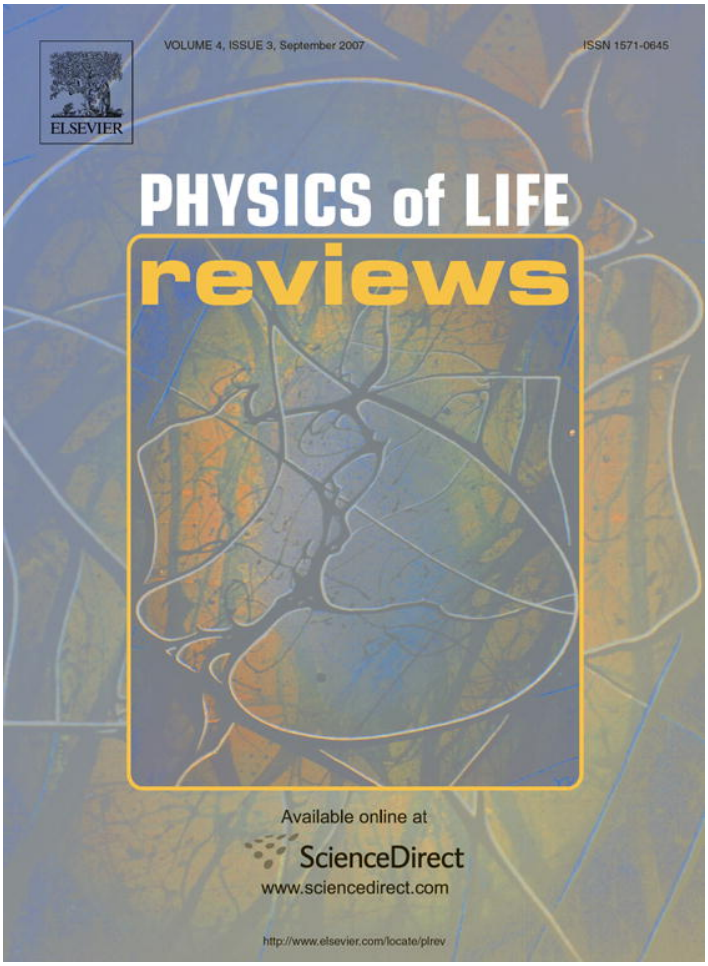


Available online at www.sciencedirect.com
ScienceDirect
Physics of Life Reviews 4 (2007) 189–222
PHYSICS of LIFE reviews
www.elsevier.com/locate/plrev

Review

# Seeing surfaces: The brain's vision of the world

Heiko Neumann [a], Arash Yazdanbakhsh [b,c], Ennio Mingolla [b,*]

[a] *Ulm University, Faculty of Engineering and Computer Sciences, Institute of Neural Information Processing, Germany*
[b] *Boston University, Department of Cognitive and Neural Systems, USA*
[c] *Harvard Medical School, Neurobiology Department, USA*

Received 9 August 2007; received in revised form 4 September 2007; accepted 4 September 2007

Available online 19 September 2007

Communicated by L. Perlovsky

**Abstract**

Surfaces of environmental objects are the key to understanding the visual experience of primates. Surfaces create structure in patterns of light available for sampling by visual systems, and delineate potential interactions that an animal can have with its environment, such as approaching goals, avoiding obstacles, grasping an object, or identifying members of a social group. Recent progress in modeling the perception of visual surfaces highlights the importance of feedforward and feedback connections in visual neural networks that segregate and group visual input into coherent regions related to corresponding surfaces in the visual world. Rich non-linear network dynamics in the brain underlie surface perception, including the detection, regularization, and grouping of visual boundaries between surfaces, the determination of "ownership" of a boundary by a closer surface that partially occludes a background, and the apprehension of a surface's visual quality, such as color or texture. Recent modeling efforts on these fronts are reviewed.


*PACS:* 87.19.L-; 84.35.+i; 87.19.lj; 87.19.lt

*Keywords:* Neural networks; Perception; Receptive field; Visual system

**Contents**



* Corresponding author.
*E-mail addresses:* heiko.neumann@uni-ulm.de (H. Neumann), Arash_Yazdanbakhsh@hms.harvard.edu (A. Yazdanbakhsh), ennio@cns.bu.edu (E. Mingolla).

1571-0645/$ – see front matter 
doi:10.1016/j.plrev.2007.09.001

## 1. The surfaces of our visual world

A key visual competency of many species, including humans, is the ability to rapidly and accurately ascertain the sizes, locations, trajectories, and identities of objects in the environment. Whether noticing a deer moving behind a thicket, locating edible plants, or steering around obstacles to reach goals, many of the tasks of vision can be understood as stemming from the need to guide behavior based on changing visual input, whose patterns of light are structured by the geometries of surfaces and illuminants. The psychologist James J. Gibson [1–3] devoted enormous energy to articulate the importance of surfaces as the interface between light and objects. Gibson noted that the patterns of light are highly specific to the layout of surface in a particular environment and thus, in principle, informative for any animal's visual system. The difficulties begin, however, when one considers the virtually infinite variation in the patterns of visual input that follow from changes of an observer's viewpoint, or from motion of objects in the environment, or from disturbances in the patterns of illumination (e.g. from a passing cloud). The pattern of visual stimulation is constantly changing, whether from changes of perspective or progressive occlusion and disocclusion of objects, while our brains somehow (correctly!) perceive a stable underlying world through all the optical transformations.

The challenge of using patterns of light to perceive a world is evident to any one who has ever tried to do computer vision. More than a quarter of a century has passed since the publication of David Marr's Vision [4], a volume that was a manifesto for a generation optimistic that artificial intelligence and computational vision would soon unlock secrets that had eluded generations of philosophers and sensory psychologists. While tremendous progress on machine vision has certainly been made in the interim, the rate of progress has undershot expectations from the 1980s. Getting a machine to perform anything like what animals do—navigating and interacting with objects based on available light from unknown sources—remains an elusive challenge.

A key thesis of this review is that much of the bottleneck is due to an initial reluctance to integrate findings from perceptual psychology and, more recently, neurophysiology. As a result, many previous models have tried to go from simple elementary features, whether "raw" pixel values or simple blobs or edges, directly to such high-level competencies as object recognition. Typically, such models would work up to a point for relatively sparse or noise-free data sets, but fail to generalize to more complex environments. The increase in computer processing power has steadily increased the size of data sets that can be processed, but fundamental obstacles still exists, and many of these concern surface perception as the key intermediate step between patterns of light and the richness of visual experience.

This article will not address all the important aspects of even surface perception. Specifically, we will not address the important issue of surfaces in motion, beyond noting some rudimentary issues in the immediately following paragraphs. Instead we will focus on the case of static surface perception. We note in passing that there is no such thing as truly "static" perception of anything, as our eyes are constantly making small movements (tremor, drift, or micro-saccades) even if we are "fixating" on a stationary target in a room, but for analytical purposes it is useful to separate out such cases from others where objects or the observer are making macroscopic movements. As we will describe, significant progress has been made in recent years in both experiments and modeling of surface perception.

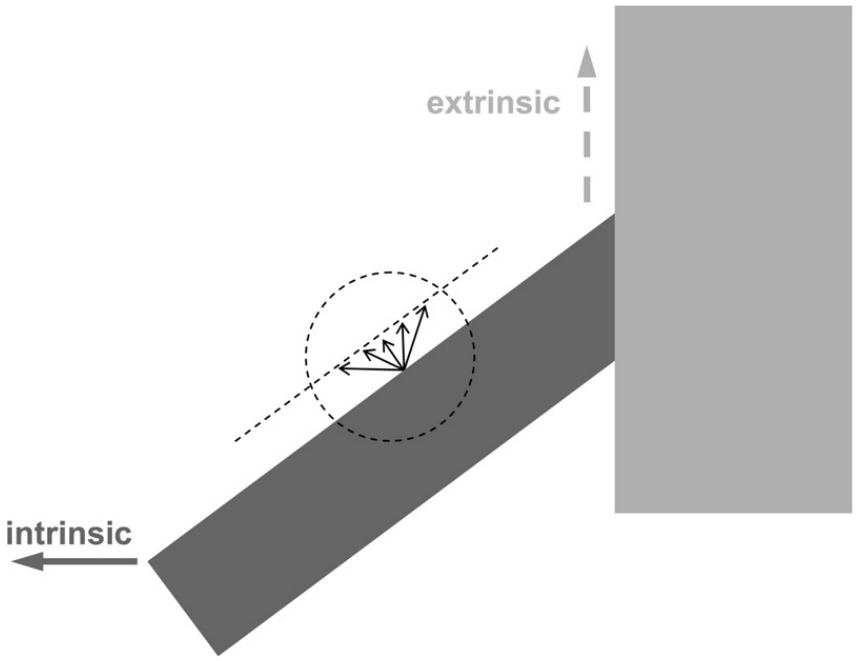



Fig. 1. A diagonally oriented dark gray rigid bar moves to the right. Any motion detector that can measure only in the zone indicated by the dashed circle can pick up at most the component of motion that is normal to the orientation of the line, a restriction known as "the aperture problem". As far as that detector is concerned, the point at the base of the five depicted vectors could be moving along any trajectory indicated by a vector ending on the dashed line. The far left corner feature of the dark gray bar, termed an "intrinsic terminator", veridically reveals the leftward motion of the dark bar. The local motion signals at the upper right of the bar, however, are termed "extrinsic", because they track upwards, along the left edge of the lighter vertical bar, which is positioned closer to the observer.

The following treatment of basic problems in the perception of moving surfaces follows that of Berzhanskaya, Grossberg, and Mingolla [5]. Wallach [6] first showed that the motion of a featureless line seen behind a circular aperture is perceptually ambiguous: for any real direction of motion, the perceived direction is perpendicular to the orientation of the line; i.e., the *normal component* of motion. This so-called *aperture problem* is also faced at a mechanistic level by any localized motion sensor, such as a neuron in the early visual pathway, which responds to a local contour moving through an aperture-like zone of visual coordinates called a *receptive field*. (The general concept of receptive fields is reviewed in Section 2.) In contrast to an extended straight line, a moving dot, line end or corner provides unambiguous information about an object's true motion direction [7]. In the example in Fig. 1, motion of the left line end corresponds to the real motion of the line. The right line end is formed by the boundary between the line and a stationary occluder, and its motion provides little information about the motion of the line, as noted by Bregman [8] and Kanizsa [9]. Nakayama, Shimojo and Silverman [10] have suggested classification of terminators as *intrinsic* and *extrinsic*: an intrinsic terminator belongs to the moving object; an extrinsic one belongs to the occluder. Motion of intrinsic terminators is taken into account in computing the motion direction of an object, while motion of extrinsic terminators is generally ignored [7,11].

The chopsticks illusion of Anstis [12], shown in Fig. 2, points out how crucial it is for the visual system to correctly assess motion of intrinsic vs. extrinsic terminators, as qualitatively different motion percepts result from displays with identical stimulus motions, but different static elements. Clearly the determination of the proper motions of surfaces in the world is inextricably bound with the problem of determining which surfaces are in front of and partially occluding others from the observer's viewpoint.

These elementary illustrations serve to point out a truism of the Gestalt movement in sensory psychology, which first broke with the elementarism of the structuralists of the 19th century: The units of visual perception are not elementary dots, whether pixels in some artificial sensor array, or the outputs of individual photoreceptors in an animals eyes. Rather, there are important organizational steps required to find consistent and coherent coalitions of stimulation over time, before the larger "meanings" of visual patterns can be determined. In this spirit the important computational manifesto of Witkin and Tenenbaum [13] proposed that chunks of visual structure extracted from visual images act as "semantic precursors" that enter the interpretation processes of scene and object recognition.

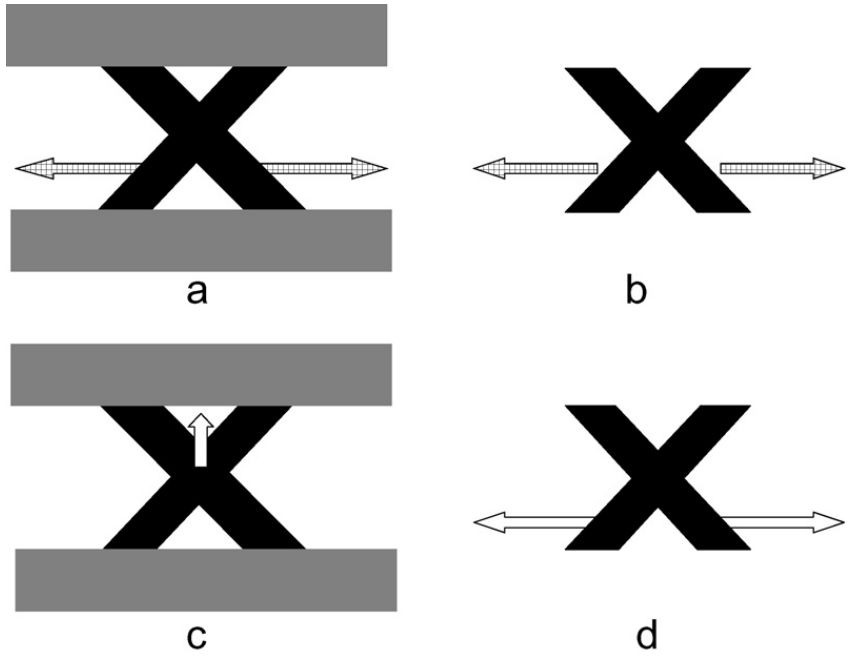



Fig. 2. In the "chopsticks illusion" of Anstis [12] the dark bars of (a) and (b) undergo identical motion, in both cases horizontally away from the common center of the two bars forming the respective "X" patterns, as indicated by the hatched arrows. (c) Indicates the percept seen for display (a), whereby signals from the extrinsic terminators of the "X" are discounted, and the common vertical components of the two bars' motions result in a percept of a rigid "X" moving upwards, partially obscured by two occluding bars. The percept corresponding to (b), however, as indicated in (d), is of two "chopsticks" undergoing a scissors-like motion, each moving independently, as indicated by the empty left and right arrows. Typically one of the chopsticks is seen as being slightly closer to the observer than the other, often with illusory contours linking the two halves of the closer chopstick.

In the second half of the 20th century the dominant conception of visual processing derived from neurophysiological evidence was a hierarchical model most closely identified with Hubel and Wiesel [14]. Their "feedforward" view was that the early stages of the visual system start by extracting relatively general and localized features of visual input, such as dots, lines, or edges, and that each subsequent stage of processing has increasing tolerance for shifts or other deformations of features preferred by the previous stages. Each next stage also prefers more specific combinations of earlier features, leading to more specialized detectors at times parodied as ending in "grandmother cells"—cells "tuned" to respond only to views of one's grandmother [15]. Interestingly, this model has more in common with 19th century associationism than with the Gestalt movement of the 20th century. This paper describes how the intuitions of the Gestaltists have influenced neural network models formulated as systems of non-linear ordinary differential equations whose components can be identified with neurophysiological mechanisms and whose dynamics describe behavioral levels of brain function. We begin by a review of some physiological preliminaries.

## 2. Brain mechanisms for surface perception

### *2.1. The concept of a neuron's receptive field*

The receptive field of a visual neuron is the region of the visual field where modulation of light can change the neuron's activity. Note that although a neuron can be physically located anywhere in the brain (including the retina), the receptive field of the neuron is always described with reference to coordinates in the visual field from the viewpoint of the observer. Although this definition seems straightforward, accumulating evidence in recent decades indicates that it is not an easy matter to describe a "fixed" receptive field for many neurons in a way that could be easily modeled by a spatial weighting function (e.g. spatially truncated Gaussian).

One practical way to find a neuron's receptive field is to use a minimal stimulus like a flashing small bar that is displayed many times in many places across the visual field. The experimenter then notes the extent of modulation in the neuron's activity as a function of the location and timing of presentations of the flashing bar (Fig. 3). Note that this procedure requires that the eye be as still as possible in its orbit, as when one stares or "fixates" on a point in a scene.

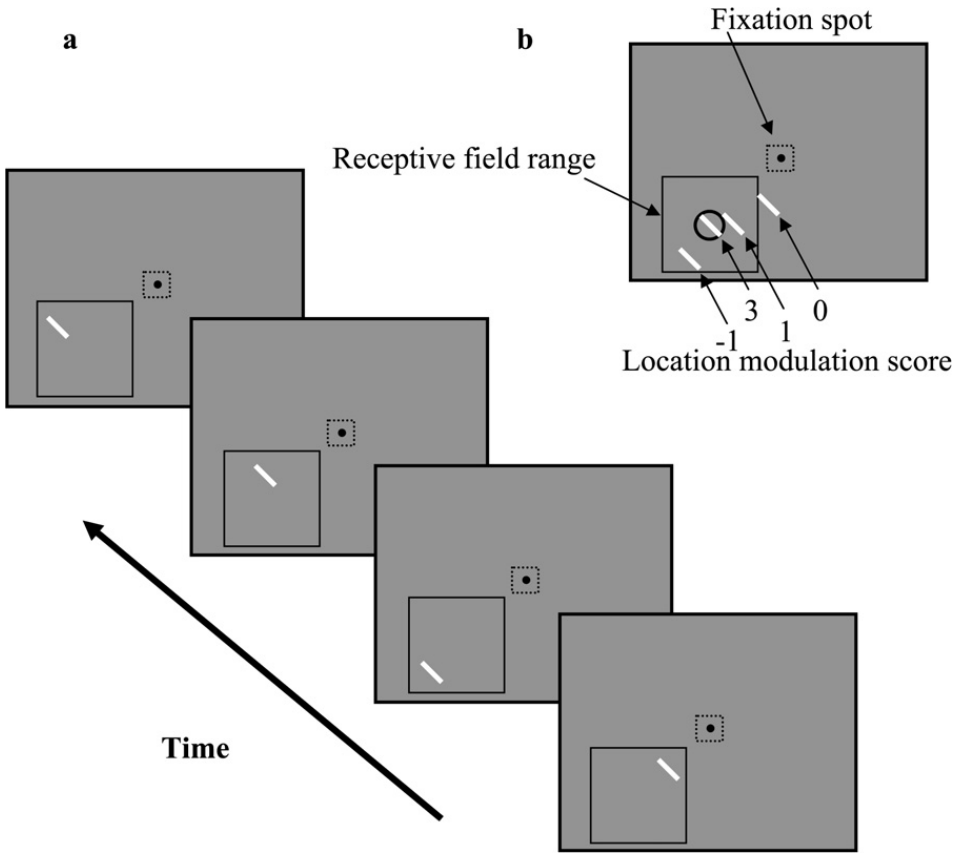



Fig. 3. (a) The sequence of randomized locations of a flashing bar for receptive field mapping. (b) Based on the modulation of the cell activity compared to the base activity for each location of bar flashing, the score is assigned.

Then each point is assigned a modulation score based on the amount that cell activity changes when the bar is flashed at that point. The activity of the cell is recorded by an electrode tip that is placed close to the cell following a surgical procedure. This method can reveal what is called the *first order* receptive field of a given neuron [16,17].

Fig. 4a, depicts the $xyz$ diagram of a first order map, where $z$ corresponds to the modulation score at each spatial $xy$ coordinate, where the degree of modulation is determined by counting the action potentials (spikes) of the cell in a given temporal interval. The receptive field has the excitatory and inhibitory parts, meaning that, for the excitatory central region (depicted in orange)[1] an increase in light stimulation is correlated with an increase in neural activity (firing rate of action potentials). Light increases in the peripheral region (depicted in blue), on the other hand, tend to depress the baseline activity of the neuron. The excitatory part has a higher activity score compared to the non-modulated part (the peripheral part of the surface diagram of Fig. 4a, where the baseline score is zero). The inhibitory part has a lower score compared to the non-modulated part (the trough of the surface diagram in Fig. 4a). Note that Fig. 4b is a schematic simplification. Whereas excitatory and inhibitory zones are shown as uniform regions, variations of excitatory and inhibitory activations exist throughout actual neural receptive fields.

