

To appear in: The Visual Neurosciences (L.M. Chalupa and J.S. Werner, eds.), Cambridge: MIT Press.

Image Parsing Mechanisms of the Visual Cortex

Rüdiger von der Heydt

Krieger Mind/Brain Institute and Department of Neuroscience,
Johns Hopkins University, Baltimore, MD 21218

Short title: same as above

Correspondence:

Rüdiger von der Heydt
Krieger Mind/Brain Institute
Johns Hopkins University
3400 North Charles Street
Baltimore, MD 21218
Phone: 410 516-6416
Fax: 410 516-8648
E-mail: von.der.heydt@jhu.edu

1. Introduction

In this chapter I will discuss visual processes that are often labeled as “intermediate level vision”. As a useful framework we consider vision as a sequence of processes each of which is a mapping from one representation to another. Understanding vision then means analyzing how information is represented at each stage and how it is transformed between stages (Marr, 1982). It is clear that the first stage, the level of the photoreceptors, is an image representation. This is a 2-dimensional array of color values, resembling the bitmap format of digital computers. Retinal processes transform this representation into a format that is suitable for transmission through the optic nerve to central brain structures. A radical transformation then takes place in the primary visual cortex. At the output of area V1 we find visual information encoded as a “feature map”, a representation of local features. The two dimensions of retinal position are encoded in the locations of the receptive fields of cortical neurons, but each neuron now represents not a pixel, but a small patch of the image, and several new dimensions are added to the three color dimensions, such as orientation, spatial frequency, direction of motion, and binocular disparity.

While the local feature representation of V1 is known in detail, we still do not have a good understanding of the nature of processing in the extra-striate areas. At first glance, the visual properties of the neurons in area V2 are rather similar to those of their input neurons of V1, except that V2 neurons have larger receptive fields (Zeki, 1978; Burkhalter and Van Essen, 1986). Many neurons of area V4 again have properties that can be found already in V1 and V2, such as selectivity for color, orientation, and spatial frequency (Zeki, 1978; Desimone et al., 1985; Schein and Desimone, 1990). Some neurons in the extrastriate areas show selectivity for increasingly complex features (see Connor, Chapter xx, this volume).

Here, I will review evidence that the extrastriate areas provide a new stage of processing that may be described as “image parsing”. This stage appears as a mediator between the local feature representation of V1 and the processes of attentional selection and object recognition. From the enormous amount of information that streams in through the optic nerves at each moment the visual system selects a small fraction, and in general precisely what is relevant for a given task. This amazing performance indicates powerful mechanisms for organizing the incoming information. The gestalt psychologists first pointed out that the visual system tends to organize elemental visual units (such as points and lines) into larger perceptual units, or figures, according to certain rules called “gestalt laws” (see Spillmann and Ehrenstein, Chapter xx, this volume). Much of this organization occurs independently of what the subject knows or thinks about the visual stimulus. Kanizsa illustrated this “autonomy of perception” with a painting that is supposed to show a knife behind a glass, but is perceived, instead, as a transparent knife passing in front of the stalk of the glass (Kanizsa, 1979, Fig. 2.19).

Grouping together features that belong to an object is a general task of perception. Specific visual problems arise from the fact that vision is based on 2D projections of a 3D world. Due to spatial interposition, parts of the scene are occluded and features of near and far objects are cluttered in the image. Projections of the same objects vary with the viewing angle, and the true 3D shape of the objects and their relations in space can only be inferred from the images. In principle, any image has an infinite number of possible interpretations in 3D, vision is an “ill-posed problem” (Poggio and Koch, 1985). In spite of this fundamental ambiguity, vision is the most reliable of our senses. Apparently, through evolution and experience, biological vision systems have learned to make efficient use of the regularities present in images and to infer the

missing information (Helmholtz, 1866; Attneave, 1954; Barlow, 1961; Marr, 1982; Poggio and Koch, 1985; Ullman, 1996).

2. Illusory Contours: Creative Mechanisms

A prominent example of this creative process is the phenomenon of illusory contours (Fig. 1). In A, the system “visibly” fills in the missing contours of an overlaying triangle. Note that the illusory contours are not just interpolations between given contrast borders, as it might seem in A, but form also in the absence of contrast borders that could be interpolated (C). In fact, when the corners of the overlaying triangle are defined by lines which could be interpolated, illusory contours do not form (B). What all illusory contour figures have in common is the presence of occlusion cues, such as terminations of lines and edges (L-junctions) (Coren, 1972). Thus, the system seems to infer an occluding object. However, this is not an inference in abstract terms. The mere expectation of a contour does not lead to perception of illusory contours (Fig. 1 D). Apparently, in forming the contours the system combines evidence from occlusion cues with rules such as the gestalt principle of good continuation.

Interestingly, illusory contours are represented in the visual cortex at a relatively early stage. In monkey area V2 many cells respond to illusory contour stimuli as if the contours were contrast borders (von der Heydt et al., 1984). Fig. 2 shows an example of a cell that was tested with a moving illusory bar. The raster plot in B shows that the cell responds when the illusory contour traverses a small region that was determined before as the cell's minimum response field (ellipse, see legend). Fig. 2 C shows a control in which the two bars were moved exactly as in B, but the open ends were closed off with thin lines. Closing lines weaken the perceptual illusion (see figure at the bottom), and they also reduce the responses of the neuron. Cells in V2 respond not only to figures with illusory bars, but also to other figures that produce illusory contours, such as a pattern of two abutting line gratings (Fig. 3, right). It can be seen that the cell of Fig. 3 responds at the same orientations for the illusory contour as for the bar stimulus; thus, it signals the orientation of an illusory contour. Using the criteria of consistent orientation tuning and response reduction by the closing lines, 30-40% of the cells of V2 were found to signal illusory contours of one or the other type, and the results obtained with the two types of contour were highly correlated (von der Heydt and Peterhans, 1989; Peterhans and von der Heydt, 1989).

As shown in Fig. 2, illusory contour responses can be evoked by stimuli which are devoid of contrast over the excitatory center of the receptive field. The inducing contrast features can be restricted to regions from which an optimized bar stimulus would not evoke any response. The cells seem to integrate occlusion features over a region larger than the conventional receptive field (Peterhans et al., 1986). Nevertheless, the extent of spatial integration is limited; for neurons with near-foveal receptive fields, the responses declined if the gap of the stimulus (Fig. 2 B) was made wider than about 3° visual angle.

V2 is one of the largest areas of the monkey cerebral cortex (Felleman and Van Essen, 1991), and the fact that so many cells of this area respond this way indicates that illusory contour stimuli probe a basic function of the visual cortex. V2 is an early stage of processing where responses are fast and highly reproducible. Illusory contour responses arise as early as 70 ms after stimulus onset (von der Heydt and Peterhans, 1989; Lee and Nguyen, 2001). This indicates that illusory contours are probably not the result of object recognition processes at higher levels, but generated within the visual cortex. Computational models have shown how such contours

might be generated (e.g., Grossberg and Mingolla, 1985; Finkel and Sajda, 1992; Heitger et al., 1998).

Illusory contours in V1. Representation of illusory contours has also been demonstrated in V1 of cat (Redies et al., 1986; Sheth et al., 1996) and monkey (Grosf et al., 1993; Lee and Nguyen, 2001; Ramsden et al., 2001). However, it is not clear if cells in V1 also generalize over the various types of illusory contour figures and if they signal the contour orientation. Sheth et al. (1996) and Ramsden et al. (2001) used a combination of optical imaging and single-unit recording to identify the illusory contour representation with the abutting-grating type of stimulus. Sheth et al. found cells with consistent orientation tuning for illusory contours in V1 of the cat. In the monkey, Ramsden et al. found that the representation of illusory contours in V1 is different from that of V2. Illusory contours *reduced* the activity in columns of the corresponding orientation, but *increased* activity in columns of the orthogonal orientation, in contrast to V2, where the same columns were activated by illusory contours and contrast borders. They conclude that V1 de-emphasizes illusory contours.

