1–30.16. <https://doi.org/10.1167/8.7.30>
- Evans KK, Treisman A (2005) Perception of objects in natural scenes: is it really attention free? *J Exp Psychol Hum Percept Perform* 31(6):1476–1492. <https://doi.org/10.1037/0096-1523.31.6.1476>
- Gibson JJ, Kaplan GA, Horace N, Reynolds JR, Wheeler K (1969) The change from visible to invisible: a study of optical transitions. *Percept Psychophys* 5(2):113–116
- Grossberg S (1993) A solution of the figure-ground problem for biological vision. *Neural Netw* 6(4):463–483. [https://doi.org/10.1016/S0893-6080\(05\)80052-8](https://doi.org/10.1016/S0893-6080(05)80052-8)
- Grossberg S (1994) 3-D vision and figure-ground separation by visual cortex. *Percept Psychophys* 55(1):48–121. <https://doi.org/10.3758/BF03206880>
- Grossberg S, Mingolla E, Ross WD (1994) A neural theory of attentive visual search: interactions of visual, spatial, and object representations. *Psychol Rev* 101(3):470–489
- Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154
- Julesz B (1971) *Foundations of cyclopean perception*. The University of Chicago Press, Chicago. ISBN 0-226-41527-9
- Kaplan GA (1969) Kinetic disruption of optical texture: the perception of depth at an edge. *Percept Psychophys* 6(4):193–198. <https://doi.org/10.3758/BF03207015>
- Ko HK, von der Heydt R (2018) Figure-ground organization in the visual cortex: does meaning matter? *J Neurophysiol* 119(1):160–176. <https://doi.org/10.1152/jn.00131.2017>
- Kogo N, Strecha C, Van Gool L, Wagemans J (2010) Surface construction by a 2-D differentiation-integration process: a neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures. *Psychol Rev* 117(2):406–439. <https://doi.org/10.1037/a0019076>
- Lamme V (1995) The neurophysiology of figure ground segregation in primary visual-cortex. *J Neurosci* 15:1605–1615
- Land MF, Nilsson DE (2002) *Animal eyes*. Oxford University Press, Oxford
- Layton OW, Yazdanbakhsh A (2015) A neural model of border-ownership from kinetic occlusion. *Vis Res* 106:64–80
- Layton OW, Mingolla E, Yazdanbakhsh A (2012) Dynamic coding of border-ownership in visual cortex. *J Vis* 12:8. <https://doi.org/10.1167/12.13.8>
- Layton OW, Mingolla E, Yazdanbakhsh A (2014) Neural dynamics of feedforward and feedback processing in figure-ground segregation. *Front Psychol* 10(5):972. <https://doi.org/10.3389/fpsyg.2014.00972>
- Lettvin JY, Maturana HR, McCulloch WS, Pitts WH (1959) What the frog's eye tells the frog's brain. *Proc Inst Radio Eng NY* 47:1940–1951
- Pitts MA, Martínez A, Brewer JB, Hillyard SA (2011) Early stages of figure-ground segregation during perception of the face-vase. *J Cogn Neurosci* 23(4):880–895. <https://doi.org/10.1162/jocn.2010.21438>
- Roelfsema PR, Lamme V, Spekreijse H, Bosch H (2002) Figure-ground segregation in a recurrent network architecture. *J Cogn Neurosci* 14:525–537. <https://doi.org/10.1162/08989290260045756>
- Rubin E (1915) *Synsoplevede figurer* (visually experienced figures). Doctoral dissertation
- Sarti A, Mallad R, Sethian JA (2000) Subjective surfaces: A method for completing missing boundaries. *Proceedings of the National Academy of Sciences*, vol 97, issue 12, 6258–6263
- Sincich LC, Horton JC (2005) The circuitry of V1 and V2: integration of color, form, and motion. *Annu Rev Neurosci* 28:303–326
- Singer W (1999) Neuronal synchrony: a versatile code review for the definition of relations? *Neuron* 24:49–65
- Treisman A, Gelade G (1980) A feature integration theory of attention. *Cogn Psychol* 16:97–136
- Tyler CW, Kontsevich LL (1995) Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. *Percept Lond* 24(1):127–154. c
- von der Heydt R (2015) Figure-ground organization and the emergence of proto-objects in the visual cortex. *Front Psychol* 6:1695. <https://doi.org/10.3389/fpsyg.2015.01695>
- von der Heydt R, Zhou H, Friedman HS (2000) Representation of stereoscopic edges in monkey visual cortex. *Vis Res* 40:1955–1967. [https://doi.org/10.1016/S0042-6989\(00\)00044-4](https://doi.org/10.1016/S0042-6989(00)00044-4)
- Wolfe JM, Horowitz TS (2017) Five factors that guide attention in visual search. [Review article]. *Nat Hum Behav* 1:0058. <https://doi-org.ezproxy.neu.edu/10.1038/s41562-017-0058>
- Wurtz RH (2015) Using perturbations to identify the brain circuits underlying active vision. *Philos Trans R Soc B* 370:20140205. <http://rstb.royalsocietypublishing.org/content/370/1677/20140205>
- Zhaoping L (2005) Border ownership from intracortical interactions in visual area V2. *Neuron* 47(1):143–153
- Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. *J Neurosci* 20:6594–6611