### 2.2. *Complexities involved with the notion of receptive field*

#### 2.2.1. *Beyond the first order response*

Even the receptive field shown in Fig. 4 has its own shortcoming for full characterization of a cell's response. Fig. 5a's left panel shows an illusory "Kanizsa square" with pacmen elements and corresponding *neon color spreading* in the right panel. The subjective components are outlined in Fig. 5b. A relevant example comes from a finding by von der Heydt, Peterhans, and Baumgartner [18], who showed that there is a contextual effect on the cell response in certain brain regions, even though the only stimulus elements that appear are *outside the first order receptive field*

[1] For interpretation of the references to color, the reader is referred to the web version of this article.

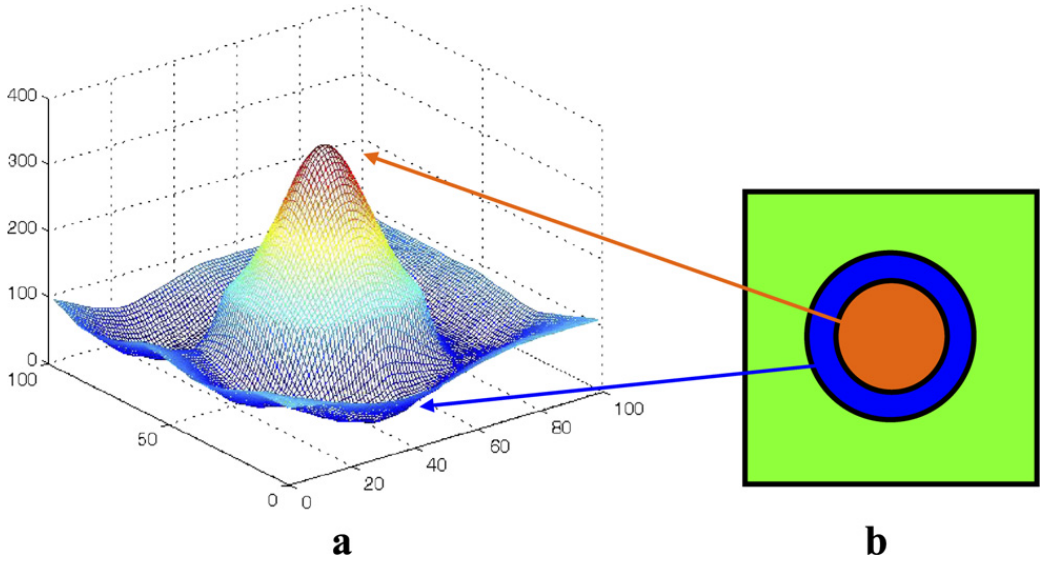



Fig. 4. (a) Surface plot for first order receptive field. (b) Schematic representation of the first order receptive field. See text for details.

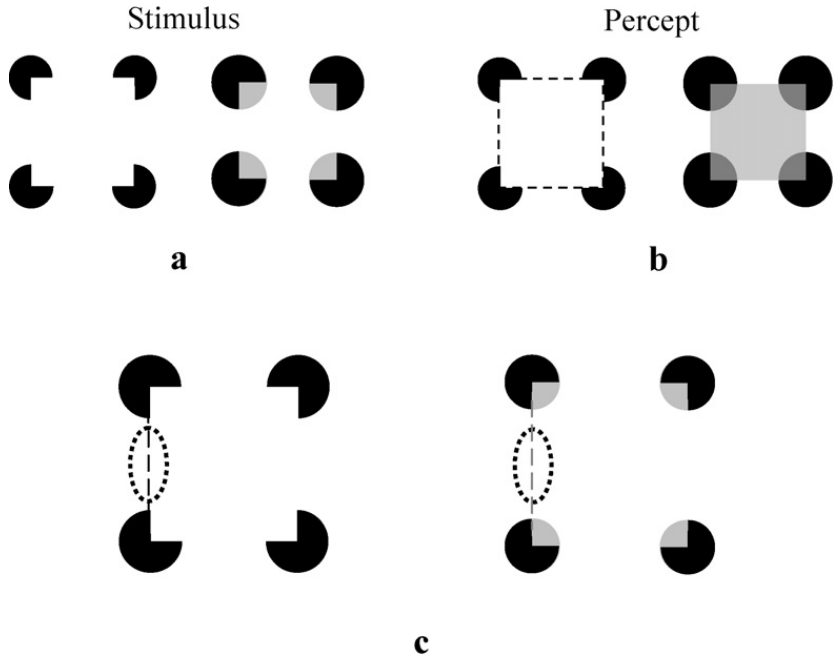



Fig. 5. (a) Left, subjective square induced by Kanizsa pacmen configuration, right, neon-color-spreading version of the Kanizsa square. (b) Outlined subjective percept in (a). (c) Certain cells in visual area V2 give a positive response, even when the pacmen are outside of the zone for obtaining a first order map (indicated by the dotted ellipses).

region (Fig. 3). Pacmen are outside the schematic receptive field, yet the cell activity is modulated by the simultaneous presence of the pacmen on the both sides.

Interestingly, for some neurons, the response modulation is maximum when the two pacmen are simultaneously present; when just a single pacmen is present, the response modulation is negligible. This fact can be described as a *second-order response* [17,19,20] which means the addition of the effects of the two elements is non-linear and this non-linearity is supra-additive; i.e., the effects of having two pacmen are much stronger then the linear addition of their individual effects. Therefore, the receptive field has response properties beyond the first order response. A modeling approach to this non-linearity has been proposed in the form of "bipole grouping", as is described Section 3.3 [21,22].

If in Fig. 3 instead of one bar per each frame, two bars in random positions are presented and the modulation score of the cell is plotted in terms of offset of the two bars, a second order map like the one in Fig. 6a will be generated.

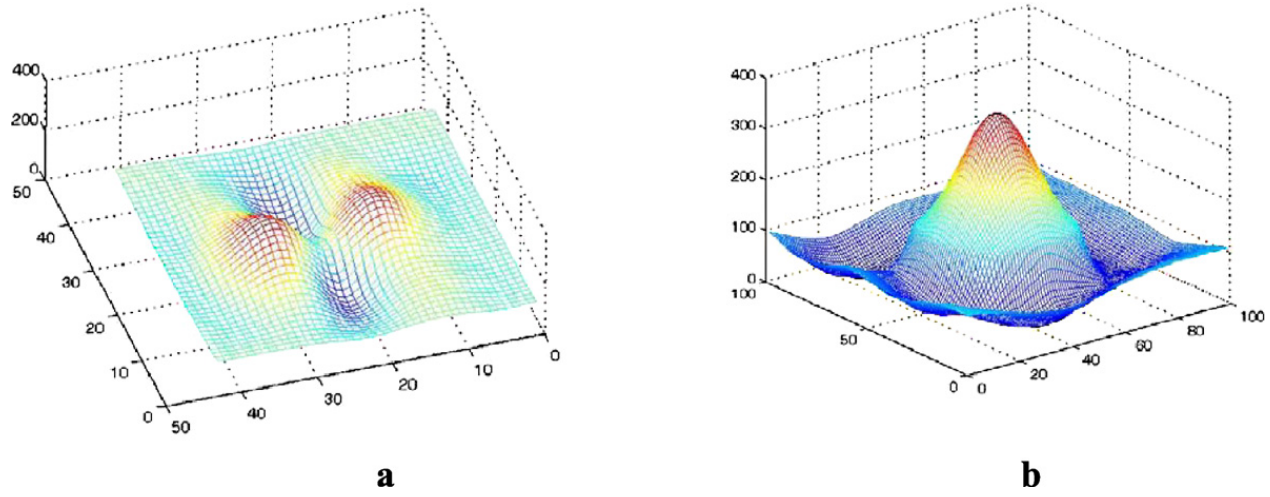



Fig. 6. (a) Second order map of a neuron. (b) First order map of the same cell.

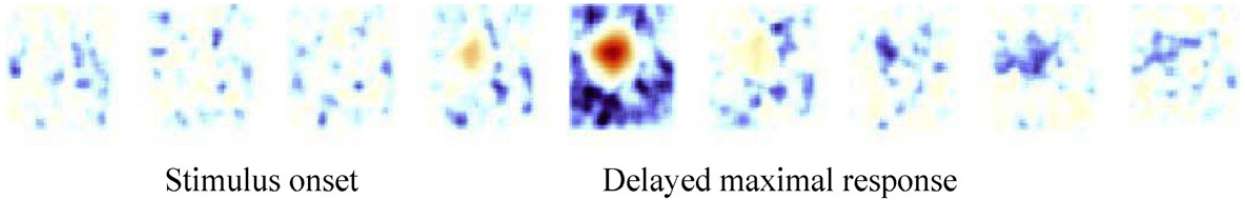



Fig. 7. Temporal evolution of the first order receptive field.

Comparing the second order map of Fig. 6a and the first order map of the same cell (Fig. 6b) shows the cell response is also tuned to the co-presence of the two collinear bars along a preferred orientation. This is a minimal example of context dependency. Notably, this orientation selectivity is barely visible in the first order map (all data maps in this section are provided courtesy of the Margaret Livingstone Laboratory at Harvard Medical School).

### 2.2.2. *A neuron's receptive field evolves in time*

Figs. 4 and 6 may give the impression that the receptive field is constant in time. Fig. 7 shows a time sequence of first order receptive field evolution. It can be seen that the maximal response has a delay from the onset of the stimulus.

This view of the evolution of a neuron's receptive field adds a new, temporal dimension to response of cells that are described in many textbooks only as spatial weighting functions of the distribution of light. There are related models that suggest ways of understanding the temporal evolution of such receptive fields. In the next section we briefly review a possible approach to understanding these findings.

## 2.3. *A basic modeling approach to the concept of receptive field*

Fig. 8 shows a simplified neural model of neural connections, depicting feedforward, feedback, and lateral connections within three layers of neurons.

In this conception, neural connectivity has delays stemming from two sources, synaptic connections (where the tip of arrows point the cells), and the limited conduction velocity of the wiring (body of the arrows). This simple network conceptualizes the temporal evolution of a receptive field, including first order and second order maps, and predicts the context dependency of cell response. Note that its critical novelty, as compared with the Hubel/Wiesel conception described in Section 1, is the importance of "top–down" feedback and also the related recurrent loops made by lateral connections between units [23]. Delays related to feedback connectivity underlie the evolution of each cell's response. The first order response of many neurons can be traced to the geometry of feedforward connections. The second order response evolves differently from the first order one and can be attributed to the rich feedback connectivity of the units. The time sequence of the first order map shows a delayed minor inhibition in the center of the receptive field

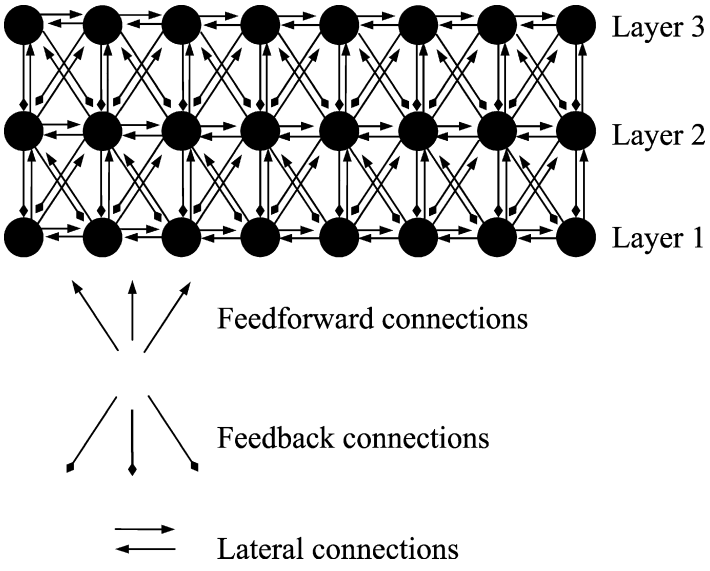



Fig. 8. A simple model for neural connectivity to describe time evolvement of a receptive field, its context dependency and late response. Feedforward, feedback and lateral connections are illustrated.

(Fig. 5; last panel on the right). One possible explanation for this extra delay can be the later feedback signal, with an inhibitory nature. As can be seen by the few examples presented above, the characterization of a cell's receptive field is subtle and varies with stimulation context.

### *2.4. Surface representations in early visual: Filling-in of featural quality*

Receptive fields in early visual areas are limited in space, though the extent of the spatial limit may be context dependent [18,24–30]. Despite the possible variability of receptive field size due to the overall context, the limited spatial extent of the receptive field poses a constraint on the representation of surfaces by neural coding. How are surfaces represented in the brain? A spatially limited receptive field as sketched in Fig. 4b can detect only the local changes (Fig. 9). A uniform surface, without a local change, can result in null activity of the cell (or spontaneous firing rate), because many neurons are "balanced" between excitation (e.g., in their center) and inhibition (in a radial surrounding region) in response to equal quantities of light throughout a patch that covers their receptive field. Is null activity enough for generating a surface representation in the brain? The null activity is the same for all surfaces irrespective to which object the surface belongs and with which boundaries the surface is surrounded. This potentially could mean that each surface in a visual scene loses its featural identity, because all of them have the label of null activity.

The problem of null activity in response to uniform regions of light by most neurons has led to a long-standing historical controversy about the surface representation. Filling-in is one of the suggested solutions to the problem. Note that if a surface belongs to one object, it should have a closed boundary. A practical way to group all points of the surface together is to fill it with a metaphorical liquid. Because the bounded region is closed, the liquid remains within the closed area. If the boundaries do not belong to a unique surface, then the metaphor of liquid is again helpful. It drains out from non-closed boundaries [31,32]. The metaphor turns out to be useful, because, otherwise categorizing the surfaces between closed and non-closed ones becomes a challenging mathematical problem to determine which boundaries are closed and which ones are not.

If the liquid metaphor has something to do with how the brain represents a surface, is there a topographical assembly of neurons to represent the filled-in area? The question has a hidden assumption, namely, that of an isomorphic surface representation [33]. The locations of visual features can be represented in our brain in a one to one and continuous way, with every location in a topographic brain map having a corresponding location in visual coordinates. What is the evidence for this assumption? Interestingly, until today, the findings regarding the question of filling in are heterogeneous and inconsistent. While some studies show the evidence of filling in certain visual areas others studies assert the lack of such phenomenon in the same visual areas. Komatsu [34] reviews a good deal of the recent psychophysical and neurophysiological evidence supporting the filling-in proposal. Another review of this topic can

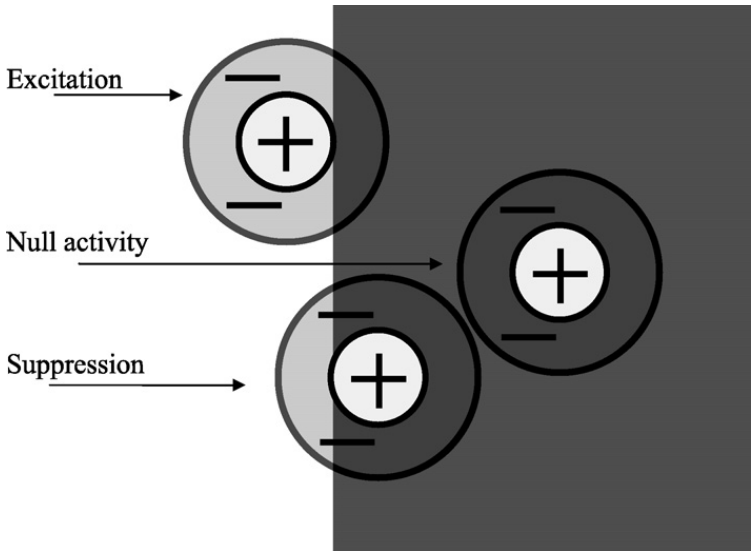



Fig. 9. The limited spatial extent and structure of excitatory and inhibitory subregions of a receptive field enables it to better respond to edges rather than uniform surfaces.

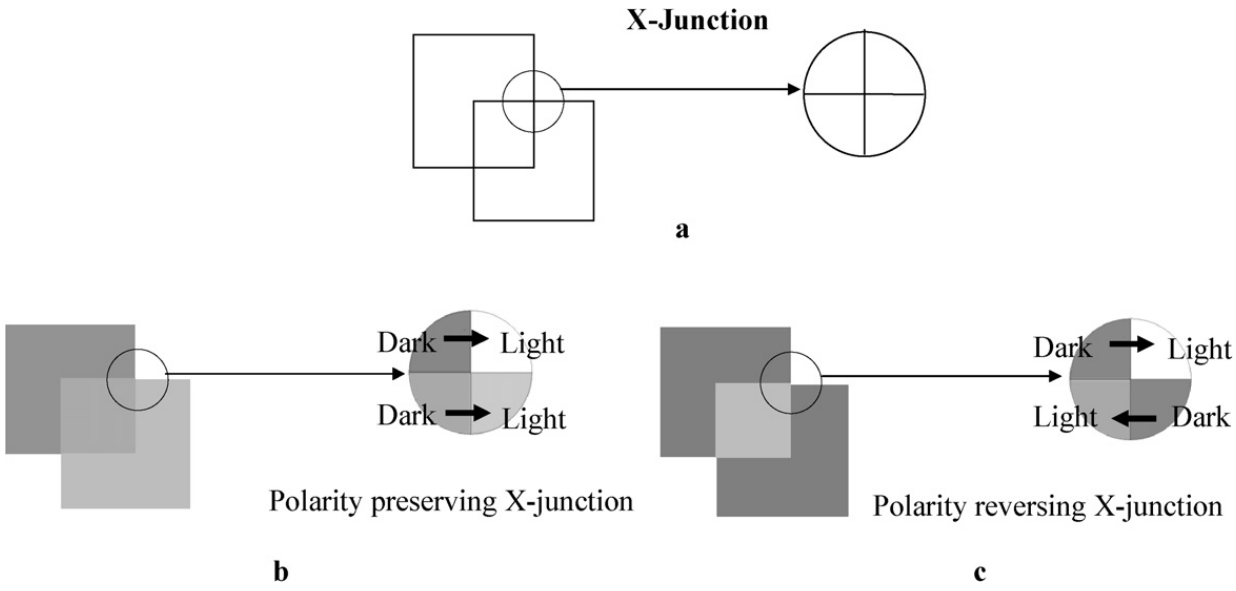



Fig. 10. (a) Two partially overlapping square regions create a characteristic local geometry, termed an "X-junction". (b) The bottom square is perceived as a transparent one over the top square. The X-junction is polarity preserving. (c) No overlap of one square over the other is evident; the percept rather is like a central bright small square in the middle of the two thick L-shaped darker parts. Note that the geometry of edges is identical in (b) and (c).

be found in Neumann and Mingolla [35]: in comparison to our understanding of processes of perceptual boundary formation, the situation with respect to surface quality perception is considerably less developed. Many models are based on some mechanism to laterally propagate activities to from sparse estimates or surface brightness or color derived near boundaries. Improved empirical techniques juxtaposed with hypotheses developed by computational modelers offer the hope that coming years will see progress in this area, as has already occurred in the field of contour representation.

### *2.5. Surface depth percept via transparency: Contrast relations within X-junctions*

The perception of transparency from a 2D image is a good example of how our visual system is able to exploit 2D cues to stratify the image in depth and give us the impression of partial occlusion: In Fig. 10b the bottom square is

perceived as a transparent square over the top square. Interestingly, different contrast relation within the same geometry (as outlined in Fig. 10a) gives dramatically different depth stratification as if a small bright square is surrounded by two darker L-shaped flankers at the same plane (Fig. 10c). It was in 1974 that *Metelli* [36] introduced the scission hypothesis of transparency: the visual system split the luminance of a 2D surface into separate depths. There are locations in Figs. 10b and 10c with the richest information content to uniquely determine the surface grouping. These areas provide 2D local cues for surface grouping in depth and are historically termed junction zones. The zoomed areas in Figs. 10b and 10c are called *X-junctions*. In terms of relative darkness and lightness, the transparent case X-junction is polarity preserving (Fig. 10b), and that of non-transparent one is polarity reversing (Fig. 10c).