Studies that compared both areas invariably found marked differences between V1 and V2 in the frequency of cells that signaled illusory contours, the signaling of orientation, and the degree of cue invariance (von der Heydt and Peterhans, 1989; Leventhal et al., 1995; Sheth et al., 1996; Bakin et al., 2000; Ramsden et al., 2001).

Correlation of physiology and perception. Varying the configurations and spatial parameters of the displays shows a tight correspondence between human perception and neural responses for illusory contours generated by abutting gratings (Fig. 3 B) (Soriano et al., 1996). However, in discriminating the shape of illusory figures, the human visual system shows larger spatial integration than the neurons of monkey V2 (Ringach and Shapley, 1996). Because neurons that signal illusory contours are only a subset of the cells that signal contrast edges, orientation-dependent adaptation aftereffects should transfer from contrast-defined to illusory contours, but not in the reverse direction, and the discrimination of orientation should be less accurate for illusory contours than for contrast-defined contours. Both predictions were borne out in psychophysical experiments (Paradiso et al., 1989; Westheimer and Li, 1996). Illusory contours are usually associated with perception of overlay (Coren, 1972), and some neurons in V2 are selective for the implied direction of occlusion of illusory contours (Baumann et al., 1997). Thus, the illusory contour mechanisms may be related to the coding of border ownership discussed below.

Illusory contours are universal. Perception of illusory contours has been demonstrated in a variety of non-human species, including bee, cat, and barn owl (Srinivasan et al., 1987; Bravo et al., 1988; De Weerd et al., 1990; Nieder and Wagner, 1999; for a review see Nieder, 2002). Most elegant is the combination of behavioral experiments with single cell recordings (Nieder and Wagner, 1999).

3. Border Ownership: Image Context Integration

Illusory contours and related visual phenomena are only the tip of an iceberg of cortical processes involved in perceptual organization. Kanizsa's figure (Fig. 1 A) suggests that illusory contours are the product of mechanisms in figure-ground segregation. The system takes the peculiar arrangement of the black elements as evidence for an occluding triangle and hence creates a representation of its contours. In fact, it also creates the representation of a white opaque surface, as one can see from the subtle difference in brightness relative to the

background. The illusory contours appear as the edges of this surface. Similar linking of contour and surface can be observed also for sharp contrast borders. Perception tends to interpret such borders as occluding contours and assigns them to a surface on one or the other side of the border. This compulsion of the visual system is demonstrated by Rubin's vase figure (Fig. 4 A). The borders are perceived either as the contours of a vase, or as the contours of two faces. Each border is perceived as belonging to one or the other side, but rarely to both. In the case of a simple figure such as the white square of Fig. 4 B, the contrast borders are "of course" perceived as the contours of the square. They seem to belong to the enclosed light-textured region. The surrounding gray, which does not "own" these borders, is perceived as extending behind the square, forming the "background". Perception of border ownership is a subtle phenomenon that remained long unnoticed until it was discovered by the gestalt psychologists (Rubin, 1921; Koffka, 1935).

One could argue that even the display of Fig. 4 B is ambiguous. With some effort, the square can also be perceived as a window, and the border then appears as the edge of the frame. Completely unambiguous displays can be produced by means of random-dot stereograms, as shown in Fig. 4 C. When binocularly fused by crossing the eyes (see legend) the top pair shows a tipped square floating in front of a background plane, while the bottom pair shows a square window through which a background plane can be seen. In the first case, the stereoscopic borders are perceived as the edges of the square, in the second case, as edges of the window frame. Perception of border ownership cannot be reversed in this stereogram.

In perceptual experiments we observe the tip of the iceberg. By recording signals in visual cortex we should be able to explore also the depth of it. Contrast borders are represented in the visual cortex by signals of the orientation selective cells discovered by Hubel and Wiesel. Do these signals also represent the relationship between border and surface? This idea can be tested with a simple experiment (Zhou et al., 2000). Light-dark borders are placed in the receptive field of a neuron at optimal orientation (Fig. 5), and the same border is either presented as the right side of a light square (e.g., A1), or the left side of a dark square (B1). Columns A2 and B2 show a similar test with displays of reversed contrast, and columns 3-4 and 5-6 show the same kind of test with squares of larger sizes. The bar graph at the bottom represents the responses of a cell of V2. If we compare the responses to the corresponding displays in A and B we see that in every case the neuron responds more strongly when the edge in the receptive field belonged to a square to the left than a square to the right, despite locally identical stimulation.

Note that the corresponding displays in rows A and B are identical over the entire region occupied by the two squares (as one can see by superimposing them). Thus, if a neuron responds differently, it must have information from outside this region. Thus, by varying the square size we can reveal the extent of image context integration. In this example, square sizes of 4° , 10° , and 15° were tested, and in each case the responses differed depending on the location of the figure. By contrast, the size of the minimum response field of this cell was only 0.4° , which is typical for V2 neurons of the foveal representation. Thus, although the cell can "see" only a small piece of contrast border through the aperture of its receptive field, its responses reveal processing of an area of at least 15° in diameter.

What might be the mechanism of side-of-figure selectivity? For a single square figure on a uniform background relatively simple algorithms would be able to discriminate figure and ground. The convexity of the figure area could be used, or simply the orientation of the L-junctions (corners) on either side of the receptive field, or the fact that the figure is a region of one color enclosed by a region of a different color ("surroundedness"). Any of these strategies

would work for the isolated square. However, for other displays in which border ownership is also perceptually clear, mechanisms based on one simple strategy would fail to produce the right answer. We have used two other configurations besides squares to see how well the neural responses correlated with perception, a C-shaped figure as shown in columns 3 and 4 of Fig. 6, and a pair of overlapping squares as shown in columns 5 and 6 of the same figure. For the C-shape, convexity is not valid, and the L-junctions next to the receptive field are reflected to the other side in comparison with the square, but surroundedness would still be a valid cue. For the overlapping squares, surroundedness is violated, while convexity and orientation of L-junctions are valid.

Fig. 6 shows data of another neuron of V2. Columns 1 and 2 show the same test as described in Fig. 5. This cell was selective for contrast polarity, responding to light-dark edges as shown in A1 and B1, but hardly at all to dark-light edges as shown in A2 and B2 (the actual colors in the experiment were violet, shown here as light gray, and gray; the cell was strongly color selective). The cell was also side-of-figure selective, with a preference for figure location on the left side of the receptive field (display A1). Columns 3 and 4 show a test with a C-shaped figure. It can be seen that the cell “correctly” preferred the display in which the C-figure was located on the left of the field (display B3), although the L-junctions next to the receptive field would rather suggest a figure on the opposite side.

Columns 5 and 6 show a test with two overlapping figures. These displays are fairly symmetric about the receptive field as far as size of regions and distribution of colors are concerned, and neither of the figures is surrounded by uniform color. Nevertheless, the cell preferred display A5 in which the border in the receptive field belongs to the lower left figure. In this case the T-junctions might account for the emergence of the occluding square as a figure (but convexity might also contribute because the overlapped region has a concavity whereas the overlapping region does not). Thus, the response of this cell are entirely consistent with the perception of border ownership. Not all cells tested showed this pattern, but the example is not unusual. About half of the cells with a side-of-figure effect for single squares exhibited the corresponding side preference for overlapping figures, while the others showed no significant response difference. And vice versa, the overlapping figure test predicted the single-figure result in about half of the cases. When tested with the concave side of C-figures, about one third of the cells with a side-of-figure effect for single squares showed preference for the C on the same side, the others were indifferent. Cases in which the side preferences were “contradictory” (as judged by perception) were rare.

Cells with a response preference for one or the other side of the figure were found for any location and orientation of receptive field, and side-of-figure preference was invariant throughout the recording period. The responses of these cells seem to carry information not only about the location and orientation of the contours, but also about the side to which they belong. About half of the orientation selective cells of areas V2 and V4 were found to be side-of-figure selective by the test of Fig. 5. In 32% of the V2 cells the ratio of the responses to preferred and non-preferred sides was greater than 2, and ratios as high as 10 were not unusual. For comparison, by the same criterion, 29% of V1 cells are direction selective (De Valois et al., 1982), 50% of upper-layer V1 cells are opponent color selective (from Fig. 9 of Leventhal et al., 1995). We found side-of-figure selectivity also in V1, but in a smaller fraction of the cells. These results show that the side-of-figure test probes an important aspect of the cortical representation.