## 3. Surface boundary formation and chunking by grouping

### *3.1. Visual boundary generation in vision—issues and problems*

How does the brain proceed to form surface representations from initially detected feature items of moving objects as motivated in the previous section? In monocular vision, changes in luminance (e.g. illumination changes or shadows), abrupt changes in texture gradient (when different materials are juxtaposed), or mutual occlusions between surfaces seen from a distinct vantage point often lead to vivid *contrasts* in one or several feature domains, e.g., luminance or brightness, orientation, etc. Also, contrasts occur for stereoscopic depth changes (depth discontinuities or changes in surface orientation) as well as in cases of discontinuities in groups of moving items (kinetic contours). Fig. 11 (left) sketches a typical scene configuration that leads to such discontinuities in luminance images. Due to the coherence of surface material properties, image features that are sampled by the retina from neighboring regions are not mutually independent from each other. For example, local contrasts are often aligned along a given image orientation [37] and can thus readily be combined (or grouped) to form local contour fragments that act as precursors to code prototypical object configurations [13], again emphasizing the importance of the rules of Gestalt organization (see Section 1).

The visual inference of surface *boundary* information requires more than simple contrast detection and local grouping. The formation of boundaries refers to the segmentation of a scene into apparent surfaces and their apparent qualities. This requires global information about the actual scene arrangement and the associated visual context. Many demonstrations show that often local cues that were distant from each other can be combined to form a coherent surface representation although no apparent contrast is present between the localized items. For example, Nakayama [38] in a series of experiments demonstrated vivid examples of formation of continuous apparent surfaces from locally disparate items and also how apparent surface qualities are perceptually created through an emergent process of surface completion (see Fig. 11b). This suggests that visual boundary formation is critical for the segmentation of complex scene arrangements into surfaces and visible objects. Evidence suggests that boundaries are a necessary prerequisite to assign perceptual surface qualities, such as transparent color, or surface lightness [39,40], which are strongly influenced by more global context effects.

As already motivated in Section 1, building surface representations necessitates distinguishing boundaries that are *intrinsic* to a surface and those that are *extrinsic*. The Gestaltists in the 1930s [41] observed this phenomenon and suggested that a surface boundary carries a *boundary ownership* property. In other words, a surface boundary is assigned to the surface whichever side of a boundary that created the perception of one surface in front of another.[2] Occlusion boundaries depend on the current viewpoint of the observer and the presence of multiple objects in the scene and their possible motion trajectories, as discussed in the previous section. Fig. 12 shows configurations in which three surfaces partially occlude each other all situated in front of a homogeneous background, leading to image boundary configurations with various perceptual ownership assignments. Mutual occlusions often generate *junctions* in the luminance pattern through a sudden change in boundary orientation where the background object becomes visible because it is no longer occluded by the surface in front. Surface occlusions often generate junctions of the so-called T- or X-types, depending on whether the occluding surface is opaque or transparent. (See Section 2.5 for the

[2] This is true for surface boundaries which are created by surface segments that are separated in depth and therefore disconnected from surface regions that are neighboring in the two-dimensional projection obtained from the point of view of the eye. In the case of discontinuities in surface orientation, i.e. edges formed by dihedral angles for which both surfaces are visible, the assignment is somewhat arbitrary, however.

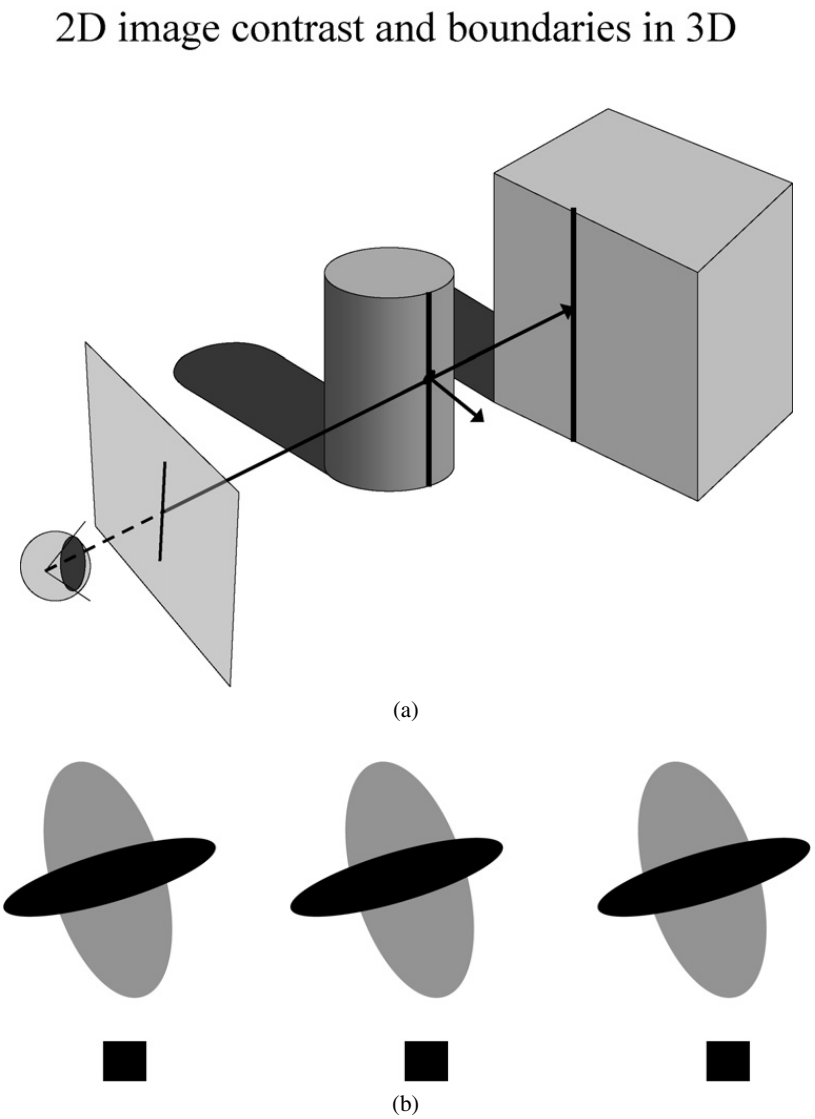



Fig. 11. Boundary generation in single luminance images. (a) Boundaries can be generated by discontinuities in one or multiple physical parameters describing combined surface and photometric properties in the viewed scene and their interaction with the relative viewing direction. As an example a luminance contrast is generated by a distant object (plane surface) that is partially occluded by a (smoothly curved) surface of an object that is situated closer to the observer. (b) Boundaries are not necessarily congruent with physical changes in the scene that correspond to luminance or color contrasts in the image. Fusion of the stereogram leads to a percept of two elliptical mutually occluding shapes hovering at different depths. In the case where the small dark ellipse appears in front the percept is one of an opaque surface patch occluding a lighter surface in the background. In the case where the depth arrangement is reversed the lighter surface appears in front and partially occludes the more distant dark surface. In this case now, the appearance of the light surface changes into transparency and a faint bounded light gray film appears in the occlusion region. This demonstrates that boundary and surface appearance can change depending on the interpretation of the scene arrangement. (*Notes*: The stereogram has been generated after K. Nakayama. Stereoscopic surface perception. *Proceedings of the National Academy of Science*, 24: 919–928, 1999; the small squares at the bottom have been included to help free fusing the images.)

discussion of transparency.) Similar considerations also hold true when boundaries are created by textured surfaces. In cases of mutual occlusions the projected textures undergo sudden changes in the statistics (when different types of textures occlude), their orientations (occlusion of oriented surface textures) of the projected patterns, or in texture gradient (mutual occlusion of curved surfaces).

This all demonstrates that the neural mechanisms of boundary integration are multi-faceted in their contribution and that cell responses to localized inputs can be altered by context information such as those generated by apparent surface properties. In particular, the region of support for such grouping could be the result of distinct patterns of interacting components. It is, however, far from clear how remote activities in a network influence localized spatial activation patterns of a target through integration. An intuitive approach suggests that a measure of local elements is

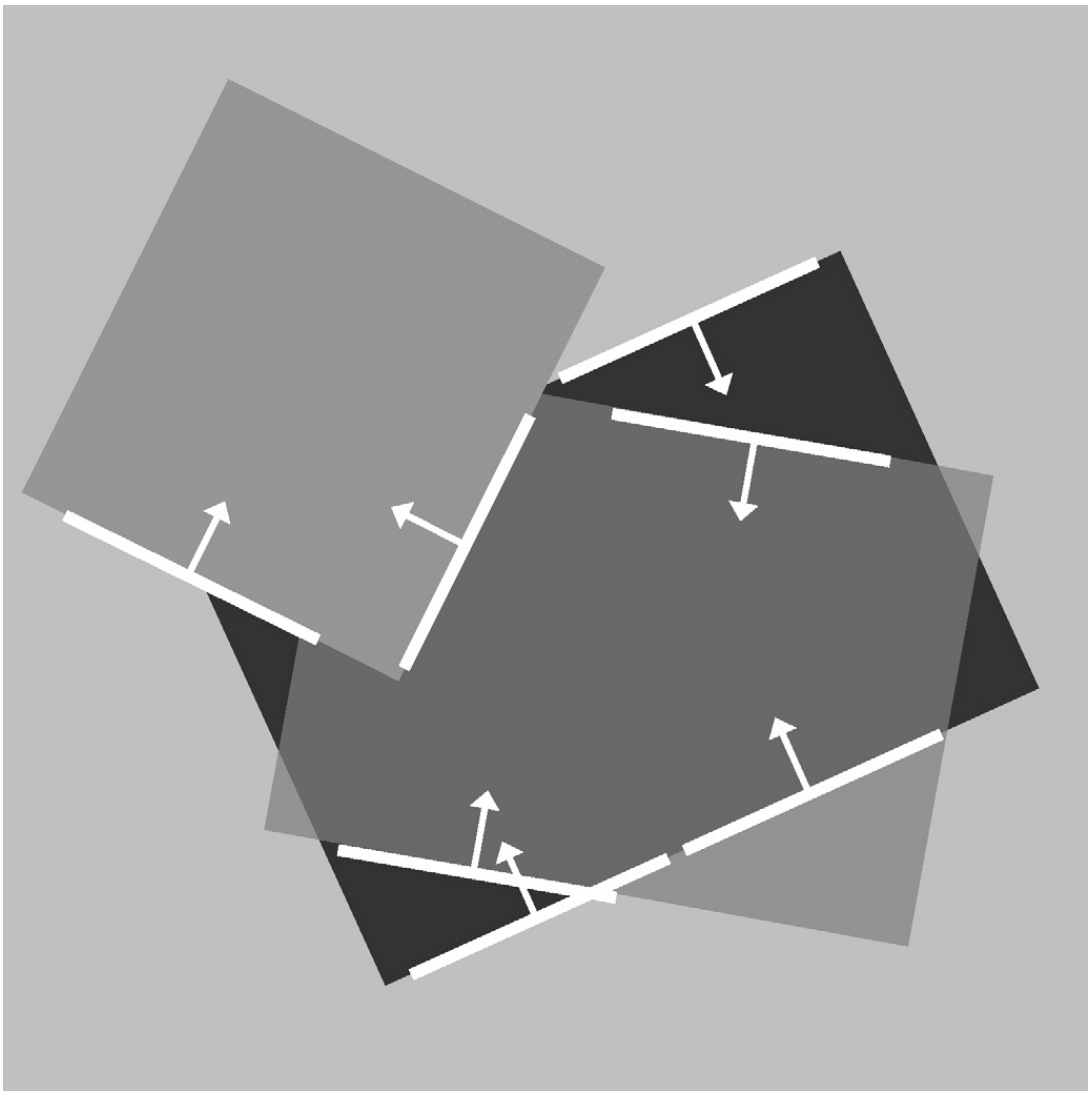


Fig. 12. Border ownership, occlusions and transparency. Three mutually overlapping rectangular surface patches are situated on a light background surface. The gray patch in front of the dark patch appears to be transparent. The key issue is the assignment of the boundary segments to those surface regions that they delineate from the background or occluded surfaces (border ownership). The white bars selectively depict segments of surface boundaries with the arrows pointing towards the surface region that "owns" the boundary segment. Mutual occlusions often generate junctions of different configuration. In the display the occlusion junctions that occur where the transparent surface occludes an opaque patch are of the X-type with smooth continuations of the collinear arms that belong to the same surface. When an opaque surface occludes patches T-junctions occur in the monocular image.

aggregated to form larger more complex visual elements over a sequence of hierarchically organized stages. This view was mainly supported by the seminal work of Hubel and Wiesel [14] that indicated that cells respond to increasingly complex patterns at each successive stage of the cortical architecture (compare Section 2). For tasks demanding fast processing and decision making it seems that mainly this forward sweep of processing is sufficient for arriving at basic scene categorizations [42]. However, for tasks, which require more time to come to a conclusion, such as in figure-ground segregation of camouflaged objects, initial responses need to be further elaborated. This suggests that the picture of a primarily hierarchically organized processing architecture needs to be changed. In a nutshell, initial elements measured locally based on cells' receptive field properties are integrated through convergent signal transmission along feedforward pathways. However, activity at later processing stages in the brain can influence these initial responses, such that the existence of a reverse hierarchy of information flow has been suggested (see [43] for a synopsis). In particular, the basic principles of cortical mechanisms for combining and subsequent influencing earlier responses are implemented in

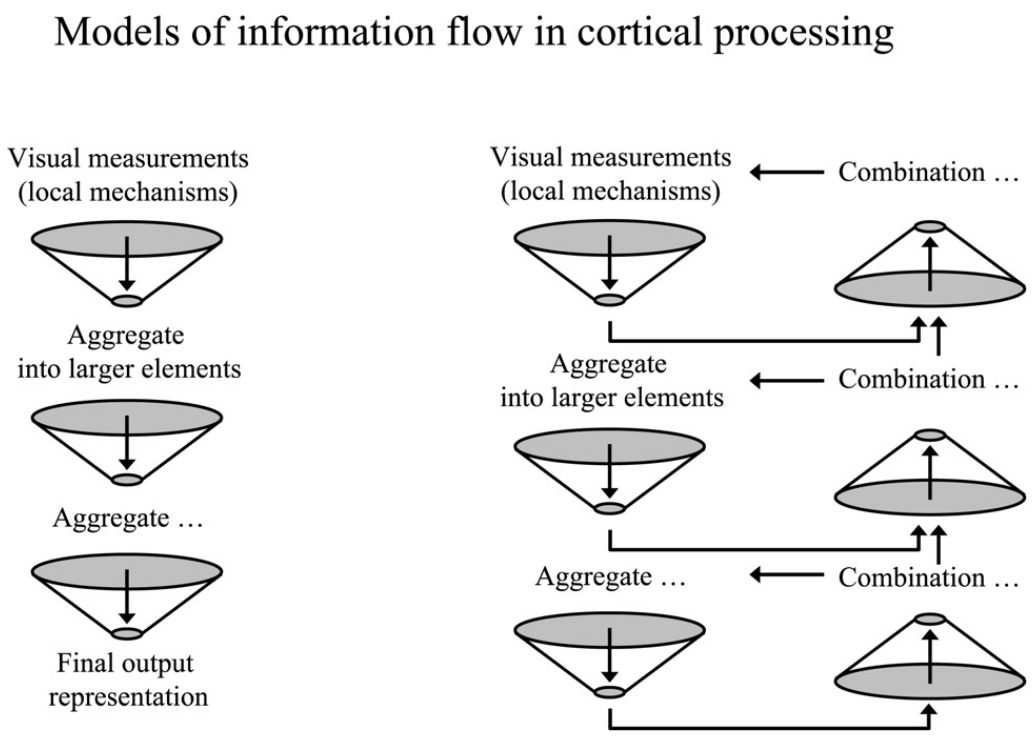



Fig. 13. Major principles of information flow in cortical processing. The primary processing is organized along hierarchical feedforward pathways in which local processing results are aggregated into larger items of increasing configural complexity. The aggregation is achieved by convergent (weighted) integration or selection over a larger neighborhood in spatial or feature domains. In the superficial layers of cortical areas anisotropic lateral slow-conducting long-range connections exist that allow the coupling of cells responsive to similar features, e.g., orientation. Fast feedforward processing is augmented by feedback paths that allow counter-streaming activation from higher stages to be re-entered and combined at earlier areas. Receptive fields at different areas monotonically increase in size the more distal they are from the sensory level. Thus, the resulting processing can be considered as building hierarchies of information flow in reverse directions allowing to link gross coarse-scale neural representations (which consider a larger context) with localized small-scale measurements at the earliest sensory-driven levels.

- lateral long-range integration (as in the superficial layers 2/3 of cortical areas) that feed recurrent loops of local interaction, and
- feedback from areas at higher-level stages of cortical processing.

Fig. 13 shows a sketch of some suggested principles of cortical architecture. An early stage integrates information on a temporal scale of approximately 80 ms; this occurs after onset of a presented stimulus. Next there are additional processes acting on a temporal scale of 120 up to 200 ms after stimulus presentation. These separate, but closely intertwined, phases of processing led Lamme and Roelfsema [44] to propose two phases (or sweeps) of processing, namely feedforward and feedback. This enables cortical neurons not only to code certain stimulus features but also to contribute to other analyses of the configuration and at different temporal moments.

In the following subsections we will focus on two related issues and their neural computational mechanisms. First, we outline the key elements of long-range grouping mechanisms for boundary integration, related empirical evidence and corresponding neural computations to implement the functionality of context-dependent feature measurement and evaluation. Second, we discuss the computations necessary for surface boundary formation and border-ownership assignment.

### *3.2. Evidence for computational long-range grouping mechanisms*

The investigation of grouping mechanisms for boundary formation has been mainly influenced by Gestalt principles of *proximity* and *good continuation* for perceptual organization [41], whereby those elements in a spatial arrangement tend to group when they are placed closer together (other features being equal). Also, for oriented items there is a tendency that groupings are preferentially formed when the items are arranged along smooth outline contours (Fig. 14). The formation of so-called *illusory contours* has also added evidence of perceptual formation of visual boundaries

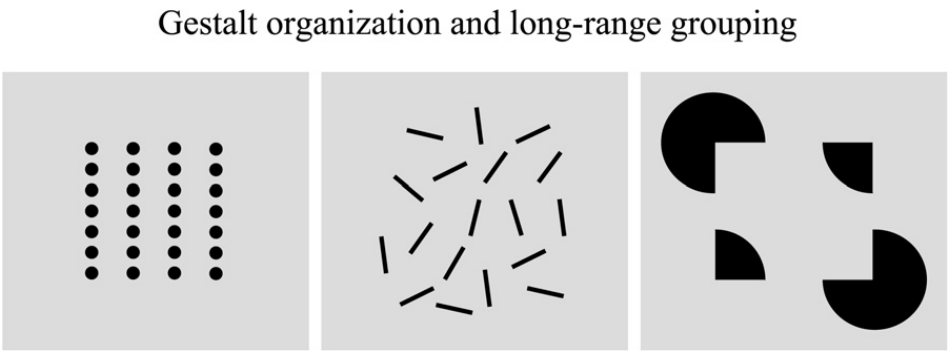


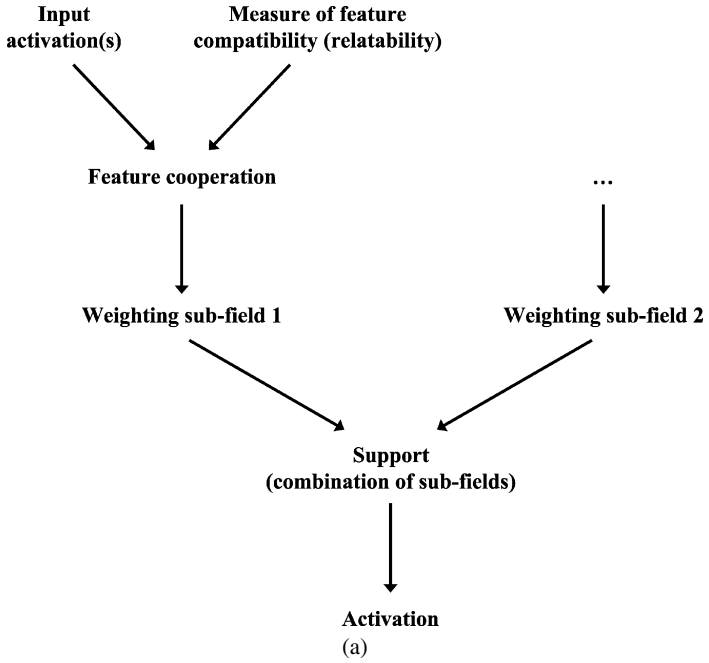

Fig. 14. Gestalt organization and long-range grouping. Local image items are spontaneously grouped into coherent perceptual units. *Left image*: In the arrangement of black dots those elements that are proximal are spontaneously grouped into columns. *Center image*: The configuration of oriented bar items leads to spontaneous grouping into a smooth continuous curve. *Right image*: Local contrasts of similar orientation but of opposite polarity and different figural context can be grouped into boundary segments bridging gaps without local luminance contrasts thus generating illusory contours.

[45]. Their investigation particularly helped to reveal the feature input arrangements that are necessary to generate perceptual contours even without luminance contrasts present in throughout the boundary. A study by Lesher and Mingolla [46] demonstrated that such boundaries vary in their perceptual strength depending on, e.g., the density and contrast of their inducers, and cannot be merely attributed to an all-or-none process of their generation. The examination of the underlying mechanisms for such grouping has received increasing attention since Field, Hayes and Hess [47] popularized the notion of an association field, a figure-eight shaped zone, along which facilitatory perceptual interactions with other perceptual boundary segments tend to occur. This figure-eight shape resembles a spatial field of integration that is composed of bipolar sub-fields (called 'bipole' in the following) to integrate responses of oriented cells in the spatial neighborhood as suggested in earlier proposals for modeling contour grouping [21]. This basic mechanism was subsequently proposed as a model of horizontal long-range integration in visual cortical area V1 and V2 [48].

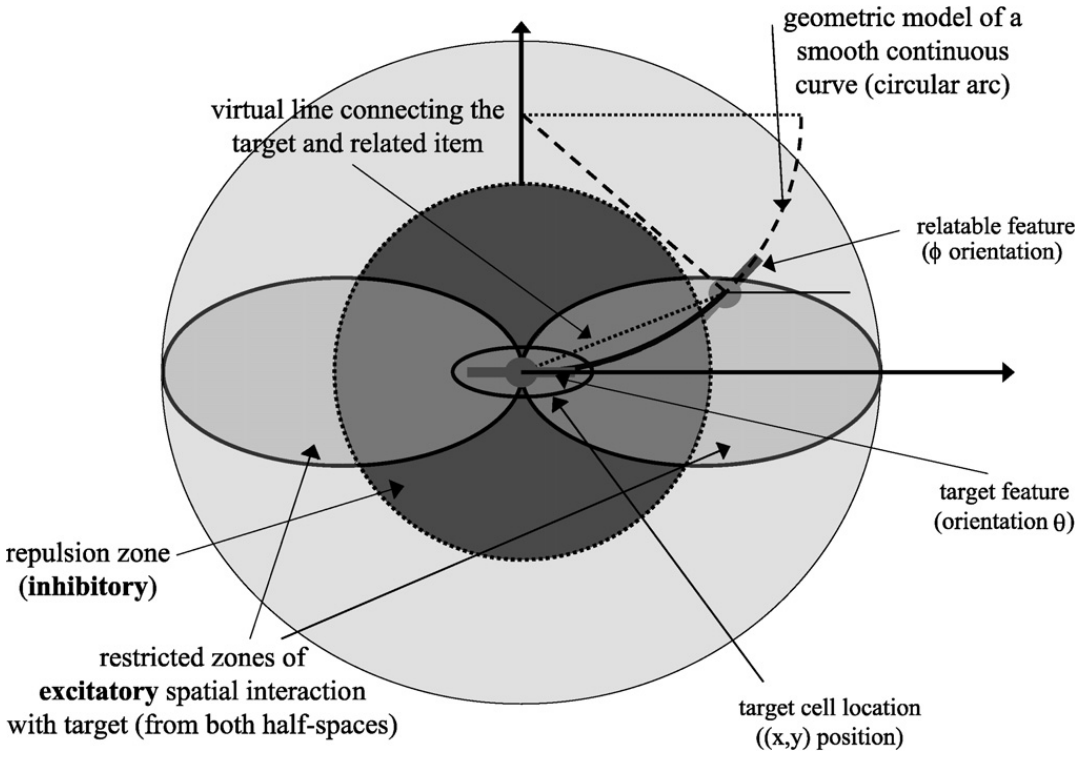
The neural mechanism of grouping local oriented contrasts can be linked to orientation selective cells in the primary visual cortex, namely area V1. In particular, the strikingly systematic columnar organization of different cortical areas and the laminar specificities suggest that local facilitation might be realized by long-range axonal connections in layers 2/3 in area V1 [49]. Cells in layers 2/3 preferentially contact cells via axonal horizontal connections along the axis of orientation preference that share the same orientation preference as the target cell. More recent investigations by Kapadia, Westheimer and Gilbert [50] arrived at a bipolar integration field from experiments in which observers were asked to adjust the orientation of a test line in the presence of flanking lines. The results revealed zones of attraction, viz regions of mutually facilitatory interaction, oriented co-axially along the target cell's orientation. Zones of repulsion, i.e. regions of mutually inhibitory interaction, exist in regions orthogonal to the axis of preference, and both zones are in the range of 8 minutes of perceptual angle. Note however, that as explained in the caption of Fig. 15, at the mechanistic (as opposed to perceptual) level, the inhibitory zone of a neural "bipole" extends radially in all directions from a common center. The neural origin of generating illusory contour responses might be attributed to processes at higher-order stages in cortex, namely V2 and beyond, which also integrate responses from co-axially arranged sub-fields of the local surrounding. Strong evidence for a role of V2 contour cells being involved in this process has been provided by physiological investigations by von der Heydt and collaborators [18,51]. The knowledge about the cortical areas involved in illusory contour processing was recently updated by a study by Mendola, Dale, Fischl, Liu and Tootell [52] using functional imaging methods showing that cells in area V2 but also in area V4 and beyond contribute to such boundary formation processes.

### *3.3. Neurocomputational mechanisms*

Based on the empirical evidence discussed above several neural computational models for feature integration and grouping have been developed in the past two decades. Many models share several key underlying computational mechanisms used to account for experimental data. We will start describing the basic component mechanisms of spatial long-range integration for grouping and boundary formation. Based on this outline, we sketch the computational elements of generating cell responses in response to figural outlines and surface boundaries.




(a)




(b)

Fig. 15. Spatial feature integration: Taxonomy and mechanism. (a) The mechanisms of feature integration can be organized into a tree structure of contributions or structural elements. The spatial weighting defines the structure of lateral spatial connectivity, or kernel, which can be subdivided into field elements. The feature compatibility function defines the representation of relatable visual structure, or context, which encodes the model of the expected most likely visual arrangement in the environment. The product of spatial weighting and compatibility function defines the total coupling strength for support local measures in the space-feature domain that were represented as activities, e.g., mean firing rate. (b) Models of long-range feature integration can be depicted by the bipole icon for modeling spatio-temporal grouping. The bipole zone of figure-eight shape represents the region of support denoting the contribution of activation in an oriented unit (light horizontally aligned ellipses) to the activation of another unit, denoted by the small ellipse in the center. The figure-eight shape of a bipole expresses relations among three fundamental quantities, namely (i) the distance between the centers of the two ellipses, (ii) the angle formed by a ray passing through the centers of the two locations indicated by the oriented gray bars (principle axes) and their orientations, and (iii) the difference in orientation of the principle axes. The measure of relatability can be expressed by a model of, e.g., a circular outline (see text for details). The mutually excitatory interactions are balanced by a shorter-range zone of inhibitory interaction (dark gray region that extends radially from the center and overlaps with the prolate excitatory zones).

*3.3.1. Outline of a framework of long-range grouping*

Grouping mechanisms invoke principles of selective integration of more localized measurements along any of several possible feature dimensions. To simplify the description we consider here features such as 2D image positions and orientation (in the range of 360 degrees). Spatial integration mechanisms, therefore, define the support of a given target activation in space-orientation domain, $\langle x, y, \theta \rangle$, to determine the new activation at the target location afterwards.[3] While the present formula for integration considers the inward flow of network activity from the viewpoint of the activation where perceptual completion should occur, this process could also be formulated from the perspective of outward flow, or voting, from regions with "bottom-up" contour information to support the assertion of a perceptual contour at a place that does not have adequate "bottom-up" support. In this view a cell casts votes to support the likelihood of cells' responses at different locations in the space-feature domain.

A *spatial neighborhood* and *weighting*, or kernel, function defines the range of influence of other elements in the vicinity of the target element and the (excitatory) weighting[4] (see Fig. 15, left). In the case of the long-range grouping mechanism, which utilizes the bipole structure, a spatial neighborhood of excitatory inputs extends along the preferred axis of orientation and ignores locations orthogonal to the feature orientation. This spatial region leads to the figure-eight, or butterfly, shape of excitatory influence (Fig. 15, right). Furthermore, depending on the logic of the bipole integration, the neighborhood function can be decomposed into sub-fields. The result of integration of items from different sub-fields needs to be combined to compute the *support* for the target activation. In case of two separate sub-fields, as in the bipole, each oriented along the preferred target orientation, the support is determined by

$$\text{support}_{\mathbf{x},\text{feature}} = \text{input}^{\text{left}}_{\mathbf{x},\text{feature}} \circ \text{input}^{\text{right}}_{\mathbf{x},\text{feature}} \tag{3.1}$$

where '$\circ$' denotes some operation of combining the sub-fields. This allows implementing simple linear as well as non-linear integration rules such as an AND-like combination of sub-fields for illusory contour formation, which fire only if flankers at both ends provide an input. Subscript '$\mathbf{x}$' denotes spatial locations in a $(x, y)$ 2D retinotopic map and 'feature' denotes, e.g., orientation, to identify the feature dimension involved in the grouping process. In the bipole model the segregated lobes of the integration field define the support of a target item. The resulting *activation* of the target cell is computed as a function of the support such that

$$\text{activation}_{\mathbf{x},\text{feature}} = f(\text{support}_{\mathbf{x},\text{feature}}). \tag{3.2}$$

The function $f(\bullet)$ might be, for example, the identity or a compressive non-linearity (e.g., sigmoid). The formal description of the mechanisms underlying the computation of activations from 'left' and 'right' sub-fields, respectively, necessitates the detailed specification of the interaction of activities in grouping. A spatial weighting, or kernel, function specifies the mutual influence between two localized features. Items that are closely spaced are in general more likely to be grouped than candidates that are located far apart. The underlying spatial neighborhood function often selectively facilitates a sector of a spatial surround to define an anisotropic coupling that is compatible with the feature domain. Elementary features along dimensions such as, e.g., (tangential) orientation, motion direction, and disparity provide the dimensions of the visual representation space. The feature *relatability*, or compatibility (see [53] and [54]) between stimulus items is defined along these dimensions. Herein, the most typical or likely appearance of meaningful structure is somehow encoded to represent "what feature goes with what". The resulting model representation is encoded in the spatial weights of the connectivity pattern of a target cell with its surround and thus defines a spatial filter (cf. [55]).

In most models the connectivity pattern of the relatable features is pre-specified, or programmed. These are encoded in closed-form mechanisms that define a coupling between tuples, e.g., pairs, of feature measurements at given locations. In this sense, they are designed a priori. To date, only few approaches investigate the possible self-organization of such lateral interactions in a neural architecture (cf. [56]). Note that the spatial weights and the feature relatability define the net coupling strength. This separable function specifies a metric for the similarity measure in the $\langle x, \text{feature} \rangle$

[3] The focus on discussing grouping within the orientation domain is no restriction of the basic concept. For example, grouping in the motion domain can be formulated in a four-dimensional space-velocity domain, $\langle x, y, u, v \rangle$ which requires specification of mutually cooperative interactions in space and between direction and speed measures.

[4] The scheme of mutual interaction of the cell at the target location and its surround is composed of excitatory as well as inhibitory contributions (see [50]). Here, the facilitatory grouping mechanism is considered only, which directly influences an on-center mechanism. We suggest the surround inhibition is mediated by a separate process in formulating the cell response properties.

space and thus defines the distance function for the clustering of a visual pattern to favor relatable items. This measure contributes to the definition of feature cooperation. Two different types of interactions can be distinguished:

(1) a convergent feedforward mechanism, when the bottom–up input is integrated at the target location, whereas
(2) a mechanism of (non-linear) lateral, or recurrent, interaction when activity is horizontally integrated within a neural layer.

The support of an oriented feature at a particular location is proportional to the activity of the feature integration process using the bipole mechanism (cf. Eq. (3.1)):

$$\text{support}_{\mathbf{x},\theta} = \sum_{\mathbf{x}'\phi} \text{act}_{\mathbf{x}'\phi} \cdot \text{relate}_{\mathbf{x}\mathbf{x}'\theta\phi} \cdot \text{weight}^{\text{left}}_{\mathbf{x}\mathbf{x}'\theta} \circ \sum_{\mathbf{x}'\phi} \text{act}_{\mathbf{x}'\phi} \cdot \text{relate}_{\mathbf{x}\mathbf{x}'\theta\phi} \cdot \text{weight}^{\text{right}}_{\mathbf{x}\mathbf{x}'\theta} \tag{3.3}$$

where 'weight' denotes the spatial weighting kernel, 'relate' the feature relatability, 'act' the input activations, and '$\circ$' denotes a combination of terms (Fig. 15, left). Coordinates in the space-feature domain are denoted by bold Latin letters for spatial locations, $\mathbf{x} = (x, y)$, and by Greek letters for orientation features, $\theta$. Other parameters refer to the specific location $(\mathbf{x}', \phi)$ in the space-orientation neighborhood (see Fig. 15, right). Models vary in the definition of these components.

Consider two tangent orientations, $\theta$ and $\phi$, at locations $\mathbf{x}$ and $\mathbf{x}'$, respectively, which can be connected by a smooth circular outline connecting $\mathbf{x}$ and $\mathbf{x}'$.[5] The parameters which define the connecting circular arc are given by the circle radius, $r = l/(2 \cdot \sin\theta)$, and the arc length connecting the two locations, $s = \theta \cdot l/\sin\theta$ with $l$ denoting the metric distance between the two locations $\mathbf{x}$ and $\mathbf{x}'$ (Fig. 15, right). Following the proposed scheme that separates spatial weightings from feature interactions (relatability), components for a weighting along a gauge orientation $\theta$ can be defined by

$$\begin{aligned} w^{\text{rad}}_{\mathbf{x}\mathbf{x}'} &= \exp\bigl(-\bigl(D(\mathbf{x},\mathbf{x}') - \mu\bigr)^2/(2\sigma^2)\bigr) \\ w^{\text{rad}}_{\mathbf{x}\mathbf{x}'} &= 1 + \cos^q\bigl(\bigl|\alpha(\mathbf{x},\mathbf{x}') - \theta\bigr|\bigr) \end{aligned} \tag{3.4}$$

with $D(\cdot) = \|\mathbf{x} - \mathbf{x}'\|$ and $\mu$ denoting the spatial offset of the Gaussian with respect to the target (center) location. The function $(\cdot)$ computes the direction of the virtual line between two locations $\mathbf{x}$ and $\mathbf{x}'$, $\alpha = \tan^{-1}((y - y')/(x - x'))$. The parameter $p$ in the second equation steers the sharpness of the angular weighting. The combined weighting is then defined by

$$\text{weight}^{\text{left/right}}_{\mathbf{x}\mathbf{x}'\theta} = \max\bigl(\pm \text{w}^{\text{ang}}_{\mathbf{x}\mathbf{x}'\theta} \cdot \text{w}^{\text{rad}}_{\mathbf{x}\mathbf{x}'}, 0\bigr), \tag{3.5}$$

with the left/right distinction controlled by the sign of the angular weighting. The max operation implements a half-wave rectification to prevent negative neural responses.

The relatability, or compatibility, of two segregated features in orientation space is encoded by a function $\text{relate}_{\mathbf{x}\mathbf{x}'\theta\phi}$. The geometric relation models the most likely geometrical shape in the visual world. This representation is sampled at discrete locations. Maximal relatability occurs at orientations tangential to the model curvilinear segment at the sample locations. Shape segments such as straight lines, circular arcs, splines, or parabolic curves are often used. Given two oriented items, one at the target location $\mathbf{x}$ and the other at a location $\mathbf{x}'$ in the vicinity, smooth continuations can be defined using differential geometry [57]. For example, a relatability function for items that on a circular arc,

$$\text{relate}^{\text{circ}}_{\mathbf{x}\mathbf{x}'\theta} = \cos^q\bigl(2\alpha(\mathbf{x},\mathbf{x}') - (\theta + \phi)\bigr), \tag{3.6}$$

with $q$ defining the sharpness of the fall-off for deviation in orientation. In order to fulfill the smooth relatability constraint of Kellman and Shipley [54], the angular values are constrained by the relations $\theta < \alpha(\mathbf{x},\mathbf{x}') < \phi$ (or $\theta > \alpha(\mathbf{x},\mathbf{x}') > \phi$) and $|\theta - \phi| \leqslant 90°$.

[5] The smooth tangential continuation by circular arcs leads to a compatibility measure that has been called a measure of co-circularity in [53].

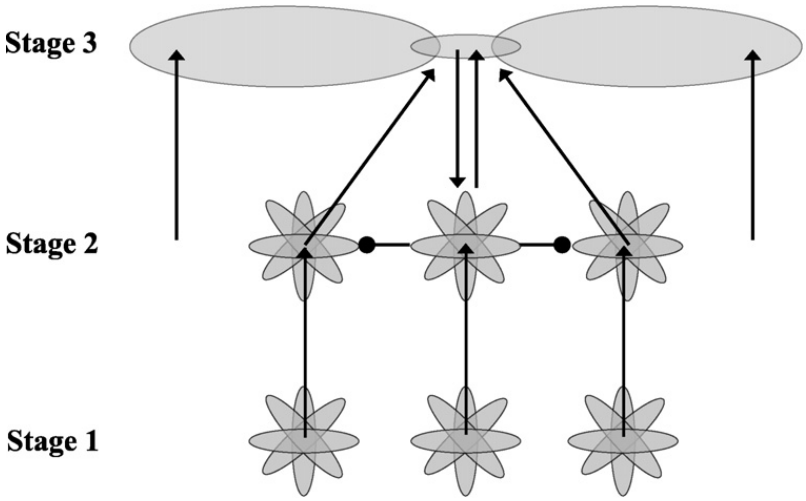



Fig. 16. Model of short-range competitive and long-range cooperative computation. Sketch of a boundary grouping model, called boundary contour system (BCS) and the macro-circuits therein. The three main levels of processing in a hierarchy are sketched. Stage 1 consists of oriented contrast-selective filtering (simple cells) of pre-processed luminance functions. Stage 2 depicts orientation sensitive filtering but polarity insensitive processing (complex cells) with lateral competition in space-orientation domain. These cells receive feedback from stage 3 long-range cooperative cells. These cells integrate input from orientation sensitive cells of co-axially aligned position and corresponding orientations. Together these two stages form a competitive-cooperative (CC) loop of recurrent boundary processing (see text for details).

### *3.3.2. Neural modeling of long-range grouping for boundary formation*

The components of feature integration mechanisms sketched above need to be combined with the neurodynamical mechanisms to generate cell responses of grouping. A larger modeling framework called FACADE (Form-And-Color-And-DEpth) theory was developed by Grossberg and coworkers as an attempt to explain a wealth of perceptual and cognitive phenomena [58,59]. Within FACADE, two sub-systems were proposed which essentially provide the mechanisms to create perceptual surface outlines and grouping (the Boundary Contour System, or BCS; [21]), while invariant surface properties are perceptually filled-in and completed afterwards (the Feature Contour System, or FCS; [31]). More recently, these basic principles have been mapped to basic laminar structures of visual cortical areas to explain interactive effects of feature extraction, perceptual grouping, learning, attention mechanisms which utilize the same cortical circuits [60,61]. This modeling framework was named LAMINART to highlight the laminar cortical structures and principles of adaptive resonance theory, or ART [55].