Experiments as shown in Figs. 5-6 suggest the existence of cortical mechanisms that use figure-ground cues to assign border ownership. In other words, the signals of orientation

selective cells in V2 might represent not only the location, orientation, and contrast of pieces of contour, but also the side of ownership.

4. Stereoscopic Depth and Monocular Form Cues

A crucial test of the hypothesis of border ownership coding is to examine the responses of orientation-selective cells to contrast-defined and disparity-defined figures. A contrast-defined square is generally perceived as “figure”, with the borders assigned to the square (see Fig. 4), while a corresponding region in a random-dot stereogram is perceived either as a figure, if its disparity is “near”, or as a window, if its disparity is “far”, relative to that of the surrounding region. In the stereogram, the nearer surface always owns the border. Thus, the random-dot stereogram is the “gold standard” of border ownership perception.

Binocular disparity is represented extensively in the monkey visual cortex (Poggio, 1995; Cumming and DeAngelis, 2001) and cells that signal edges in random-dot stereograms exist in area V2 (von der Heydt et al., 2000). These cells are orientation selective and respond to disparity-defined edges as well as to contrast borders (Fig. 7). Most of them are selective for the depth order of the stereoscopic edge, responding, for example, to a vertical edge if the front surface is on the right side, but not if the front surface is on the left side (cell 3 of Fig. 7).

If border ownership is represented in V2, then some of the cells there should combine monocular shape cues with binocular disparity information, and the side-of-figure preference should agree with the preferred depth order. That is, the preferred figure side should be the “near” side of the preferred step edge. This experiment is illustrated in the cartoon of Fig. 8. With the random-dot stereogram, the cell is activated by the left edge of the figure and the right edge of the window, but not by right edge of the square or the left edge of the window. From this we conclude that activation of this cell means border assignment to the surface on the right. Therefore, the responses to the contrast defined square (A stronger than B) show that the cell “correctly” assigns the border to the square, so the square is interpreted as a figure. If the cell responded more strongly to B than A, this would mean that it assigns the border to the frame and the square would be interpreted as a window.

Out of 27 cells recorded in V2 that signaled depth order for random-dot edges and side of figure for contrast-defined figures ($p < 0.05$ in each case), 21 responded according to the “figure” interpretation, and 6 according to the “window” interpretation (Qiu et al., 2001). This result is in agreement with the tendency in human perception to interpret compact uniform regions in the image as objects.

We speculate that the minority of “window” responses might be not just aberrant signals, but the representation of a valid alternative parsing solution. Occasional dissident votes were also recorded when the side preferences for single squares were compared with that for overlapping figures. Monocular form cues are usually ambiguous; a square can be perceived as a window, and even the display of two overlapping squares can alternatively be perceived as an L-shaped object adjacent to a square. It seems plausible that the visual cortex represents several alternative 3D interpretations if the image is ambiguous.

Linking contour and surface attributes. The convergence of stereoscopic edge mechanisms and side-of-figure processing in single cells strongly supports the conclusion that side-of-figure selective cells code for border ownership. Fig. 9 illustrates, for a pair of overlapping squares, how border ownership information is represented together with information about other contour features such as orientation, and color and luminance contrast. Each piece of

contour is represented by two pools of neurons, one for each side of ownership. In analogy to the opponent coding of direction of motion we assume that border ownership is encoded in the relative strength of activity in pairs of neurons of opposite side preference, but otherwise identical receptive fields. This scheme of coding allows the linking of contour and surface attributes. Location and orientation of contour is coded by virtue of orientation selectivity and small size of response fields. Color and brightness of object surface are coded by means of color and contrast polarity selectivity of cells with the corresponding border ownership pointer. Depth of surface is encoded similarly in the activity of stereo edge selective cells.

5. A Case for Low-level Mechanisms

An interesting point is the time course of the border ownership signals. Fig. 10 compares the averaged neuronal responses for figure on the preferred side (thick line) and figure on the nonpreferred side (thin line). Data from areas V1, V2, and V4 are shown. It can be seen that a differentiation of responses occurs soon after the stimulus onset and well before the peak of the responses. Note also that the response difference in V2 neurons remains constant during the remainder of stimulus presentation. If the effect of side of figure were due to feedback from areas of much higher order, we would probably see a delay. The immediate differentiation suggests that the mechanisms reside in these lower order visual areas. Also Bakin et al. (2000) who studied neural correlates of contour saliency, illusory contours, and depth capture in V1 and V2 found no increase of latency for these image parsing processes. Lee and Nguyen (2001) found illusory contour responses in V2 with latencies as low as 50 ms, and differentiation between modal and amodal completion by 70 ms. All these results argue for fast processes, implicating highly parallel mechanisms in the lower-order cortical areas.

Because of the ambiguity of monocular form cues, and the ill-posed nature of the vision problem in general, image segmentation is usually regarded as a task that cannot be solved by low-level computations, but requires the use of stored representations in memory. Why, then, would the visual system use low-level mechanisms at all to resolve figure-ground relationships? Since memory will eventually be used to recognize objects, one may wonder what is the advantage of low-level mechanisms.

In the case of a single border, as in Fig. 6, columns 5-6, assigning figure-ground direction reduces the number of shapes that have to be compared with memory from two to one (by eliminating the inverted L shape created by occlusion), which may not appear as a great saving. However, it is important to recognize that the problem is generally more complex. As an example, consider the display of Fig. 11 A, which might be perceived as two elongated objects occluding one another, or, in keeping with the macaque perspective, a branch in front of a tree stem. Contrast borders divide the display into seven regions of different shapes. Since the contrast borders may be occluding contours, most of these shapes are meaningless because they are surfaces of partly occluded objects, that is, regions that do not own the borders. There are 10 segments of borders (not counting the frame) each of which could belong to one of the adjacent regions, creating a total of $2^{10} = 1024$ possible depth configurations. Each depth configuration defines a different set of shapes. To help the reader to see this, I have illustrated two of the possible 3D decompositions in Fig. 11 B. Most of these configurations are generally not perceived. The point is that there is a large number of shapes that could give rise to an image like that of Fig. 11 A. All of these would have to be searched in memory if the system were not able to assign the borders beforehand. If borders are assigned, only the two bars in front of a

white background have to be processed further. Thus, low-level border assignment reduces the load on the memory matching process enormously in this example, and this is probably similar in images of natural scenes which are generally complex.

Note that the advantage of low-level processes does not depend on their ability to find a unique solution for each parsing problem. On the contrary, the results on border ownership coding sketched above indicate that the visual cortex can simultaneously represent alternative solutions. A representation like that of area V2 which contains on the order of 100 million cells has an enormous capacity. V1 and V2 can represent more than one orientation per image point (see, for example, simultaneous representation of two superimposed gratings in V1 (Movshon et al., 1985)). Therefore, it is plausible to hypothesize that the visual cortex can represent several alternative parsing results from which cognitive routines can select. Compared to the huge numbers of possible interpretations of an image this would still be a very specific representation.

There are theoretical and empirical arguments for this hypothesis. On the theoretical side it was pointed out that it would be a disadvantage if the system would lock in on one solution, given the ambiguous information available at the pre-cognitive level (principle of least commitment, Marr, 1982). Experimental studies of binocular rivalry show that the duration of dominance of the stimulus in one eye depends on the suppressed stimulus in the other eye (Levelt, 1968). Accordingly, the unperceived stimulus in rivalry is represented in the neural activity in visual cortex (Logothetis and Schall, 1989; Leopold and Logothetis, 1996).