The focus in this section is on grouping mechanisms, which utilize long-range integration of relatable contrasts. The necessary network computations include lateral excitation and the space-feature inhibition. For the sake of readability, we decompose the description of computations for grouping into different stages, which are derived from original work reported in Mingolla, Ross and Grossberg [62], Neumann and Sepp [63] and Grossberg and Raizada [22] (for an overview, see [23]), Fig. 16. Their functionalities are as follows:

- *Stage* 1: Input to the system is generated over a series of individual sub-stages, which can be summarized as a (non-linear) filtering of the luminance function. Contrast enhancement from on-center/off-surround (ON channel) and off-center/on-surround (OFF channel) interactions is generated sub-cortically at the retina and LGN in the thalamus. The half-wave rectified responses of these ON and OFF cells compensate for variable illumination by computing locally normalized contrast ratios throughout the image. The ON and OFF cell responses together drive the activation of oriented simple cells at a level that is associated to primary visual cortex. Simple cells compute the local image gradient magnitude and orientation, i.e. measure change in the luminance along different orientations. Next, similarly oriented simple cells sensitive to opposite contrast polarities (dark-to-light, or light-to-dark) pool their activations at complex cells that detect contrast regardless of its direction.
- *Stage* 2: Complex cell activations compete through on-center/off-surround processing across the feature dimensions considered, namely image locations and orientation. These competitions capture the functionality of lateral inhibition across a cortical map in which nearby cells tend to be sensitive both to contrasts at neighboring image locations and similar boundary orientations, competition sharpens boundary localization and orientations. It also helps detecting the ends of a line. Competition also normalizes the overall responses within a neighborhood. The

stage 2 dynamics of interacting neurons, with excitatory and inhibitory inputs results in

$$\tau^{\mathrm{ex}}\frac{\mathrm{d}}{\mathrm{d}t}u_{\mathbf{x}\theta} = -A\cdot u_{\mathbf{x}\theta} + B\cdot \mathrm{input}_{\mathbf{x}\theta} - C\cdot u_{\mathbf{x}\theta}\cdot v_{\mathbf{x}\theta}$$
$$\tau^{\mathrm{in}}\frac{\mathrm{d}}{\mathrm{d}t}v_{\mathbf{x}\theta} = -v_{\mathbf{x}\theta} + D\cdot\sum_{\mathbf{x}'\phi}\mathrm{input}_{\mathbf{x}'\phi}\cdot\Lambda^{-}_{\mathbf{x}\mathbf{x}'\theta\phi} \tag{3.7}$$

where $\Lambda$ denotes a weighting kernel, such as a Gaussian, $\tau^{\mathrm{ex}}$, $\tau^{\mathrm{in}}$ denote time constants of the excitatory and inhibitory cells, respectively, which are related to the circuit property of the model membrane; A, B, C, and D are constants. (See Appendix A for a brief introduction into the modeling and notation.) Note that excitatory and inhibitory groups of cells are distinguished.[6] Here, when the inhibitory activation $v$ quickly equilibrates, then the (excitatory) cell activity can be expressed by one equation lumping together the two components in Eq. (3.7)

$$\frac{\mathrm{d}}{\mathrm{d}t}u_{\mathbf{x}\theta} = -A^*\cdot u_{\mathbf{x}\theta} + B^*\cdot\mathrm{input}_{\mathbf{x}\theta} - C^*\cdot u_{\mathbf{x}\theta}\cdot\sum_{\mathbf{x}'\phi}\mathrm{input}_{\mathbf{x}'\phi}\cdot\Lambda^{-}_{\mathbf{x}\mathbf{x}'\theta\phi} \tag{3.8}$$

where the constants are modified for the combined equations. In order to better understand the computational properties of Eq. (3.8), we display the equilibrium response[7]

$$u_{\mathbf{x}\theta,\infty} = \frac{B^*\cdot\mathrm{input}_{\mathbf{x}\theta}}{A^* + C^*\cdot\sum_{\mathbf{x}'\phi}\mathrm{input}_{\mathbf{x}'\phi}\cdot\Lambda^{-}_{\mathbf{x}\mathbf{x}'\theta\phi}}. \tag{3.9}$$

This interaction implements an inhibition (in the denominator) such that the activation at a target location is normalized against the sum of weighted input activities from a local neighborhood. This inhibition is generated in Eq.(3.8) by weighting, or *shunting*, the summed input. Shunting inhibition has been proposed by Sperling [64] and Grossberg [65] to account for non-linear effects in stimulus processing, namely activity normalization.[8] In such a network shunting inhibition is utilized as a key building block to achieve contrast enhancement. Later Grossberg and Mingolla [21,66] employed this normalization property for orientation-sensitive neurons in their Boundary Contour System model. More recently, it has been proposed for enhancing responses of oriented cells (e.g., [67]). Such an operation guarantees that activities are kept within bounds and increased input levels lead to saturation of individual activities but at the same time keeping the overall network activation (in the considered pool) sensitive to input changes.

- *Stage* 3: Normalized activities feed into long-range cooperation to generate consistent surface properties by completing fragmented boundaries even in cases of camouflaged object depiction. This cooperation is achieved by bipole cells as sketched above (see Eq. (3.1)). A simple integration mechanism sums activities from relatable configurations. This computation is reminiscent of lateral long-range integration in visual cortical area V1. A mechanism that accounts for illusory contour completion only fires if both halves of their receptive fields are sufficiently activated by appropriately oriented input contrasts from complex cells, thus resembling a type of statistical AND gate.

  A simplified version of boundary integration which sums input from both sub-fields (instead of ANDing input from the two lobes; [22]) leads to the following equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}z_{\mathbf{x}\theta} = -A\cdot z_{\mathbf{x}\theta} + (B - z_{\mathbf{x}\theta})\cdot\left[C\cdot\max(u_{\mathbf{x}\theta},0) + \mathrm{support}_{\mathbf{x}\theta}\right] - (D + z_{\mathbf{x}\theta})\cdot\sum_{\phi} w_{\mathbf{x}\phi}\cdot\Psi_{\theta\phi}$$

[6] It should be noted that other models integrate activities laterally from the excitatory layer for generating inhibitory action, thus, denoting a stage of recurrent inhibition. In this case the dynamics is more complex and might also predict different results. It is beyond the scope of this overview to discuss all the details of different model approaches and thus refer the reader to the original publications.

[7] In the case of more complex computations and interactions which do not allow lumping together the two separate equations in (3.8) an analysis of equilibrium points in state space based on Eq. (3.7) and their stability is necessary. A detailed discussion on such issues is beyond the scope of this overview. The interested reader will find proper introductions into the problem in overviews and textbooks on neural networks and computational neuroscience, such as, e.g., Grossberg [22].

[8] Sperling [64] proposed a retinal on-center/off-surround interaction with divisive inhibition generated by the surround to account for adaptation and Weber law effects at different illumination levels. Grossberg [65] investigated non-linear networks with recurrent on-center off-surround interactions to study how activity patterns could be dynamically stored (short-term memory).

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{\mathbf{x}\theta} = -w_{\mathbf{x}\theta} + \text{support}_{\mathbf{x}\theta} - w_{\mathbf{x}\theta} \sum_{\phi} w_{\mathbf{x}\phi} \cdot \Psi_{\theta\phi} \tag{3.10}$$

where the input activity of the mutual support is defined by the (thresholded) activities of the integration layer, $\text{act}_{\mathbf{x}'\phi} = \max(z_{\mathbf{x}'\phi} - T, 0)$ (see Eq. (3.3)), and the operator for sub-field integration, where '$\circ$' = '$+$' in this case. In the first equation, a cell at the target location is excited by feeding activity from the previous stage complex cell responses. This feeding input is supported by activities of cells of the same orientation at co-axial locations within the integration field of the bipole. The second equation balances the total energy of activation from long-range integration in the same layer.

The processing along Stages 1–2–3 is driven in a feedforward fashion. The output of Stage 3 enhances salient contrast configurations for delineation of surfaces. This enhancement can be achieved via lateral *recurrent* interaction of grouping cells (connecting cells by long-range horizontal connections) or via intra-areal excitatory *feedback* of grouping activities back to Stage 2 cells enhancing their on-center activations. In addition, feedback from higher areas can also be "retro-injected". As discussed shunting inhibition normalizes an individual response. Now, an excitatory feedback which enhances the target activity at the center location in turn leads to inhibition of the other activities in the surround. Thus, this feedback simultaneously enhances contrast and suppresses noise. The grouping of co-aligned contour segments preserves boundary strength while avoiding saturation.

Two different mechanisms were suggested for the combination of feeding input and enhancement through feedback activation. Grossberg and Raizada [22] used an *additive combination* of signal streams. Here, feedback signal activation undergoes separate shunting competition (at layer VI of model area V1) such that the strength of the additive feedback stays within bounds. The driving feedforward signal and the normalized feedback combine additively at layer IV and undergo a shunting inhibition of the type (3.8). This guarantees that the feedback does not lead to uncontrolled activity generations which otherwise causes "hallucinations" of visual structure. An alternative utilizes a signal transformation with saturating property, such as a sigmoid. *Modulatory feedback* has been suggested on the basis of empirical findings to selectively enhance feeding inputs in a multiplicative fashion, thus, implementing a soft-gain control mechanism [63,68,69]. The mechanism must allow (i) the feeding input to survive in case no feedback is provided, (ii) in a situation where feedback is generated but feeding input is non-existing no activity should be generated, and, finally, when (iii) both input and feedback are available the input signal level should be enhanced (see [70] for a summary). This can be achieved by

$$\text{input}_{\mathbf{x}\theta} \cdot (1 + E \cdot z_{\mathbf{x}\theta}), \tag{3.11}$$

which allows direct enhancement of feeding input via matching top-down feedback encoded in $z$ activation. This enhancement could be included in a stage prior to the shunting inhibition in Eq. (3.8) such that the modulatory feedback is cascaded with a normalization stage. We now get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} s_{\mathbf{x}\theta} &= -A \cdot s_{\mathbf{x}\theta} + B \cdot \text{input}_{\mathbf{x}\theta} \cdot (1 + E \cdot z_{\mathbf{x}\theta}) \\ \frac{\mathrm{d}}{\mathrm{d}t} u_{\mathbf{x}\theta} &= -A \cdot u_{\mathbf{x}\theta} + s_{\mathbf{x}\theta} - u_{\mathbf{x}\theta} \cdot \sum_{\mathbf{x}'\phi} s_{\mathbf{x}'\phi} \cdot \bar{\Lambda}_{\mathbf{x}\mathbf{x}'\theta\phi} \end{aligned} \tag{3.12}$$

The overall functionality leads to a "biased competition" as observed in recent investigations of selective attention [71]. In a nutshell the enhancement of a particular activity in turn increases the inhibitory power in the subsequent competition stage. A bias is created that favors the enhanced activity to more strongly suppress the other activities in the local pool.

### *3.3.3. Results of computational studies*

Simulation results of these mechanisms highlight the functionality of the networks and their account for perceptual phenomena as well as dealing with complex images of real world objects. The first example in Fig. 17 shows the effects of grouping and illusory boundary formation in the Kanizsa figure and a variant in which the inducers are composed of concentric rings instead of solid circular patches (left and middle figure). The result of grouping for the middle figure is shown on the right demonstrating the demarcation of surface regions by boundaries due to luminance contrasts and illusory contours. The gray activities encode the strength of the activity as generated by the model

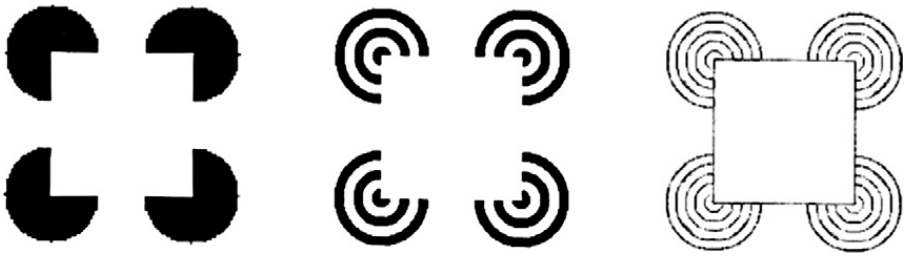


Fig. 17. Competitive-cooperative computation for boundary formation: Illusory contours. Two input luminance images one with four three-quarter black circles and an apparent white square disk in the center (Kanizsa square, left panel) and a variant in which the homogeneous circular regions were replaced by concentric circle lines of various radii (Varin figure, center panel). Simulation results for the Varin figure input where contours were generated along light–dark and dark–light contrasts as well as groupings to generate illusory boundaries bridging the gaps between circular line items and completing the fragmented outline of the central square (from S. Grossberg, E. Mingolla, W.D. Ross. Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neuroscience*, 20: 106–111, 1997, reprinted with permission).

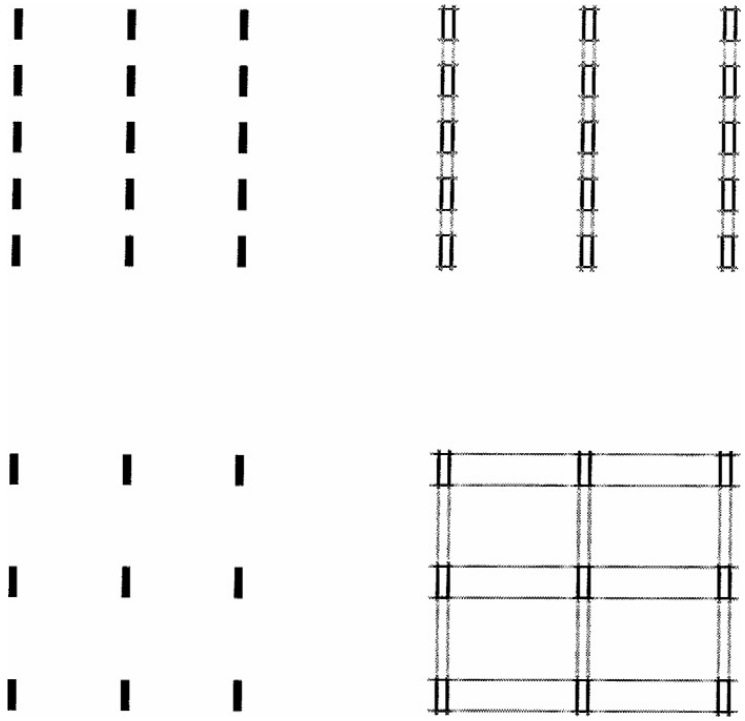


Fig. 18. Competitive-cooperative computation for boundary formation: Gestalt laws of continuity and proximity. The figure shows simulation results of grouping oriented bar items and the formation of figural (de-)compositions based on mechanisms that implement Gestalt laws of smooth continuation and element proximity. Two input luminance images consisting of horizontally oriented bars (top). In the left panel the center-to-center spacing of the bars is identical in horizontal and vertical orientation resulting an equally salient grouping, or segmentation, along the horizontal and the vertical directions. In the right panel the density has been doubled along the horizontal direction, which immediately yields a strong formation of horizontal fragmented line patterns. Simulation results show the same type of categorical switch in the generation of a web of boundary activations (bottom). In the left configuration both groupings lead to similar activations that survive local competitions and receive long-range support of approximately equal strength. The long-range groupings formed by long-range integration in the right configuration dominate weaker candidates from aligned line endings, which cannot properly establish (from S. Grossberg, E. Mingolla, W.D. Ross. Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neuroscience*, 20: 106–111, 1997, reprinted with permission).

simulation of Grossberg, Mingolla and Ross [48]. The second example in Fig. 18 shows the effects of grouping together figural items into configurations of boundary fragments depending on the items' parameters (orientation) and their spatial density and position. This exemplifies how some of the empirically defined laws of Gestalt organization are characterized on the circuit level of neural computation and how the balanced arrangement of neural activation is influenced by changes in the spatial configuration of figural inputs. The input configuration (top, left) is composed of equidistant items that are spatially separated horizontally and vertically by the same amount and, therefore, an ambiguous perception emerges which is simulated by the model of long-range integration and competition (bottom,

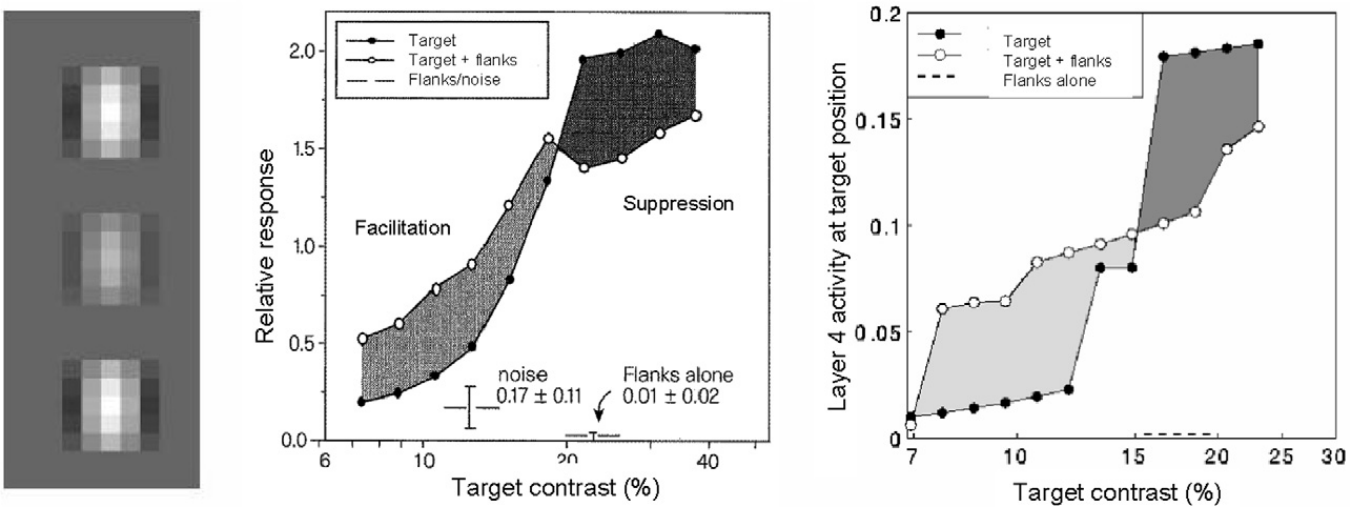



Fig. 19. Enhancement and suppression effects in the grouping of aligned items. The figure shows different effects in grouping of co-aligned items given two high-contrast flankers and a target in the middle having different contrast against the background (flankers and target items are generated by Gabor patches of equal orientation; left panel). Psychophysical investigations have shown a facilitatory effect, i.e. enhancement, when the target is a low contrast item. This could be interpreted such as long-range grouping with high-contrast items enhances the target as if it were of higher visual contrast or saliency. The situation changed when the contrast of the target patch approaches that of the flankers. In this case mutually competitive effects dominate the grouping such that the saliency is less vivid in the center (center panel). Model simulations show quantitatively the resulting activities of oriented cells of model cortical layer 4 (corresponding to stage 1 cells in Fig. 16; right panel). The same cross-over behavior is replicated by the model demonstrating how excitatory and mutually competitive interactions balance each other (from S. Grossberg, R.D.S. Raizada. Contrast-sensitive grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40: 1413–1432, 2000, reprinted with permission).

left). When the spatial input configuration is manipulated such that the horizontal item density is doubled (top, right) the appearance changes into an unambiguous percept of strong horizontal groupings. Since the integrated neural activity in the horizontal orientation is now much stronger it also strongly weakens the vertical candidate groupings for aligned bar endings. As a consequence these groupings were extinguished and only the horizontal groupings survive (bottom, right).

The third example in Fig. 19 illustrates the grouping of oriented items depending on their luminance contrast as investigated in a psychophysical study by Polat, Mizobe, Pettet, Kasamatsu and Norcia [72]. An oriented item (Gabor patch) of variable contrast is centered at a V1 cell's classical receptive field (CRF). Additional coaxially positioned and collinearly oriented flankers are positioned outside the CRF. These are fixed at high contrast (left). The neural responses measured at the target cell as a function of stimulus contrast were compared to the response when the central item is augmented by the two flankers (middle). The results demonstrate that the neural activity of the single item presented alone increases from weak to medium item contrasts and saturates for further contrast increases due to the effects of conductance-based (shunting) neural mechanisms. The strengths of enhancement change when two high-contrast flankers were added. For a low contrast situation the target cell's response is facilitated through item grouping while the interaction becomes more suppressive when the central item's contrast exceeds a certain level. This has been successfully replicated in a simulation of the model proposed by Grossberg and Raizada [22]. The reason for this "turnover" behavior lays in the non-linear excitation and inhibition properties at different stages of the layered architecture. In a nutshell, low-contrast feedforward activity is combined with feedback activations from long-range integration (generated from the integrated high-contrast flankers) and increases at a faster-than-linear rate. This increase is only slightly counteracted by lateral shunting inhibition, since the target contrast is low. The net response in the target-plus-flanker situation is thus increased in comparison to the target presented alone (Fig. 19, right, light gray zone). The inhibitory action is increased significantly when the target contrast is increased. At the same time the enhancement through the flanking items levels off for higher contrasts due to the saturation effect of the shunting mechanisms. As a net effect the neural activity at the center of the target-plus-flanker condition is lower than it would be for presentation of the high contrast target item alone (Fig. 19, right, dark gray zone).

The final example in Fig. 20 shows the generation of a representation of 3D intrinsic shape properties from images of shiny surfaces. In the mirror-like surface coatings the structures resulting from the different projected environments

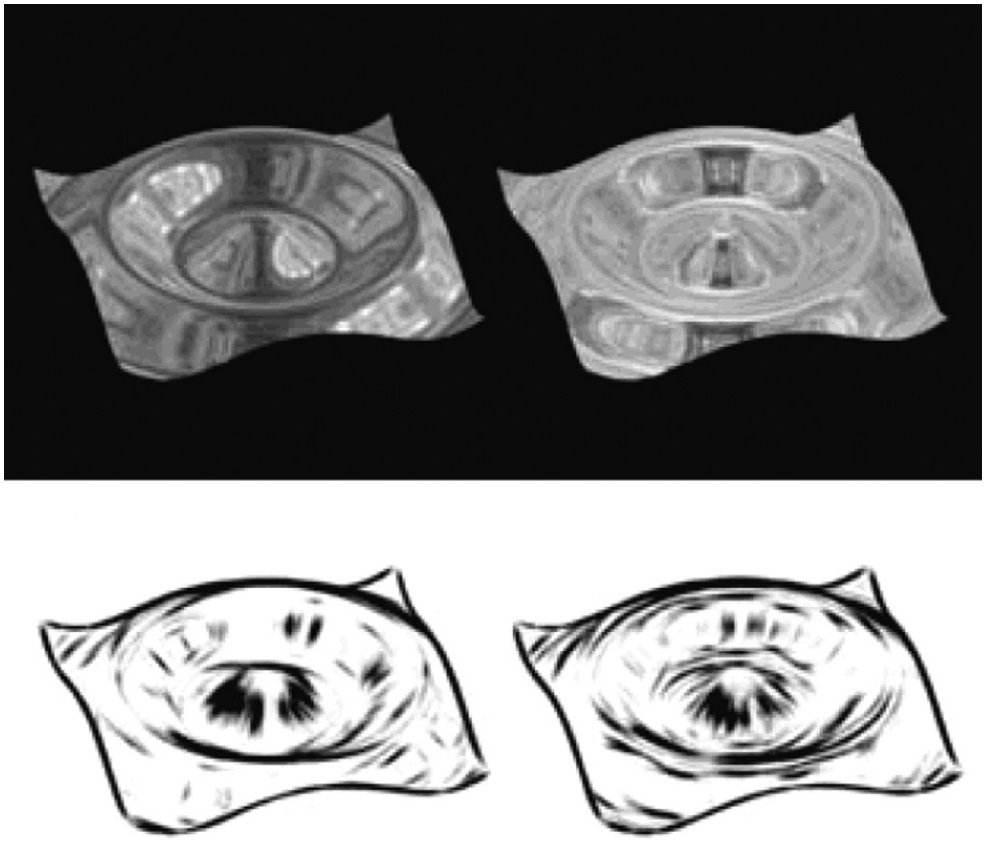


Fig. 20. Grouping and competition in the depiction of invariant properties of surface layout. The input luminance images depict two exemplars of a metallic 3D curved surface geometry with two different environmental maps projected onto the surface. Due to their local geometry the surface generates a deformation of the spectral content of the original environment texture with a bias along the direction of the lower principle curvature. If the absolute values of the two principal curvatures differ significantly (strong anisotropic surface curvature) the texture deformation leads to locally parallel oriented contrasts, whereas in patches with more or less isotropic surface curvature the texture is stretched or compressed equally likely along all directions. Model mechanisms of spatial competition and oriented long-range grouping enhance those regions of locally oriented contrast (corresponding to elliptic surface patches, mutual self-occlusions, and occluding boundaries) while mainly suppressing activities in regions with isotropic surface curvature. This leads to a sketch-like enhanced contrast pattern of activity highlighting the intrinsic surface geometry irrespective of the initial environment map and its resulting texture pattern (from U. Weidenbacher, P. Bayerl, H. Neumann, R. Fleming. Sketching shiny surfaces: 3D shape extraction and depiction of specular surfaces. *ACM Trans. on Applied Perception*, 3: 262–285, 2006, reprinted with permission).

are selectively deformed in accordance to the local differential geometry of the surface patches (top). For example, in a patch of weak and locally isotropic surface curvature the image of the projected pattern of environment structure appears without major deformations. In segments of strong anisotropy, however, a pattern of deformation appears in which the projected pattern is compressed along the direction of larger surface curvature while it is stretched along the direction of minimal curvature. The degree of anisotropy is encoded by the pattern of initial orientation selective contrast filtering (anisotropic bandpass filters) and, thus, generates a distributed representation of local qualitative surface geometry. The spatial arrangement of filter activations for different orientations is the input to the mechanism of oriented grouping. The model proposed by Weidenbacher, Bayerl, Neumann and Fleming [73] (which is based on the approach of Neumann and Sepp [63]) generates strong groupings for the anisotropic patterns while only weak groupings are generated for isotropic structures. The subsequent center-surround normalization via shunting competition further enhances the groupings for anisotropic patterns to signal salient features of smoothly curved surfaces (e.g., ridges, cusps, bounding contours, and self-occlusions; compare [74]) while the unspecific groupings are almost suppressed. This leads to results showing activation pattern almost invariant of the input luminance pattern (bottom).

### *3.4. Figural surface boundaries and border-ownership computation*

The goal of visual processing considered here seeks building neural representations of perceptual surfaces. We have discussed so far, how established neural mechanisms contribute to the formation of surface boundaries. In Section 4

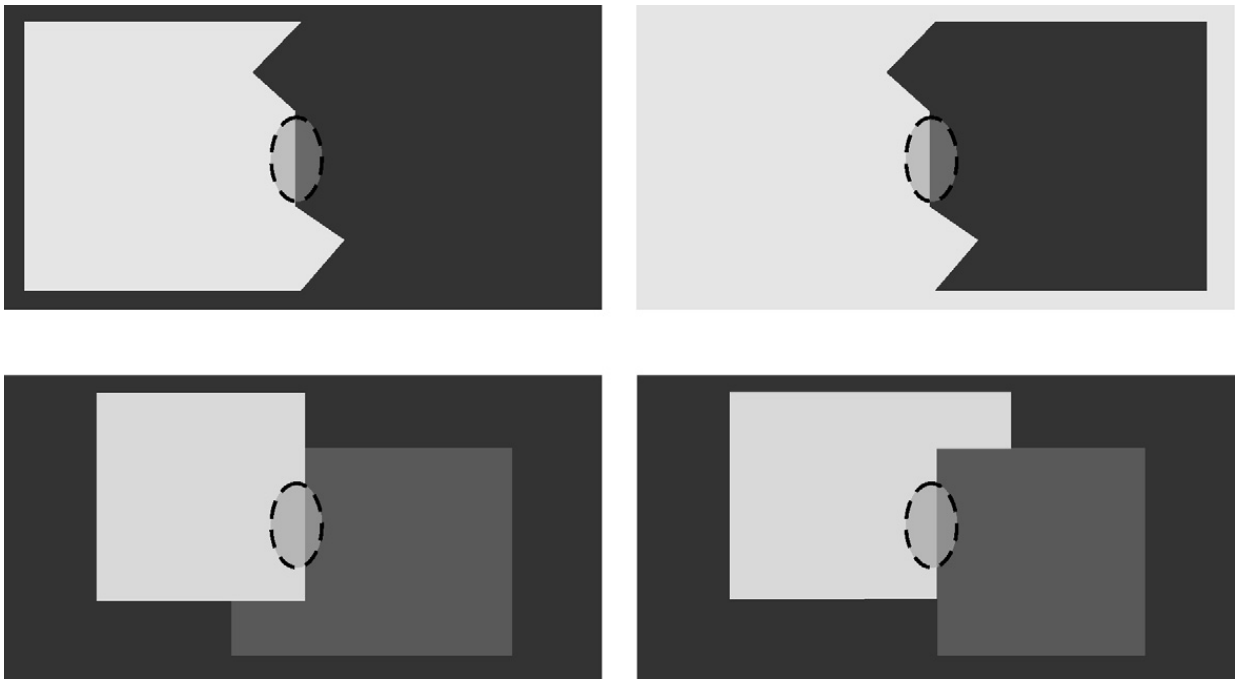


Fig. 21. Border-ownership: Local contrast patterns with different figure-ground direction. The mechanisms of local contrast detection and grouping into boundary fragments cannot locally decide on which side the belonging surface is located that generates the border contrast. This decision involves more global information by somehow tracing the contours, their arrangements, and the lightnesses of the surface regions to decide the surface and its border "ownership". The upper two panels show a light-dark contrast as part of a vertically oriented contour segment. The figural outline of the full contour far outside the contrast selective cell in the center (dashed ellipse) is different such that a bright figure is situated to the left in the left panel and a dark figure is situated to the right in the right panel. A similar situation is generated in the surface arrangement with mutual occluding patches depicted in the bottom panels. Here the bright rectangular surface appears in front partially occluding the medium gray patch and the dark background (left) whereas the gray patch appears in front partially occluding the light patch and the dark background (right).

we will discuss the interaction between boundaries and the representation of homogeneous surface layout, e.g. the computation of perceptual attributes like invariant lightness from luminance, or transparency. Before doing so, we will here consider another step that is prerequisite to the segmentation of surfaces and the segregation of figures from complex background, namely, the determination of the so-called *ownership* of a boundary by neighboring surfaces. Pictorially, while walking along a boundary border ownership (BO) assigns a label indicating which side (left or right) is the surface that generates, or owns, the perceptual boundary. This has also been coined *belongingness* in earlier literature and was already identified by the Gestaltists in the 1930s [41].[9] Recently, Zhou, Friedman and von der Heydt [76] measured cell responses in areas V1, V2 and V4 along the ventral pathway that is mainly concerned with the processing of static forms. Cells were probed with contrasts oriented along the axis of preferential cell selectivity and a given contrast polarity. While this local image property was kept constant the figural organization was manipulated in different stimuli such that the contrast at the target cell either belongs to a figural surface region to the left or to the right of the boundary. Fig. 21 shows few stimulus arrangements with a figural shape on a homogeneous background (top row) and configurations of two overlapping shapes in which several remote image cues signal occlusion boundaries that were owned by the surface of the shape that is in front (bottom row). The authors found that more than 50% of the cells investigated in areas V2 and V4 differentially modulated their response depending on the direction of ownership, while only 18% of cells in V1 showed this behavior. This was taken as direct evidence that cells in areas V2 and V4 explicitly code BO. Of those cells identified a majority was sensitive to local contrast polarity while the rest was insensitive. Interestingly, the response modulation occurred already in the rising phase of the response after stimulus onset and was virtually independent of shape size, which in turn was considered as argument against time-consuming iterative computations.

[9] In the original scientific literature the observation of border ownership was identified with the "Einseitigkeit der Grenzlinien". Indeed the "Gesetz der Innenseite" is one of the primary laws in Gestalt organization that are involved in the figural organization (see [75]).

Models have been proposed to account for these and earlier findings of segregating figure from ground. Several qualitative criteria can be identified that give hints for BO direction given a location on an oriented boundary segment. These criteria inspired many of the proposed neural model mechanisms. They are summarized as follows:

1. Real luminance contrasts as well as illusory boundaries are involved in the assignment of ownership; the visibility of illusory boundaries generated by abutted line ends in monocular images requires input from both sides of the boundary contour and, in addition, the inducer elements need to be extended as dots representing the endings will be grouped but fail to generate an illusory contour.
2. Traveling along the boundary of a closed figural shape the curvature directions and angles at junctions sum up to a value of $2\pi$.
3. A figural region that owns the boundary measured at a particular location mainly extends to the side of the ownership direction (this is a more qualitative expression of the sum of angles principle stated above).
4. T-junctions and X-junctions (as well as line endings that were aligned in one direction) signal potential spatial occlusions; in the case of T-junctions occlusions arise from opaque surface regions at the opposite side of the T-stem, while X-junctions often occur when transparent surfaces overlap each other (see Section 4).

Heitger, von der Heydt, Peterhans, Rosenthaler and Kübler [77] suggested that the formation of illusory contours by long-range integration of end-stop cell responses is directly involved in the assignment of figure-ground direction (see criterion 1 above). The mechanism of long-range integration is of the type depicted in Figs. 15 and 16, which relates to cells reported in area V2 [51]. These grouping cells were proposed to selectively integrate termination signals that are compatible in their direction of end-stopping. The signal strength which is proportional to the number of inducing line endings on either side of the boundary is taken as an indicator for the figure-ground, viz BO, direction. In Fig. 22 (left) an example case is shown that indicates the figure-ground direction along the illusory boundary of the central square (bottom). Since this approach requires separate sets of terminator signals this approach fails to account for configurations such as the variation of the Kanizsa pattern in Fig. 23 (top left). Here each inducer is defined by opposite end-stopping directions, which cannot be integrated by the proposed model (see the detailed discussion in [23]).

More recent models take several features into account and gather evidence for the presence of figural shape and, thus, BO direction. For example, Kikuchi and Akashi [79] propose a model of hierarchical processing oriented contrasts, their curvature groupings and explicit L-junction detection to assign BO tags to boundaries. The idea is based on the second criterion above that analyzes the geometric properties of closed contours. The propagation of signals along contours is considered as neural implementation of an angular integration of the boundary path that is further augmented by signals at sharp corners, or L-junctions. The model was tested for rather simple shapes only and it remains unclear whether the approach could also explain more complicated shapes tested by Zhou et al. and how it performs when localized cues such as corners were selectively eliminated (as in Fig. 23, top right). Several approaches use cells to measure oriented contrasts that feed into grouping cells to determine BO direction. Von der Heydt and coworkers [80,81] employed mutual competition of contrast cells and feedback from grouping BO cells to contrast cells to bias the competition. In a later version of the model this feedback mechanism was supplied by a mechanism of shunting inhibition at the competitive interaction. T-junctions were detected following criterion 4 above to break symmetries in the bottom-up signals from contrasts to grouping cells. Sakai and coworkers [82,83] propose a hierarchically organized competitive-cooperative model for BO assignment integrating oriented contrasts generated by orientation selective filtering and subsequent shunting inhibition. Unlike the previous models a competitive-cooperative mechanism of surround modulation is proposed to acquire the bias information for the detection of BO. In particular, the circular surround of a target cell is sampled by excitatory and inhibitory subfields which are (i) located asymmetrically along each side of the target cell's orientation axis and (ii) employ a larger subfield region for the inhibition (positioning and sizing of sub-field regions were tested in an exhaustive search and evaluation process). Excitatory input is gathered mainly from contrasts oriented orthogonally to those preferred at the target cell while inhibition is generated by like-oriented contrasts in the surround. This integration mainly follows criterion 3 to indicate the direction of the figural extent with respect to a local oriented contrast. Subfield responses are combined linearly. In order to account for the contrast selectivity of the majority of tested cells in V2 and V4 the authors combined the output of the surround integration (for facilitation) by gating those responses by the contrast response of a cell at the same location and with corresponding orientation selectivity. The gating mechanism utilizes a slightly different version of the gain

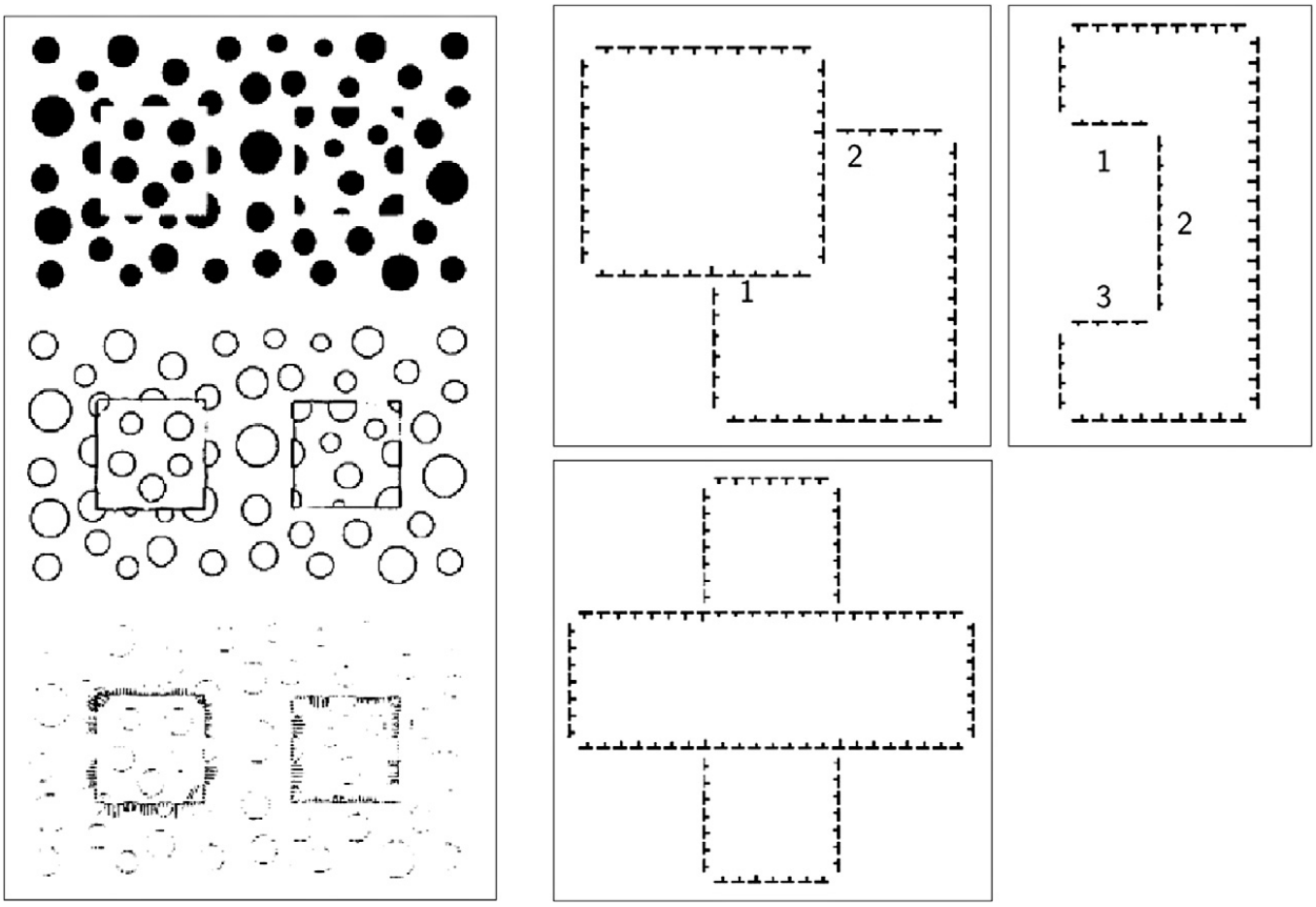



Fig. 22. Model simulations to generate border-ownership. *Left*: Results of contour grouping based on selective integration of responses of end-stop cells using long-range integration with figure-eight shape spatial integration fields. Two input patterns (top) depict an occlusion pattern generated by a near surface and an occlusion generated by a hole in the ground revealing a far surface. The final boundary representation (center) is generated by integrating local oriented contrasts and the output from grouping cells that generate illusory contours induced by the occlusions. The selectivity of the grouping scheme to the orientation of the end-stop directions the figure-ground direction (border-ownership) is signaled opposite to the side with the majority of oriented end-stop cells that respond to partially occluded texture items (circles). In the left input image all occluded circular items appear in the background which assigns the ownership direction towards the central surface region. The occluded circles in the right input image all appear at the side of the central square, which leads to the assignment of ownership to the surface region in the surround region (from F. Heitger, R. von der Heydt, E. Peterhans, L. Rosenthaler, O. Kübler. Simulation of neural contour mechanisms: Representing anomalous contours. *Image and Vision Computing*, 16: 407–421, 1998, reprinted with permission). *Right*: Results of a contour integration scheme that groups oriented contrasts of compatible ownership direction into figural layout. Model V1 cells are sensitive to oriented contrast but insensitive to ownership direction. They send their activities to pairs of like-oriented model V2 cells selective to opposite directions of border-ownership. A mechanism utilizes spatial weights along co-axial orientation for long-range integration of mutually relatable ownership activities. In accordance with the layout of spatial bipole interactions (Fig. 15b) excitatory connections yield the support of oriented ownership activity from figure-eight weighting fields along all directions sampling the full circle of local surround of a target location. At the same time a cell gathers inhibitory input from incompatible arrangements of ownership activities (see details in supplementary data in [78]). The model simulations demonstrate successful disambiguation for mutually occluding surface patches, a C-shape with convex and concave corners and boundary fragments, and the mutual occlusion of two rectangle regions. The time courses of neural responses (not shown here) replicate the evolution of the assignment decisions also observed in experiments (from L. Zhaoping. Border ownership from intracortical interactions in visual area V2. *Neuron*, 47: 143–153, 2005, reprinted with permission).

control described in Eq. (3.11). Summing pairs of cells with opposite contrast polarity both were gating the integrated facilitatory surround activation results in contrast invariant BO cell responses.