6. Conclusion

The studies reviewed above demonstrate a wealth of processes of visual organization in the cortex. While it was previously thought that contour mechanisms serve to make contrast borders explicit and “fill in the gaps”, the recent findings indicate that cortical processing goes beyond representation and completion of contrast borders. Besides border ownership, 3D surface organization influences perceptual grouping of elemental features in many ways, and modulates responses of V1 and V2 neurons accordingly (Zipser et al., 1996; Bakin et al., 2000). Changes in 3D surface organization that alter perception of border ownership can also influence the perceived direction of motion, and this influence can be traced in the motion signals of area MT (Duncan et al., 2000). The goal of this processing apparently is to make the 3D structure of a scene explicit, specifically, to represent occluding contours and how they belong to surfaces. Assigning contours to surfaces is perhaps the most important first step on the way from image to object representation, because it specifies the surfaces in the scene and their ordering in depth.

The most surprising aspect of the new findings is that neural signals at early cortical levels, despite the small size of the conventional receptive fields of the cells, reflect information about the global image context. Assigning border ownership generally requires the application of algorithms that involve image regions of at least the size of the projection of the object to be processed. The same is true for labeling textured regions as figure or ground (Lamme, 1995; Zipser et al., 1996; Lee et al., 1998). These algorithms might implement gestalt rules such as prägnanz of form, contour closure, and common fate, and the rules that relate patterns of junctions to 3D shape and layout of objects (Waltz, 1975; Adelson, 1993; Nakayama et al., 1995).

Thus, figure-ground segregation and assignment of border ownership are evidence for image parsing mechanisms and suggest networks of global scope. The organization of visual representations in terms of 3D surfaces may be a first stage in the construction of object

representations (Marr, 1982), or may provide a structure for specific processing of selected detail information which is required by most visual tasks, when attention is directed to a specific object in a scene, or to a specific feature of an object (He and Nakayama, 1992; Nakayama et al., 1995; He and Ooi, 2000). Perhaps the same networks that achieve image parsing serve also for the selection of information.

The exact nature of those image parsing mechanisms still needs to be clarified. What kind of cues are used, and how are they combined? What is the output of this stage? Probably the output is not a globally coherent representation of 3D relationships (because perception is not globally coherent; see the Penrose triangle and the work of artist M.C. Escher), but rather a patchwork of coherent domains. How large are these domains? Does the parsing stage provide a unique solution for a given image (or domain), or multiple solutions? The mechanisms might either home in on the most likely interpretation of the given optical information, or pass on several choices to the next stage. The results sketched out above on the combination of stereoscopic depth and monocular form cues suggest that V2 can represent multiple solutions in parallel, leaving the final decision to a later stage.

It has often been pointed out that vision (and any perception) is highly selective. An enormous amount of information streams in through the eyes, but only little is used. A common demonstration are the “two pictures with ten differences”. Viewers generally have to look back and forth many times between the two pictures before they find the differences. This shows that the amount of information the system can store and compare between two looks is only a small fraction of the total, even in those relatively simple drawings. This phenomenon of “change blindness” which has been documented by formal experiments (Rensink et al., 1997; see also Simons and Silverman, Chapter xx, this volume) has led to the conclusion that the visual system processes only the selected information and does not represent much more than the image. Whatever information is needed at a given moment, it is thought, can be retrieved from the image representation. However, psychological tests can only reveal processing that leads either to motor response or to a retrievable representation in memory. As I have argued, this is only the tip of an iceberg. The recordings from the visual cortex reveal sophisticated processing of gigantic amounts of information. In the experiments on border ownership coding, for example, the animal was attending to a visual task at the fixation spot and thereby -- probably -- trying to ignore all other visual stimuli as much as possible, but border ownership processing was nevertheless obvious in half of the signals recorded in V2. Thus, it seems that many locations of the retinal image are processed automatically and in parallel all the time. Whenever our eyes saccade to a new point of fixation, area V2 recomputes the figure-ground relationships, parsing a new image into object-like chunks of information. All this occurs 3-4 times per second on average. Without this extensive preprocessing the system would not be able to select information as efficiently as it does. The relatively large size of the early visual areas V1, V2, and V4, is certainly related to the computational difficulty of the image parsing task which still defies the power of supercomputers. The imperceptible function of these areas makes vision appear effortless.

References

1. Adelson EH (1993) Perceptual organization and the judgment of brightness. *Science* 262: 2042-2044.
2. Attneave F (1954) Some informational aspects of visual perception. *Psychol Rev* 61: 183-193.
3. Bakin JS, Nakayama K, Gilbert CD (2000) Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J Neurosci* 20: 8188-8198.
- *4. Barlow HB (1961) Possible principles underlying the transformations of sensory messages. In: *Sensory Communication* (Rosenblith WA, ed), pp 217-257. Cambridge: MIT Press.
5. Baumann R, van der Zwan R, Peterhans E (1997) Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. *Eur J Neuroscience* 9: 1290-1303.
6. Bravo M, Blake R, Morrison S (1988) Cats see subjective contours. *Vision Res* 28: 861-865.
7. Burkhalter A, Van Essen DC (1986) Processing of color, form and disparity information in visual areas VP and V2 of ventral extrastriate cortex in the macaque monkey. *J Neuroscience* 6: 2327-2351.
8. Coren S (1972) Subjective contours and apparent depth. *Psychological Review* 79: 359-367.
9. Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. *Annu Rev Neurosci* 24: 203-238.
10. De Valois RL, Yund EW, Hepler N (1982) The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res* 22: 531-544.
11. De Weerd P, Vandenbussche E, Debruyn B, Orban GA (1990) Illusory Contour Orientation Discrimination in the Cat. *Behavioural Brain Research* 39: 1-17.
12. Desimone R, Schein SJ, Moran J, Ungerleider LG (1985) Contour, color and shape analysis beyond the striate cortex. *Vision Res* 25: 441-452.
13. Duncan RO, Albright TD, Stoner GR (2000) Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. *J Neurosci* 20: 5885-5897.
14. Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* 1: 1-47.

15. Finkel LH, Sajda P (1992) Object discrimination based on depth-from-occlusion. *Neural Computation* 4: 901-921.
16. Grossf DH, Shapley RM, Hawken MJ (1993) Macaque-v1 neurons can signal illusory contours. *Nature* 365: 550-552.
17. Grossberg S, Mingolla E (1985) Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological Review* 92: 173-211.
18. He ZJ, Nakayama K (1992) Surfaces versus features in visual search. *Nature* 359: 231-233.
19. He ZJ, Ooi TL (2000) Perceiving binocular depth with reference to a common surface. *Perception* 29: 1313-1334.
20. Heitger F, von der Heydt R, Peterhans E, Rosenthaler L, Kübler O (1998) Simulation of neural contour mechanisms: representing anomalous contours. *Image and Vision Computing* 16: 409-423.
21. Helmholtz Hv (1866) *Handbuch der physiologischen Optik*. Hamburg: Voss.
22. Julesz B (1971) *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press.
- *23. Kanizsa G (1979) *Organization in Vision. Essays on Gestalt Perception*. New York: Praeger.
24. Koffka K (1935) *Principles of Gestalt Psychology*. New York: Harcourt, Brace and World.
25. Lamme VAF (1995) The neurophysiology of figure-ground segregation in primary visual cortex. *J Neuroscience* 15: 1605-1615.
26. Lee TS, Mumford D, Romero R, Lamme VAF (1998) The role of the primary visual cortex in higher level vision. *Vision Research* 38: 2429-2454.
27. Lee TS, Nguyen M (2001) Dynamics of subjective contour formation in the early visual cortex. *Proc Natl Acad Sci U S A* 98: 1907-1911.
28. Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkeys percepts during binocular rivalry. *Nature* 379: 549-553.
29. Levelt WJM (1968) *On binocular rivalry*. The Hague: Mouton.
30. Leventhal AG, Thompson KG, Liu D, Zhou Y, Ault SJ (1995) Concomitant sensitivity to orientation, direction, and color of cells in layers 2, 3, and 4 of monkey striate cortex. *Journal of Neuroscience* 15: 1808-1818.
31. Logothetis NK, Schall JD (1989) Neuronal correlates of subjective visual perception. *Science* 245: 761-763.

- *32. Marr D (1982) Vision. A computational investigation into the human representation and processing of visual information. San Francisco: Freeman.
- 33. Movshon JA, Adelson EH, Gizzi MS, Newsome WT (1985) The analysis of moving patterns. *Exp Brain Res suppl.* 11: 117-151.
- *34. Nakayama K, He ZJ, Shimojo S (1995) Visual surface representation: a critical link between lower-level and higher-level vision. In: *Invitation to cognitive science* (Kosslyn SM, Osherson DN, eds), pp 1-70. Cambridge, MA: MIT.
- 35. Nieder A (2002) Seeing more than meets the eye: processing of illusory contours in animals. (In press).
- 36. Nieder A, Wagner H (1999) Perception and neuronal coding of subjective contours in the owl. *Nat Neurosci* 2: 660-663.
- 37. Paradiso MA, Shimojo S, Nakayama K (1989) Subjective contours, tilt aftereffects, and visual cortical organization. *Vision Res* 29: 1205-1213.
- 38. Peterhans E, von der Heydt R (1989) Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *J Neuroscience* 9: 1749-1763.
- 39. Peterhans E, von der Heydt R, Baumgartner G (1986) Neuronal responses to illusory contour stimuli reveal stages of visual cortical processing. In: *Visual Neuroscience* (Pettigrew JD, Sanderson KJ, Levick WR, eds), pp 343-351. Cambridge: Cambridge University Press.
- 40. Poggio GF (1995) Mechanisms of stereopsis in monkey visual cortex. *Cerebral Cortex* 3: 193-204.
- 41. Poggio T, Koch C (1985) Ill-posed problems in early vision: from computational theory to analogue networks. *Proc R Soc Lond B* 226: 303-323.
- 42. Qiu FT, Macuda TJ, von der Heydt R (2001) Neural coding of border ownership: figural and stereoscopic cues. *Soc Neurosci Abstr* 27: Program No. 165.6.
- 43. Ramsden BM, Hung CP, Roe AW (2001) Real and illusory contour processing in area V1 of the primate: a cortical balancing act. *Cereb Cortex* 11: 648-665.
- 44. Redies C, Crook JM, Creutzfeldt OD (1986) Neuronal responses to borders with and without luminance gradients in cat visual cortex and dorsal lateral geniculate nucleus. *Exp Brain Res* 61: 469-481.
- 45. Rensink RA, O'Regan JK, Clark JJ (1997) To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science* 8: 368-373.
- 46. Ringach DL, Shapley R (1996) Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Res* 36: 3037-3050.

47. Rubin E (1921) Visuell wahrgenommene Figuren. Copenhagen: Gyldendal.
48. Schein SJ, Desimone R (1990) Spectral properties of V4 neurons in the macaque. *J Neuroscience* 10: 3369-89.
49. Sheth BR, Sharma J, Rao SC, Sur M (1996) Orientation maps of subjective contours in visual cortex. *Science* 274: 2110-2115.
50. Soriano M, Spillmann L, Bach M (1996) The abutting grating illusion. *Vision Res* 36: 109-116.
51. Srinivasan M, Lehrer M, Wehner R (1987) Bees perceive illusory contours induced by movement. *Vision Res* 27: 1285-1290.
52. Ullman S (1996) *High-level Vision*. Cambridge, Mass.: MIT press.
53. von der Heydt R, Peterhans E (1989) Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *J Neuroscience* 9: 1731-1748.
- *54. von der Heydt R, Peterhans E, Baumgartner G (1984) Illusory contours and cortical neuron responses. *Science* 224: 1260-1262.
55. von der Heydt R, Zhou H, Friedman HS (2000) Representation of stereoscopic edges in monkey visual cortex. *Vision Research* 40: 1955-1967.
56. Waltz DI (1975) Understanding line drawings of scenes with shadows. In: *The Psychology of Computer Vision* (Winston PH, ed), pp 19-91. New York: McGraw-Hill.
57. Westheimer G, Li W (1996) Classifying illusory contours by means of orientation discrimination. *J Neurophysiology* 75: 523-528.
58. Zeki SM (1978) Uniformity and diversity of structure and function in rhesus monkey prestriate visual cortex. *J Physiol (Lond)* 277: 273-290.
- *59. Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. *J Neuroscience* 20: 6594-6611.
60. Zipser K, Lamme VAF, Schiller PH (1996) Contextual modulation in primary visual cortex. *J Neuroscience* 16: 7376-7389.

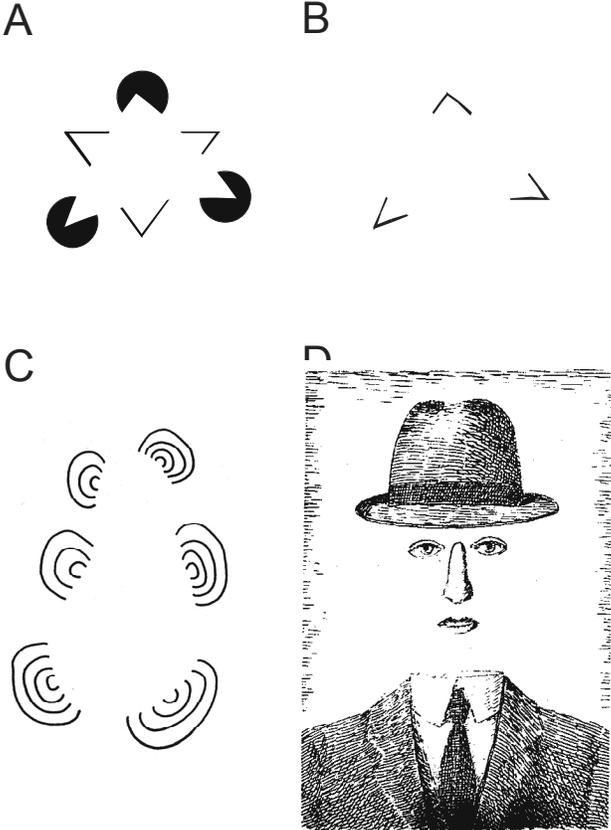


Fig. 1. Perception of illusory contours. A and C, after Kanizsa (1979). D, drawing by René Magritte.

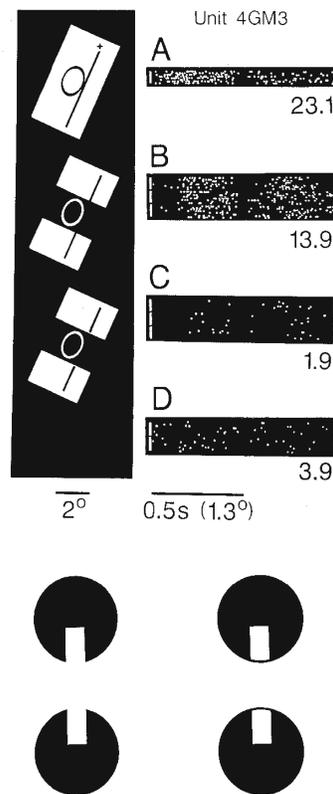


Fig. 2. Illusory contour responses in a neuron of area V2. Each line of dots in the raster plots represents a sequence of action potentials fired in response to the stimulus shown on the left. A, Responses to a moving dark bar, B, to a figure in which a moving illusory bar is perceived, C, to a modified figure in which the illusion is abolished by adding line segments. Note the reduction of responses. Figures at the bottom illustrate the perceptual effect of adding lines. D, spontaneous activity. Ellipses indicate the minimum response field of the neuron (that is, the minimum region outside which a bar does not evoke a response), cross indicates fixation point. (From Peterhans and von der Heydt, 1989, with permission)