Zhaoping [78] investigated the generation of BO signals by long-range intra-areal integration and recurrent interaction in a model of area V2 accounting for some key observations from neurophysiological experiments. The key suggestion is that no specific figural information or localized features from junction coding are necessary to account for the observed results. In this model the output of oriented contrast cells are integrated by local pairs of oriented BO-selective cells, each of which with opposite selectivity to figure side. Such cells come in mutually coupled pairs

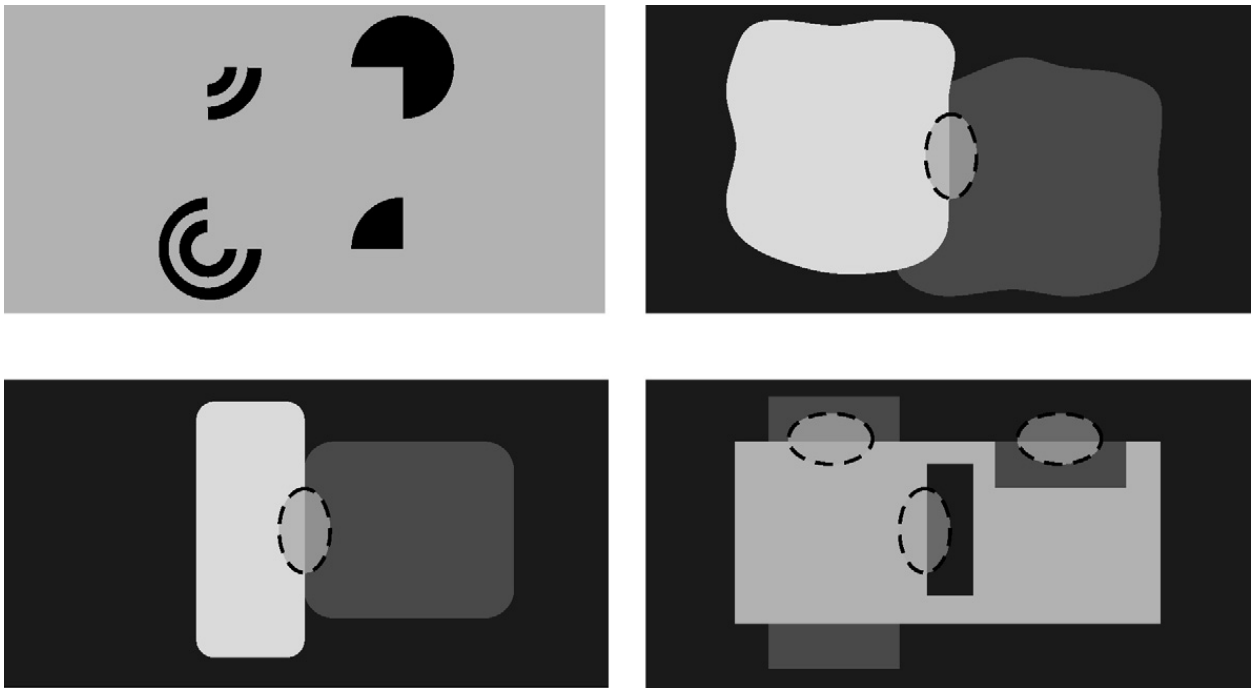


Fig. 23. Challenges for border-ownership assignments. Four different input figures are sketched that address challenges for models of border-ownership assignment. A variant of the Kanizsa–Varin figure configuration (top left) combines fragments from three-quarter circles and one-quarter circle region (filled and as concentric circles). A model that accounts for the statistics of the direction of end-stopping responses would fail to assign ownership direction since this information is perfectly balanced in this figural layout. Yet the boundary as a whole is owned by the square surface region in the center that occludes the background and the circular patterns therein. Two partially occluding surface patches with corners rounded off (top right) depict a variant of the pattern in Fig. 21 bottom left. Once again local oriented contrast detectors are depicted by dashed ellipses. Any model that relies on localized corner representations would fail to assign the ownership direction thus indicating major contributions by grouping and context information. T-junctions are indicative for mutual occlusions. However, the configuration of the surface patches and their shapes provide powerful information about interpreting the scene and the depth arrangements (bottom left). In the shown case the light and the gray region appear at the same depth level such that the ownership assignment appears to be ambiguous (although there is a slight bias in favor of the bright region to be in front). Physiological experiments probing cells and their border ownership responses so far use strictly homogeneous surface properties. If surfaces are painted such that they consist of different reflectances the situation becomes far more complicated (bottom right). The gray patch on the right belongs to the occluding rectangular surface in the middle (that owns the boundary) although T-junctions and shape configurations indicate quite the opposite. Also, the vertically oriented dark rectangle in the center is indicative of a figure in front owning the contour. However, the situation could receive a valid interpretation that is opposite to the previous one, namely if the vertical rectangle is interpreted as a hole (cut out from the light surface patch in the middle) and the dark region is part of the background surface. The two interpretations cannot be decided by taking boundary information into account only. Only when surface properties and their global arrangements are considered the interpretation can be formed.

of excitatory neurons and inhibitory interneurons. Excitatory cells integrate activities from spatially elongated regions similar to the shape depicted in Fig. 15 (right). Inhibitory cells integrate excitatory BO cell activations from a "bow-tie" zone that is oriented orthogonal to the axis of the excitatory integration field such that the two weighting fields do not overlap. This slightly deviates from the repulsion zone sketched in Fig. 15 (right), which is depicted rotationally symmetric in shape and thus overlaps with the prolate zone of excitatory weights in accordance to anatomical data. Each repulsion zone samples from orientation selective neurons of the same BO direction preference as the integration field for the excitatory target neuron. Figure-eight subfields for excitatory integration are rotated to sample over the full 180° range of the circular neighborhood of the target cell. This scheme of integration utilizes criterion 3 indicating the direction of figural extent. Symmetry breaking for BO preference is achieved by tonic low noise inputs. Firing rates of excitatory and inhibitory neurons were determined by non-linear signal transforms of the type discussed in Section 3.3.2. The particular shape of the non-linear signal functions as well as the precise weightings in the integration fields were derived from earlier investigations of network stability [84]. The approach successfully replicates cell response behavior for various shapes and sizes as used in the Zhou et al. study and also shows similar reductions in response strength for more complex shapes (see Fig. 22, right).

While these models battle for their explanatory power trying to identify the basic principles underlying the neuroscience data, they might still be incomplete to account for various interactions of neural mechanisms in generating perceptual surface representations in three dimensions and with the intrinsic qualities such as apparent lightness, color, transparency, color, and depth. In addition, neither model has so far been tested under conditions where the input is perturbed by noise or in more complex stimulus situations such as natural scenes. In this respect, the aims of the authors to exclude specific inputs from detection of geometric cues (junctions) or feedback signals from higher areas cannot exclude their possible role. For example, various demonstrations have supported the view that feedback can enhance the sensitivity to detect the presence of figural boundaries when feature noise generates scene clutter such as in surface texture patterns. In such cases feedback helps to segregate figural shape outline which will otherwise got lost when using hierarchical feedforward processing only (see [85]). Also in the case of texture boundaries the location of a region boundary is imprecise. The larger context that is considered by using larger integration fields (increased receptive field size) that exceed the lateral intra-cortical integration fields may also help to stabilize the processing.

The localized consideration of individual cell properties, such as responding to BO direction, might fall short of fully capturing the generation of coherent neural surface representation. Grossberg [58] proposed the FACADE framework to account for various perceptual phenomena involved in figure-ground segregation in spatial vision. For example, the perceived lightness of a surface region might change for different (monocular) depths and also surface lightness and contrast are affected by surrounding surface patches. Furthermore, responses of differently scaled receptive fields can be converted into ordinary depth. The arrangements of surface patches in turn influence their appearance of ownership of a boundary. Consider, for example, the configurations in Fig. 23 (bottom, left and right). In the configuration on the left the patches only touch each other and are indeterminate owning the shared boundary. Yet, depending on the spatial extent of the regions a local decision mechanism for BO direction will eventually assign one or the other direction.[10] Now consider the case shown on the right where three types of contour segments appear which are made explicit by the border of an oriented receptive field. The left cell at the upper contour should signal the BO direction downward while the right cell should signal the same direction irrespective of the opposite shape boundary configuration in both situations. Dependent on the perceptual judgment, the figure direction either points to the right (dark patch perceived in front of the larger light rectangle), points to the left (dark region depicts the background seen through a hole in the center of the occluding light rectangle), or no direction is signaled (dark region is a paint on the surface patch). The interpretations are biased by the shape of the region outlines (again, as observed by the Gestaltists) and could be manipulated by an attentional bias. Simulations of related configurations that were motivated by psychophysical investigations have been described in Kelly and Grossberg [86].

The examples discussed above highlight that surface perception that need to be signaled by a distributed network of cells might change due to configuration effects or the assignment of particular qualities. Again, utilizing illusory boundaries, Gillam and Nakayama [87] investigated the appearance of surface boundaries in depth arrangements. They showed that subjective boundaries that were generated by line terminators could be strengthened or extinguished depending on the ordinary depth relation of the apparent (smooth) surfaces that contained the lines in the images. Even more impressive was the demonstration that the subjective boundary can be generated by the occluded surface (and could be distant from the one induced by the image features) rather than the occluding surface that, in turn, owns this boundary in terms of the 3D interpretation of such a scene. This observation is puzzling when we assume that the assignment of ownership direction appears early and uses image related boundary groupings alone.

## 4. Summary and prospects: Surface perception, object recognition, and visual attention

This review began with an ambitious agenda, to describe the current state of modeling of important behavioral and physiological data concerning surface perception. Along the way we only touched on the challenges posed by the perception of surfaces in motion. Even for static surfaces, our focus narrowed to a consideration of challenges associated with the processing of boundaries between surfaces, including the detection and completion of boundaries, and the

[10] For the sake of completeness, it should be noted that such stimulus variations have not been tested in physiological experiments so far. So it remains to see whether and how cells in area V2 and V4, which have been demonstrated to signal ownership direction for simple stimuli, will respond to such variants of stimulus probes.

all-important assignment of "ownership" of a boundary between two surfaces by the surface that is nearer to the observer. Throughout we have emphasized that the challenges faced by the human visual system, operating in real time, require a rapid and dynamically reconfigurable grouping and segmentation of visual input in ways first qualitatively articulated by Gestalt psychologists, and, increasingly, captured by systems of non-linear differential equations that describe neural networks. In such models, nodes correspond to cells or populations of cells *in vivo*, and connections between cells are described by weights and indices in the equations. The important behavioral competencies of the models are expressed by its dynamical states, which are taken to correspond to perceptual states. Models involving feedback, in particular, seem to better characterized the labile and reconfigurable behavior of biological networks than earlier models that describe a rigid feedforward hierarchy. Though it should be noted that even feedback models can (and should) capture the dynamics of very rapid visual processing for simple or relatively noise-free tasks where important visual competencies, including object recognition, appear to be supported by a single feed-forward neural cascade [88,89].

One of the most important aspects of surface processing that we have not described in detail concerns the generation of neural representations of homogeneous featural quality, such as brightness, color, or even motion. One perceptual phenomenon considered as an underlying substrate for generating neural surface representations is *filling-in*. Evidence suggests that filling-in is generated by active processes that are time-consuming and co-vary as a function of spreading distance. These and other findings support the view that neural activations are generated to signal a homogeneous representation of surface layout. This contradicts other proposals to account for this phenomenon. Two recent proposals suggest an underlying mechanistic account for filling-in. The 'isomorphic theory' assumes an active spread of activation on a pointwise retinotopic map of feature representations. Another mechanism is based on a subdivision of cortical neurons; it selectively recruits neurons in a scale-sensitive fashion (see [34] for an overview and discussion). In comparison to the topics of contour detection, boundary completion, and border ownership, the evidence gathered from empirical findings is more diverse and less conclusive at the moment (see [33,35]).

Border ownership discussed in Section 3.4 opens the door to consideration of the three-dimensional layout of surfaces, to the extent of which of two different surfaces is closer to an observer. This can be seen in the stereoscopic display shown in Fig. 11b in which a left-right reversal of image pairs not only changes relative depths of surfaces but also their lightness appearance such that the lighter surface appearing in front is now *transparent*. Recent investigations begin to shed some light on the neural mechanisms, which contribute to these surface representations [17,32].

Another aspect of three-dimensional static surface perception that we touched on only briefly concerns perception of solid shape, as from shading or texture. The shading example described in Section 3.3.3 hints at some of the difficult issues of measurement of image gradients and grouping of regions necessary for solid surface perception. A related recent treatment of shape-from-texture is provided by Grossberg, Kuhlmann and Mingolla [90].

We have not discussed how objects are recognized from their surface appearance. While this is clearly of paramount importance to vision, it must be emphasized that one can "see" an object or surface perfectly well without recognizing it. Section 2.4 gave some pointers to issues related to the perception of the featural quality (e.g., brightness or color) of surface regions. Conversely, the demands of recognition place different constraints on neural networks than do the problems involved in *seeing*, Carpenter [91].

Finally, the "elephant in the room" that we have ignored until this point is *attention*. No part of our visual consciousness is unaffected by the need for our organism to select from among the myriad patterns available for sampling by our eyes at any moment. Consider what happens when we first look at an object that is not instantly recognizable. Typically, we make scanning eye movements, directing our fovea—the central region of our vision that is most sensitive to fine detail—around to a variety of points of interest on the object. Cortical magnification, which describes the relative proportion of cortical tissue that is devoted to processing various parts of the visual field, is such that the pattern formed in the brain's first cortical visual area, V1, is a greatly distorted version of what is on the retina—with the central, foveal region receiving a disproportionately large portion of the V1 resources, making the problem of reconciliation of multiple views challenging. The brain must combine the information from several views into a single category that is invariant to where we happen to be gazing at the moment, so that on future presentations of the same object, recognition can be efficient, without requiring a painstaking revisiting of all points of earlier foveation.

Though initially strange to contemplate, a visual system that is trying to integrate information from multiple foveations directed at different locations on a surface that has yet to be recognized cannot logically rely on its recognition component to delineate the extent of the object in the visual field. Even to ensure that a foveation has successfully moved from one part of a given object's surface to another part of that same object, as opposed to some other object's

surface, the visual system must rely on some "data-driven" representation within which a current surface-of-interest can be "tagged" for preferential treatment as a target for several foveations. Itti and Koch [92] have investigated the *saliency* of locations in a visual scene in this regard. Recently Fazl, Grossberg and Mingolla [93] have proposed that the idea of an "attentional shroud," first proposed by Tyler and Kontsevich [94], can be of help. An attentional shroud interacts with surface representations to select a particular object's coherent surfaces for privileged access to the detailed processing afforded by conscious awareness. The shroud can loosely conform to the three-dimensional structure of surfaces, including curved surfaces that are not confined to single "depth planes". The shroud notion may thus be a bridge between our increasingly detailed understanding of surface perception and the rich tradition of study of attention that has characterized the more "cognitive" aspects of psychology over the years.

**Acknowledgements**

HN was supported in part by the European Union (EU FP6 IST Cognitive Systems Integrated project: Neural Decision-Making in Motion; project number 027198). AY was supported in part by the National Eye Institute (EY13135). EM was supported in part by the National Science Foundation (NSF SBE-0354378), and the Office of Naval Research (ONR N00014-01-1-0624).

## Appendix A. Brief introduction to neurodynamics and notational formats

The transmission of activity in neurons is denoted in terms of potential changes across the membrane of a cell. Single cell dynamics can be described at various levels of detail, e.g., at the level of multi-compartments, as single compartment entity or as a cascade model ([95]; see Fig. 24). Here, we utilize single compartment models of neurons, which are essentially point-like representations of a neuron neglecting influences from widespread dendrites and related non-linear interactions.

The cell membrane is composed of layers of proteins and lipid molecules that separate the internal and external conducting solution. The membrane acts as a capacitance to build a charge at both sides of the membrane. Without any input current the cell membrane is in a state of dynamic equilibrium in which currents are flowing across the membrane that balance each other, resulting in zero net current flow. The membrane acts as a resistor that blocks ions of different type to freely pass across the barrier. Proteins of different type act as gates that have constant or activity dependent conductances allowing different amounts of ions passing the membrane. A simple description of a piece of membrane takes into account the conductance $C$, the resistance $R$ and the resting potential $v$, resembling an $RC$ circuit, Fig. 25 (left). Applying Kirchhoff's second law describes the dynamics of the membrane potential (voltage) given an arbitrary input current $i$ (that is injected into the soma of the cell)

$$\tau \frac{\mathrm{d}v(t)}{\mathrm{d}t} = -v(t) + v_{\mathrm{rest}} + R \cdot i(t) \tag{A.1}$$

with $\tau = RC$ and a resting potential $v_{\mathrm{rest}}$ that is in simulations set to zero for convenience.

If we take into account excitatory and inhibitory synaptic input that is delivered by fast chemical synapses then the respective synaptic currents need to be incorporated in the dynamic voltage equation. This leads to the following dynamics

$$\tau \frac{\mathrm{d}v(t)}{\mathrm{d}t} = -v(t) + R \cdot g_e \cdot \big(E_e - v(t)\big) - R \cdot g_i \cdot \big(v(t) - E_i\big) \tag{A.2}$$

where $g_e$ and $g_i$ denote functions that represent time and input dependent membrane conductances (separate for excitatory and inhibitory synapses, respectively), and $E_e$ and $E_i$ denote saturation points defining the respective reversal battery potentials (see Fig. 25, right).