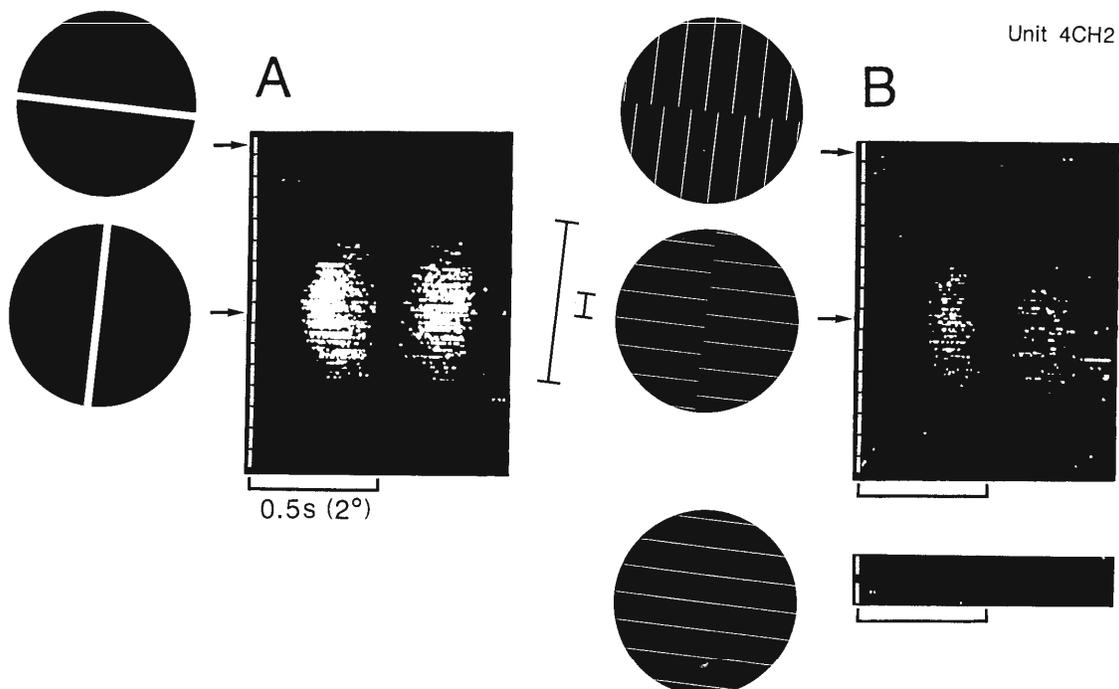


Fig. 3. Illusory contour responses in another neuron of V2. In A, bars, in B, the border between two gratings, were moved across the receptive field at 16 different orientations spanning 180°. The neuron responds at the same orientations for bar and illusory contour. Bottom right, control: grating without border of discontinuity. (Modified from von der Heydt and Peterhans, 1989, with permission)

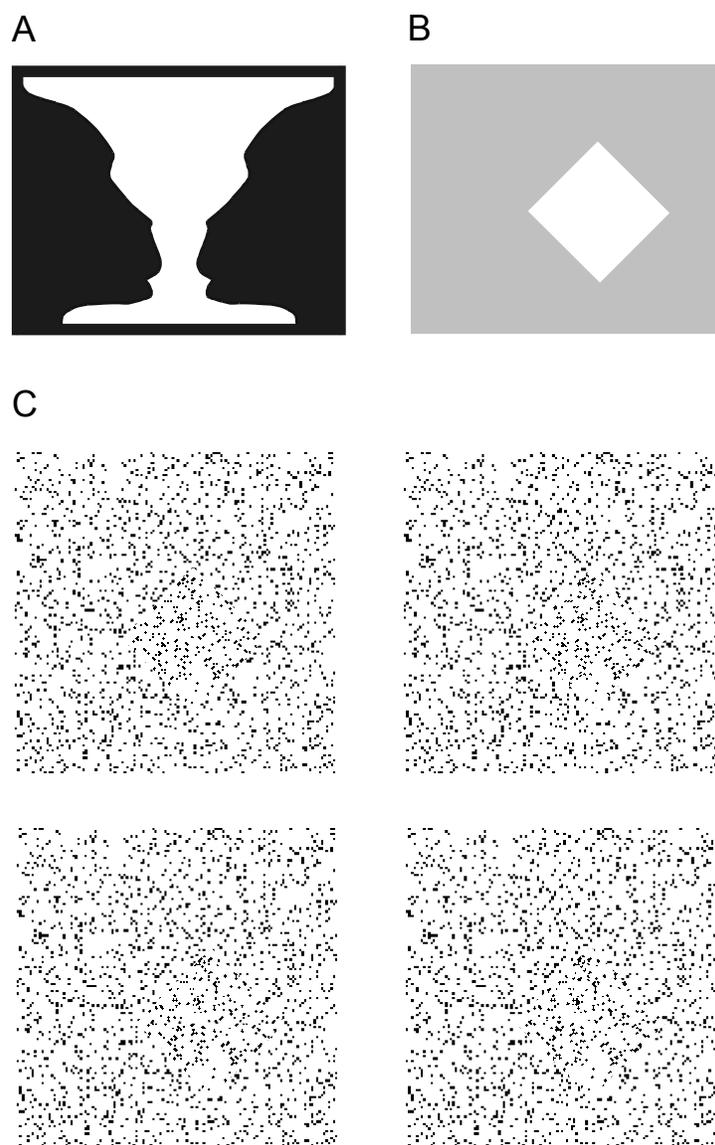


Fig. 4. Perception of border ownership. A, Physiologist's version of Rubin's vase. The black-white borders are perceived either as contours of a vase, or as contours of faces. B, White square. The contrast borders are generally perceived as contours of the square. C, Stereograms. Left and right textured square fields can be fused, for example, by squinting (try crossing the lines of sight of the two eyes until three fields are perceived instead of two; the center field then shows the result of binocular fusion). On fusion with crossed eyes, the top pair shows a square figure, while the bottom pair shows a square window. In the former, the 3D edges belong to the figure, in the latter, to the surround.

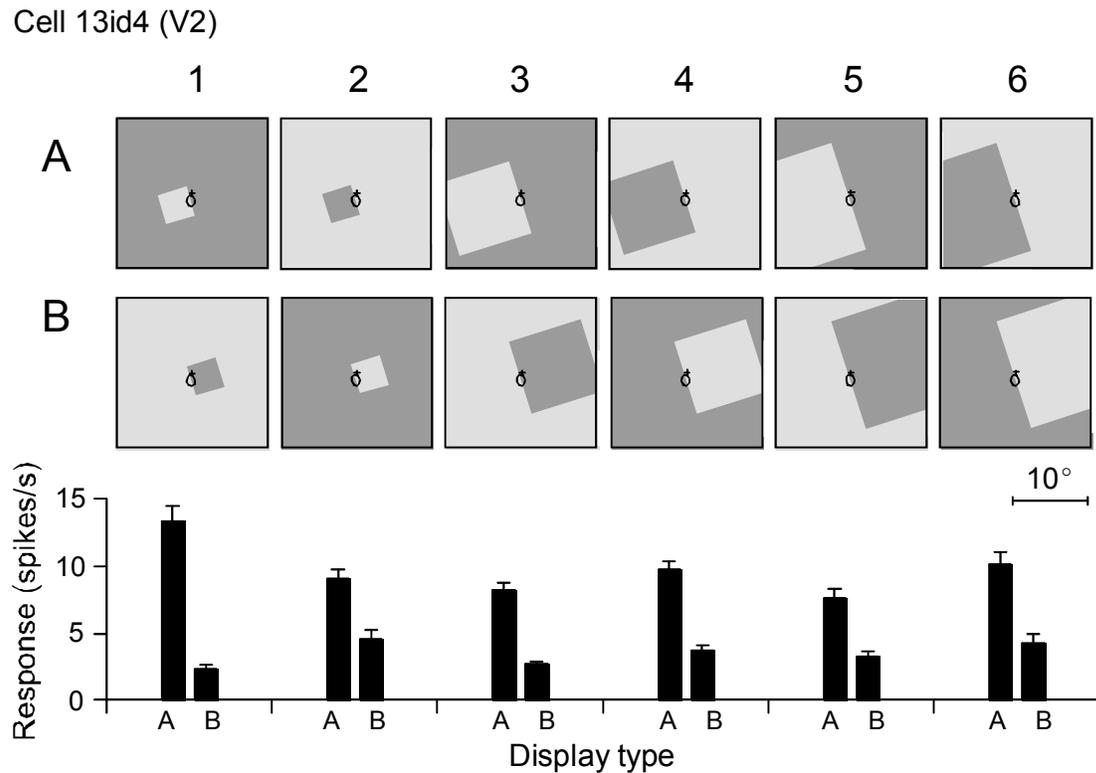


Fig. 5. Selectivity for side of figure in a neuron of area V2. Edges of squares were tested so that the square was either on the left side (A) or on the right side (B) of the receptive field (ellipses show minimum response field, cross marks fixation point). Note that corresponding displays in A and B are identical over the combined area of the two squares. Tests with square sizes of 4° , 10° and 15° are shown. Bar graph represents mean firing rates with S.E. In every case, the neuron responds more strongly when the figure is on the left side, despite locally identical stimulation. (From Zhou et al., 2000, with permission)

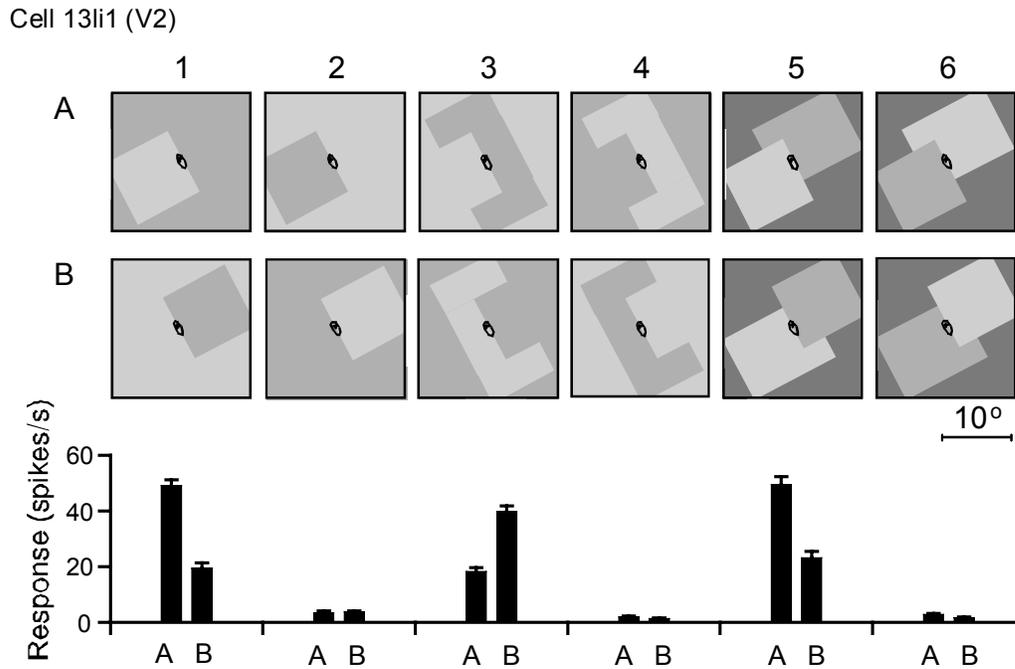


Fig. 6. Example of a V2 neuron tested with squares, C-shaped figures, and overlapping figures. The neuron was color selective with a preference for violet (depicted here as light gray). In the test with single squares (1-2) the neuron is selective for side of figure and local contrast polarity, responding best to the edge of a violet square located on the lower left hand side of the receptive field (A1). With C-shaped figures (3-4), the neuron responds better to B3 than A3, pointing to the lower left as the figure side, in agreement with perception. With overlapping figures (5-6), the neuron responds better to A5 than B5, assigning the edge to the figure that is perceived as overlaying. (From Zhou et al., 2000, with permission)

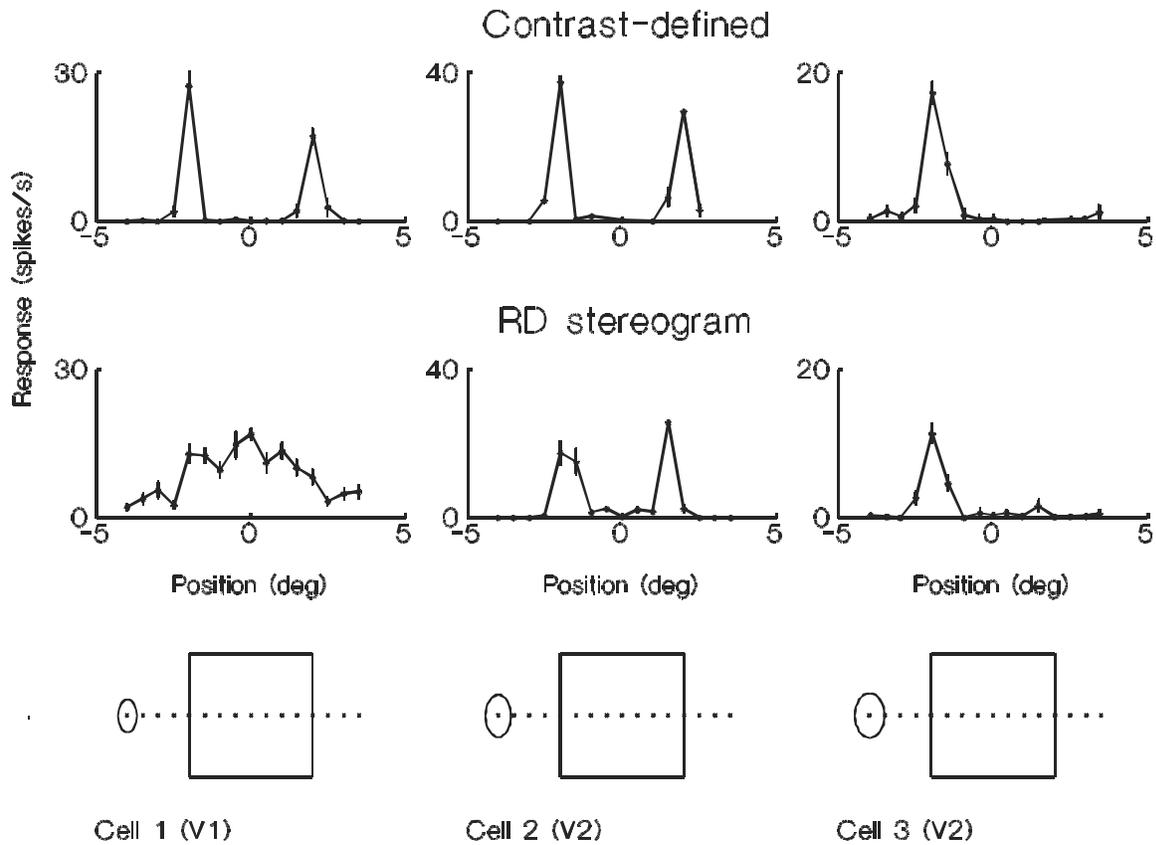


Fig. 7. Stereoscopic edge selectivity. Responses of a neuron of V1 (left) and two neurons of V2. Contrast-defined squares are compared with squares portrayed in random-dot stereograms (Julesz, 1971). Each figure is presented at 16 different positions relative to the receptive field, as shown schematically at the bottom. Small rectangles represent the minimum response fields, the large squares represent the test figures. Data points in the graphs correspond to the dotted positions below. The plots show that cell 1 is edge selective for contrast-defined figures, but responds all over the surface of the figure for random-dot stereograms, whereas cells 2-3 are edge selective for both types of figures. Most of these cells respond asymmetrically to the two sides like cell 3. (From von der Heydt et al., 2000, with permission)