In order to account for different details of observed membrane dynamics different forms of $g_e$ and $g_i$ appear in different models. For example, Hodgkin and Huxley in their Nobel awarded investigations studied the dynamics of the conductances in the spike generating activity in the giant squid axon and identified the voltage dependency of the dynamics of specific ion channel gatings. The models proposed by Sperling [64] and Heeger [67] suggest a gain control mechanism that varies the inhibitory conductance channel by the low-pass filtered luminance level (Sperling) or cortical simple cell activity by overall activation of the weighted average of neural activity from a

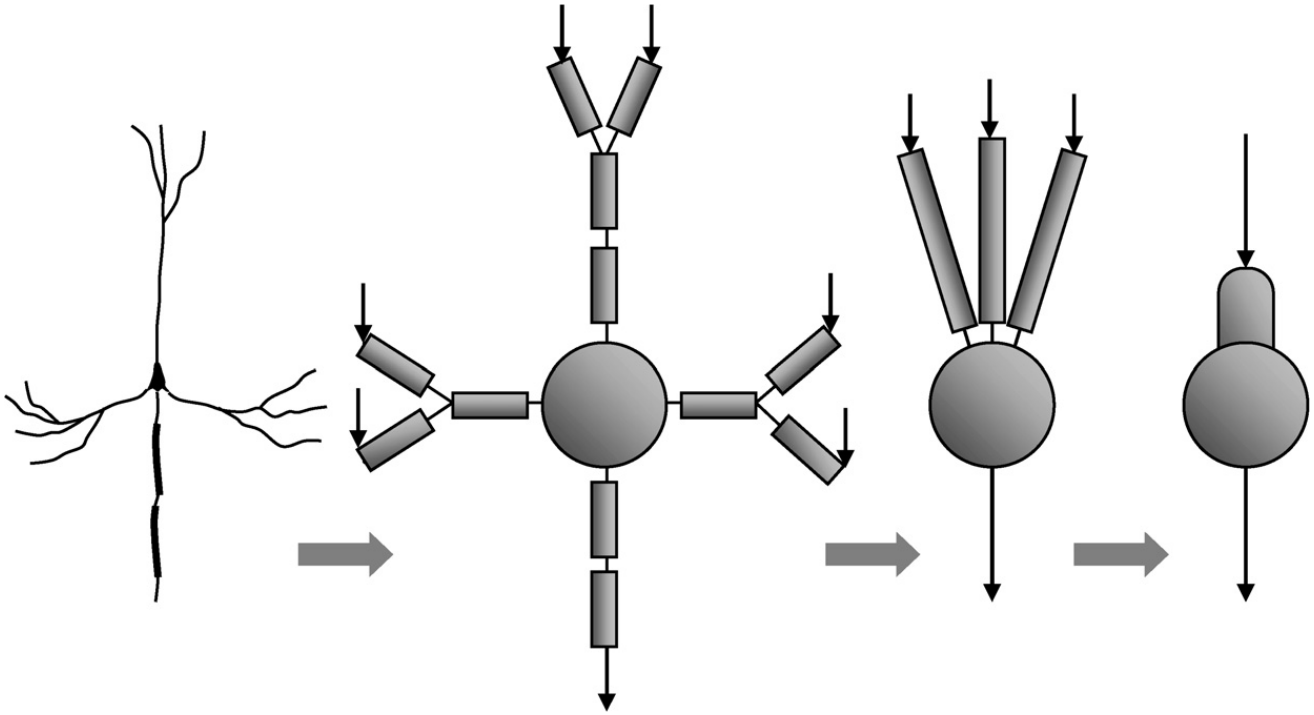


Fig. 24. Different levels of abstraction of nerve cells as computational entity. A pyramidal cell in the neocortex is sketched (left panel) with the cell body (soma) with the apical dendrites (upper segments), the basal dendrites (the two branches left and right), and the axon (lower segment). Detailed compartment models try to detail the morphology in a realistic fashion taking into account the spatial structure of a neuron and the dynamics (second panel). Reduced compartment models focus on the somato-dendritic interaction by modeling the dendritic tree by a few compartments connecting with the cell body compartment (third panel). Single compartment model consider a cell as a point process neglecting the spatial structure of the neuron. Soma and the dendrite are modeled as one homogeneous compartment (right panel).

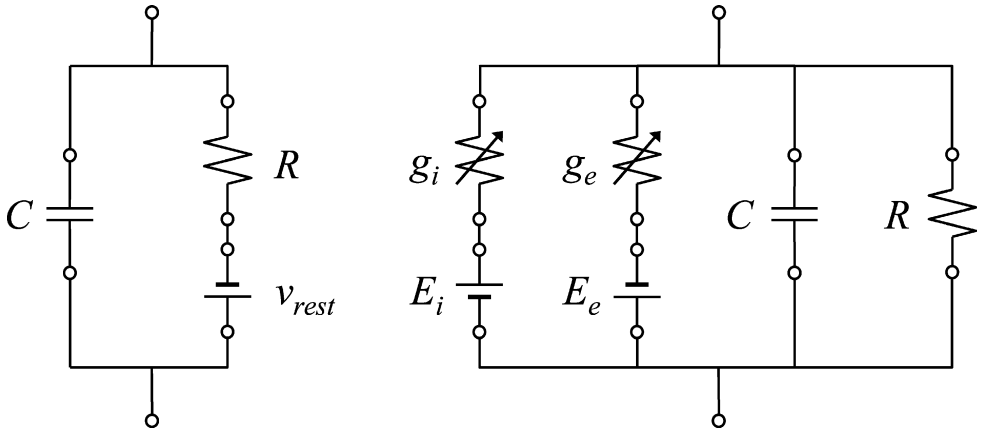



Fig. 25. Passive circuits modeling the dynamics of the membrane potential. Simple single compartment models of neurons describe the membrane as a layered patch of phospholipid molecules that separate the internal and external conducting solution acting as an electrical capacitance. In its simplest description the membrane is a passive electrical device consisting of the capacitance, $C$, a specific membrane resistance (barrier for ions to flow through the membrane, $R$) and a resting potential driven by a battery ($v_{\text{rest}}$, $RC$ circuit; left panel). A more detailed model of single compartment neuron models takes into account different synaptic interactions, namely excitatory and inhibitory input currents (right panel). The membrane resistance is dependent on the excitatory and inhibitory inputs, respectively (indicated by the arrows; conductance $g$ is inverse proportional to the resistance, $g = R^{-1}$).

pool of neighboring cells (Heeger). Grossberg [55] summarized and unified various proposals by using a generalized notation of the membrane equation, namely

$$\tau \frac{\mathrm{d}x(t)}{\mathrm{d}t} = -A \cdot x(t) + \big(B - C \cdot x(t)\big) \cdot \mathrm{net}^{\mathrm{ex}} - \big(D + E \cdot x(t)\big) \cdot \mathrm{net}^{\mathrm{in}} \tag{A.3}$$

which is the basis for the notational format used in this review article (compare Eq. (3.7) and the following ones). The introduction of parameters $C$ and $E$ allows transforming parts of this generic equation into additive components by eliminating the shunts, such as in the case of additive center-surround interaction or shunting inhibition. Saturation properties can be investigated by the steady-state solution (for simplicity, we assume here that the net inputs are generated by feedforward feeding signals). We get

$$x(t) = \frac{B \cdot \text{net}^{\text{ex}} - D \cdot \text{net}^{\text{in}}}{A + C \cdot \text{net}^{\text{ex}} + E \cdot \text{net}^{\text{in}}} \tag{A.4}$$

The limits for increasing excitatory input, $\lim \text{net}^{\text{ex}} \to \infty$, determine an upper bound $x_{\uparrow}(t) = B/C$, while the limits for increasing inhibitory input, $\lim \text{net}^{\text{in}} \to \infty$, determine a lower bound $x_{\downarrow}(t) = -D/E$.

## References


[1] Gibson JJ. The perception of the visual world. Boston: Houghton Mifflin; 1950.
[2] Gibson JJ. The senses considered as perceptual systems. Boston: Houghton Mifflin; 1966.
[3] Gibson JJ. The ecological approach to visual perception. Boston: Houghton Mifflin; 1979.
[4] Marr D. Vision. New York: W.H. Freeman; 1982.
[5] Berzhanskaya J, Grossberg S, Mingolla E. Laminar cortical dynamics of visual form and motion interactions during coherent object motion perception. Spatial Vis 2007;20(4):337–95.
[6] Wallach H. On the visually perceived direction of motion. Psychol Fors 1935;20:325–80;
English translation by Wuerger S, Shapley R, Rubin N. Perception 1966;25:1217–367.
[7] Shimojo S, Silverman GH, Nakayama K. Occlusion and the solution to the aperture problem for motion. Vision Res 1989;29(5):619–26.
[8] Bregman AL. Asking the "what for" question in auditory perception. In: Kubovy M, Pomerantz JR, editors. Perceptual organization. Hillsdale, NJ: Earlbaum Associate; 1981.
[9] Kanizsa G. Organization in vision: essays on Gestalt perception. New York: Praeger Press; 1979.
[10] Nakayama K, Shimojo S, Silverman GH. Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. Perception 1989;18(1):55–68.
[11] Duncan RO, Albright TD, Stoner GR. Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. J Neurosci 2000;20(15):5885–97.
[12] Anstis SM. Imperceptible intersections: the chopstick illusion. In AI and the Eye. New York: Wiley; 1990.
[13] Witkin AP, Tenenbaum JM. On the role of structure in vision. In: Beck J, Hope B, Rosenfeld A, editors. Human and machine vision. New York: Academic Press; 1983. p. 481–543.
[14] Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 1962;160:106–54.
[15] Barlow HB. Single units and sensation: a neuron doctrine for perceptual psychology? Perception 1972;1(4):371–94.
[16] Livingstone MS. Mechanisms of direction selectivity in macaque V1. Neuron 1998;20(3):509–26.
[17] Yazdanbakhsh A, Livingstone MS. End stopping in V1 is sensitive to contrast. Nat Neurosci 2006;9(5):697–702.
[18] von der Heydt R, Peterhans E, Baumgartner G. Illusory contours and cortical neuron responses. Science 1984;224:1260–2.
[19] Livingstone MS. Two-bar interactions in space and time: evidence for common mechanisms in stereopsis and direction selectivity. Soc Neurosci Abstracts 1999;25:1934.
[20] Pack CC, Born RT, Livingstone MS. Two-dimensional substructure of stereo and motion interactions in macaque visual cortex. Neuron 2003;37(3):525–35.
[21] Grossberg S, Mingolla E. Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentation. Perception Psychophys 1985;38:141–71.
[22] Grossberg S, Raizada RDS. Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. Vision Res 2000;40:1413–32.
[23] Neumann H, Mingolla E. Computational neural models of spatial integration mechanisms for perceptual grouping. In: Shipley TF, Kellman P, editors. From fragments to objects: segmentation and grouping in vision. Amsterdam: Elsevier; 2001. p. 353–400.
[24] Hung CP, Ramsden BM, Chen LM, Roe AW. Building surfaces from borders in Areas 17 and 18 of the cat. Vision Res 2001;41(10–11):1389–407;
Roe AW, Lu HD, et al. Cortical processing of a brightness illusion. Proc Natl Acad Sci USA 2005;102(10):3869–74.
[25] Levitt JB, Lund JS. Contrast dependence of contextual effects in primate visual cortex. Nature 1997;387(6628):73–6.
[26] Peterhans E, von der Heydt R. Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. J Neurosci 1989;9(5):1749–63.
[27] Roe AW, Lu HD, Hung CP. Cortical processing of a brightness illusion. Proc Natl Acad Sci USA 2005;102(10):3869–74.
[28] Rossi AF, Desimone R, Ungerleider LG. Contextual modulation in primary visual cortex of macaques. J Neurosci 2001;21(5):1698–709.
[29] von der Heydt R. Approaches to visual cortical function. Rev Physiol Biochem Pharmacol 1987;108:69–150.
[30] Zipser K, Lamme VA, Schiller P. Contextual modulation in primary visual cortex. J Neurosci 1996;16(22):7376–89.
[31] Grossberg S, Todorović D. Neural dynamics of 1-D and 2-D brightness perception: A unified model of classical and recent phenomena. Perception Psychophys 1988;43:723–42.

[32] Grossberg S, Yazdanbakhsh A. Laminar cortical dynamics of 3D surface perception: stratification, transparency, and neon color spreading. Vision Res 2005;45(13):1725–43.
[33] Pessoa L, Thomson E, Noe A. Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. Behav Brain Sci 1998;21:723–802.
[34] Komatsu H. The neural mechanisms of perceptual filling-in. Nature Rev Neurosci 2006;7:220–31.
[35] Neumann H, Mingolla E. Contour and surface perception. In: Arbib MA, editor. Handbook of brain theory and neural networks, II. Cambridge MA: MIT Press; 2003.
[36] Metelli F. The perception of transparency. Sci Am 1974;230(4):90–8.
[37] Sigman M, Cecchi GA, Gilbert CD, Magnasco MO. On a common circle: Natural scenes and Gestalt rules. Proc Nat Acad Sci 2001;98:1935–40.
[38] Nakayama K. Stereoscopic surface perception. Proc Nat Acad Sci 1999;24:919–28.
[39] Adelson EH. Perceptual organization and the judgment of brightness. Science 1993;262:2042–4.
[40] Adelson EH. Lightness perception and lightness illusions. The new cognitive neurosciences. Cambridge, MA: MIT Press; 2000. 339–351; M. Gazzaniga.
[41] Koffka K. Principles of Gestalt psychology. London: Rotledge & Kegan Paul Ltd; 1935.
[42] Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381:520–2.
[43] Hochstein S, Ahissar M. View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron 2002;36:791–804.
[44] Lamme VAF, Roelfsema PR. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 2000;23:571–9.
[45] Baumgartner G, von der Heydt R, Peterhans E. Anomalous contours: A tool in studying the neurophysiology of vision. Experiment Brain Res (Suppl) 1984;9:413–9.
[46] Lesher GW, Mingolla E. The role of edges and line-ends in illusory contour formation. Vision Res 1993;33:2253–70.
[47] Field DJ, Hayes A, Hess RF. Contour integration by the human visual system: Evidence for a local "association field". Vision Res 1993;33:173–93.
[48] Grossberg S, Mingolla E, Ross WD. Visual brain and visual perception: How does the cortex do perceptual grouping? Trends Neurosci 1997;20:106–11.
[49] Bosking WH, Zhang Y, Schofield B, Fitzpatrick D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. J Neurosci 1997;17:2112–27.
[50] Kapadia MM, Westheimer G, Gilbert CD. The spatial distribution of contextual interactions in primary visual cortex and in visual perception. J Neurophysiol 2000;84:2048–62.
[51] Peterhans E, von der Heydt R. Subjective contours—bridging the gap between psychophysics and physiology. Trends Neurosci 1991;14:112–9.
[52] Mendola JD, Dale AM, Fischl B, Liu A, Tootell RBH. The representation of real and illusory contours in human cortical visual areas revealed by fMRI. J Neurosci 1999;19:8560–72.
[53] Parent P, Zucker S. Trace inference, curvature consistency, and curve detection. IEEE Trans PAMI 1989;11:823–39.
[54] Kellman PJ, Shipley TF. A theory of visual interpolation in object perception. Cognitive Psychol 1991;23:141–221.
[55] Grossberg S. How does the brain build a cognitive code? Psychol Rev 1980;87:1–51.
[56] Grossberg S, Williamson JR. A neural model of how horizontal and interlaminar connections in visual cortex develop into adult circuits that carry out perceptual grouping and learning. Cerebral Cortex 2001;11:37–58.
[57] Zucker S. Differential geometry from the Frenet point of view: Boundary detection, stereo, texture and color. In: Paragios N, Chen Y, Faugeras O, editors. Mathematical models of computer vision: the handbook. New York: Springer; 2005.
[58] Grossberg S. 3-D vision and figure-ground separation by visual cortex. Perception Psychophys 1994;55:48–121.
[59] Grossberg S. Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. Psychol Rev 1997;104:618–58.
[60] Grossberg S. How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. Spatial Vis 1999;12:163–85.
[61] Grossberg S. How does the cerebral cortex work? Development, learning, attention, and 3-D vision by laminar circuits of visual cortex. Behav Cognitive Neurosci Rev 2003;2:47–76.
[62] Mingolla E, Ross WD, Grossberg S. A neural network for enhancing boundaries and surfaces in synthetic aperture radar images. Neural Networks 1999;12:499–511.
[63] Neumann H, Sepp W. Recurrent V1–V2 interaction in early visual boundary processing. Biol Cybernet 1999;81:425–44.
[64] Sperling G. Model of visual adaptation and contrast detection. Perception Psychophys 1970;8:143–57.
[65] Grossberg S. Contour enhancement, short term memory, and constancies in reverberating neural networks. Stud Appl Math 1973;LII:213–57; Reprinted in Grossberg S. Studies of mind and brain. Boston: Reidel; 1982.
[66] Grossberg S, Mingolla E. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. Psychol Rev 1985;92(2):173–211.
[67] Heeger DJ. Normalization of cell responses in cat striate cortex. Visual Neurosci 1992;9:184–97.
[68] Eckhorn R, Reitboeck HJ, Arndt M, Dicke P. Feature linking via synchronization among distributed assemblies: simulations of results from cat visual cortex. Neural Comput 1990;2:293–307.
[69] Gove A, Grossberg S, Mingolla E. Brightness perception, illusory contours, and corticogeniculate feedback. Visual Neurosci 1995;12:1027–52.
[70] Bullier J. Integrated model of visual processing. Brain Res Rev 2001;36:96–107.
[71] Desimone R. Visual attention mediated by biased competition in extrastriate visual cortex. Philos Trans Royal Soc London Ser B 1998;353:1245–55.

[72] Polat U, Mizobe K, Pettet MW, Kasamatsu T, Norcia AM. Collinear stimuli regulate visual responses depending on cell's contrast threshold. Nature 1998;391:580–4.
[73] Weidenbacher U, Bayerl P, Neumann H, Fleming R. Sketching shiny surfaces: 3D shape extraction and depiction of specular surfaces. ACM Trans Appl Perception 2006;3:262–85.
[74] Todd JT. The visual perception of 3D shape. Trends Cognitive Sci 2004;8:115–21.
[75] Metzger W. Gesetze des Sehens. Frankfurt/Main: Verlag Waldemar Kramer; 1975. Translated by Lother Spillmann, as "Laws of Seeing", MIT Press, 2006.
[76] Zhou H, Friedman HS, von der Heydt R. Coding of border ownership in monkey visual cortex. J Neurosci 2000;20:6594–611.
[77] Heitger F, von der Heydt R, Peterhans E, Rosenthaler L, Kübler O. Simulation of neural contour mechanisms: Representing anomalous contours. Image Vision Comput 1998;16:407–21.
[78] Zhaoping L. Border ownership from intracortical interactions in visual area V2. Neuron 2005;47:143–53.
[79] Kikuchi M, Akashi Y. A model of border-ownership coding in early vision. In: Dorffner G, Bishof H, Hornik K, editors. Proc. ICANN 2001. LNCS, vol. 2130. Springer; 2001. p. 1069–74.
[80] Schuetze H, Niebur E, von der Heydt R. Modeling cortical mechanisms of border ownership coding. J Vision 2003;3:114a.
[81] Craft E, Schuetze H, Niebur E, von der Heydt R. A physiologically inspired model of border ownership assignment. J Vision 2004;4:728a.
[82] Nishimura H, Sakai K. The computational model for border-ownership determination consisting of surrounding suppression and facilitation in early vision. Neurocomput 2005;65–66:77–83.
[83] Sakai K, Nishimura H. Surrounding suppression and facilitation in the determination of border ownership. J Cognitive Neurosci 2006;18:562–79.
[84] Li Z. A neural model of contour integration in the primary visual cortex. Neural Comput 1998;10:903–40.
[85] Thielscher A, Neumann H. Neural mechanisms of cortico–cortical interaction in texture boundary detection: A modelling approach. Neurosci 2003;122:921–39.
[86] Kelly F, Grossberg S. Neural dynamics of 3-D surface perception: figure-ground separation and lightness perception. Perception Psychophys 2000;62:1596–618.
[87] Gillam B, Nakayama K. Subjective contours at line terminations depend on scene layout analysis, not image processing. J Experimental Psychol Human Perception Perform 2002;28:43–53.
[88] Fabre-Thorpe M, Richard G, Thorpe SJ. Rapid categorization of natural images by rhesus monkeys. Neuroreport 26 1998;9(2):303–8.
[89] Grossberg S. Linking the laminar circuits of visual cortex to visual perception: development, grouping, and attention. Neurosci Biobehav Rev 2001;25(6):513–26.
[90] Grossberg S, Kuhlmann L, Mingolla E. A neural model of 3d shape-from-texture: multiple-scale filtering, boundary grouping, and surface filling-in. Vision Res 2007;47(5):634–72.
[91] Carpenter GA. Neural network models of learning and memory: leading questions and an emerging framework. Trends Cognitive Sci 2001;5:114–8.
[92] Itti L, Koch C. Computational modelling of visual attention. Nat Rev Neurosci 2001;2(3):194–203.
[93] Fazl A, Grossberg S, Mingolla E. View-invariant object category learning, recognition, and search: how spatial and object attention are coordinated using surface-based attentional shrouds. Boston, MA, CAS/CNS Technical Report #pending.
[94] Tyler CW, Kontsevich LL. Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. Perception 1995;24(2):127–53.
[95] Herz AVM, Gollisch T, Machens CK, Jaeger D. Modeling single-neuron dynamics and computations: a balance of detail and abstraction. Science 2006;314:80–5.