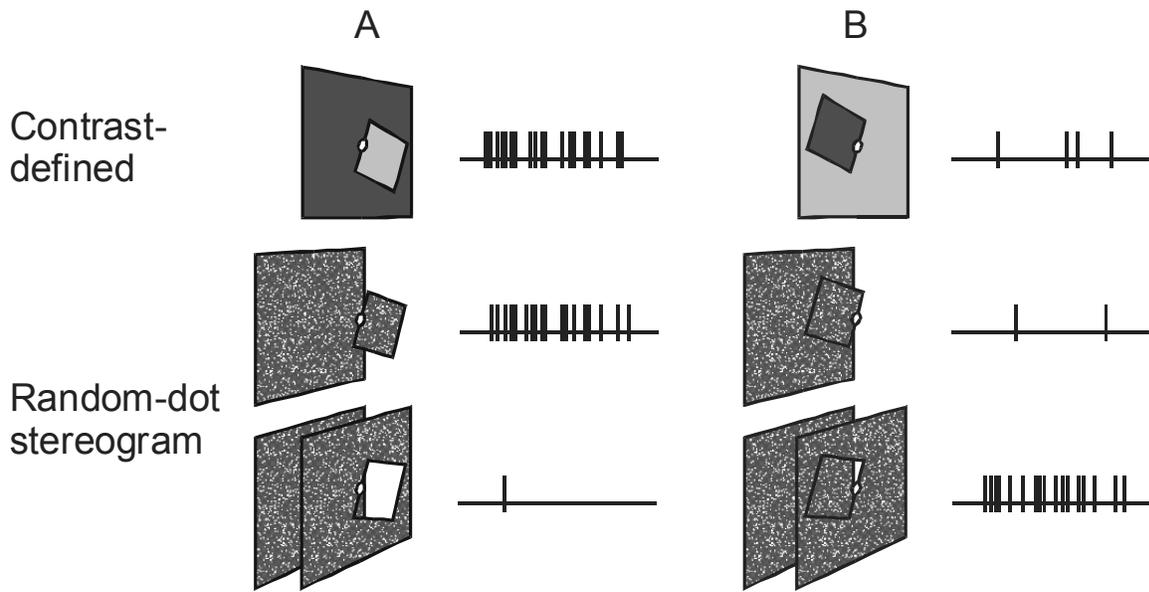


Fig. 8. A critical test of the hypothesis of border ownership coding. Side-of-figure preference and stereo edge preference are assessed for each single neuron. For contrast-defined squares, the “figure” interpretation is perceptually more compelling than the “window” interpretation. Therefore, if neurons code for border ownership, the preferred “near” side should be also the preferred side of figure. This fictitious neuron prefers the nearer surface to the right for the stereogram (no matter whether a square is displayed or a window), and prefers figure to the right (A) for the contrast-defined display, in accordance with the hypothesis.

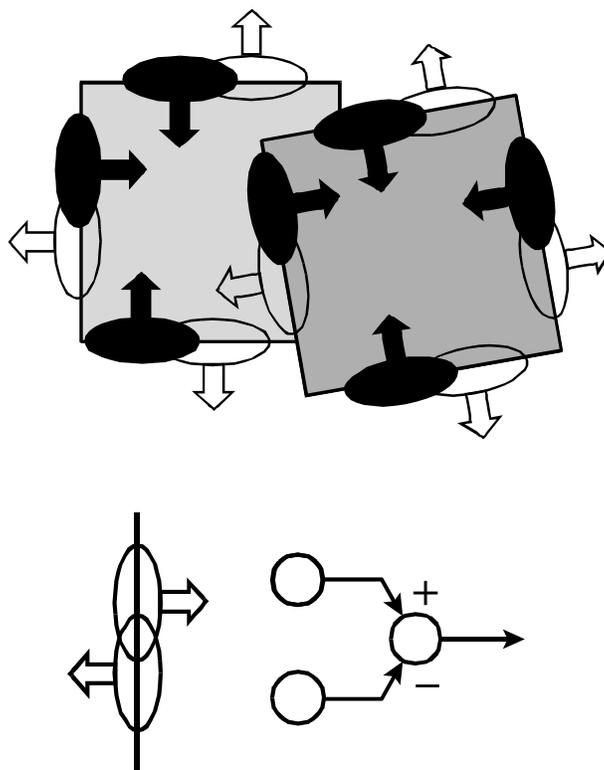


Fig. 9. Schematic illustration of the cortical representation of a pair of overlapping squares. Ellipses indicate receptive fields. Each piece of contour is represented by two pools of orientation selective neurons with opposite border ownership preference as indicated by arrows. Filled symbols indicate the neurons whose activity would be enhanced for this stimulus. Border ownership is thought to be encoded in the relative activation of the two pools.

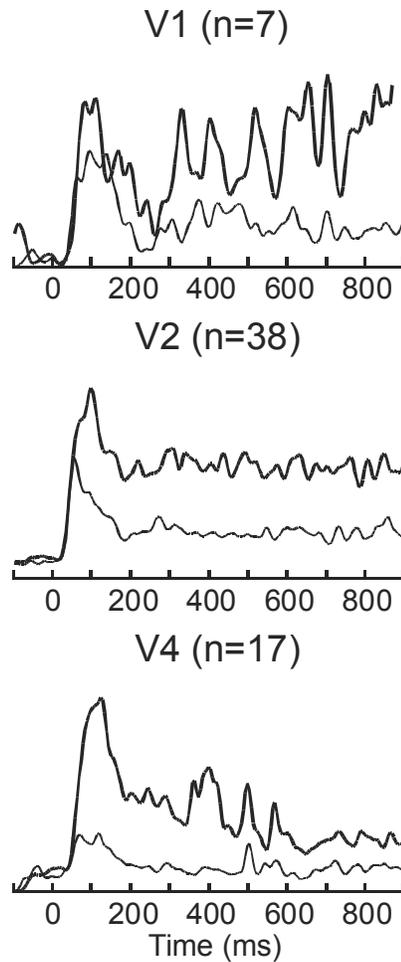


Fig. 10. The time course of the border ownership signal. The figure shows the averaged normalized responses of neurons in 3 cortical areas. Squares of 4° or 6° size were presented as in Fig. 5. Zero on the time scale refers to the onset of the display. Thick and thin lines represent responses to preferred and non-preferred sides, averaged over both contrast polarities. The delay between onset of response and differentiation of side of figure was less than 25 ms. (From Zhou et al., 2000, with permission)

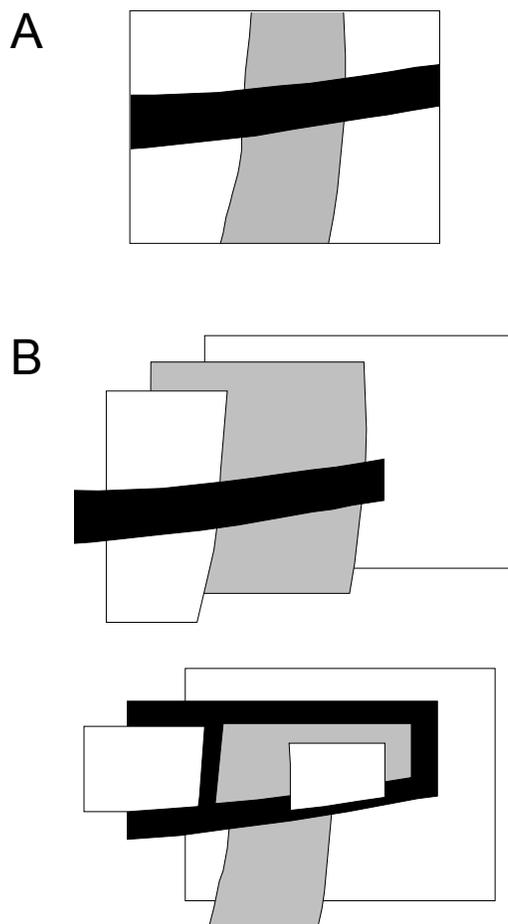


Fig. 11. The interpretation of the overlay structure of images is highly ambiguous. A, Example of a display that would generally be perceived as two elongated objects occluding one another. B, Two of about 1,000 possible interpretations of A.