

Figure and Ground in the Visual Cortex: V2 Combines Stereoscopic Cues with Gestalt Rules

Fangtu T. Qiu and Rüdiger von der Heydt*
Krieger Mind/Brain Institute and
Department of Neuroscience
Johns Hopkins University
3400 North Charles Street
Baltimore, Maryland 21218

Summary

Figure-ground organization is a process by which the visual system identifies some image regions as foreground and others as background, inferring 3D layout from 2D displays. A recent study reported that edge responses of neurons in area V2 are selective for side-of-figure, suggesting that figure-ground organization is encoded in the contour signals (border ownership coding). Here, we show that area V2 combines two strategies of computation, one that exploits binocular stereoscopic information for the definition of local depth order, and another that exploits the global configuration of contours (Gestalt factors). These are combined in single neurons so that the “near” side of the preferred 3D edge generally coincides with the preferred side-of-figure in 2D displays. Thus, area V2 represents the borders of 2D figures as edges of surfaces, as if the figures were objects in 3D space. Even in 3D displays, Gestalt factors influence the responses and can enhance or null the stereoscopic depth information.

Introduction

We perceive the world in three dimensions, although our eyes register only two-dimensional images. These images are generally cluttered, because objects occlude one another, and surfaces that are widely separated in space are projected onto adjacent image regions (Figure 1A). Thus, a fundamental task of vision is to identify the borders between image regions that correspond to different objects. These borders, also termed “occluding contours,” carry information about the form of the occluding object but are generally not related to the background objects. For example, the border between the dark and mid-gray regions in Figure 1A defines the shape of the lighter tree in the foreground, but not the shape of the partly occluded darker tree. Somehow, the brain immediately “knows” that the object corresponding to the darker region extends behind the lighter region and consequently registers the darker tree as a more or less symmetrical shape and not as a banana-shaped object (the actual form of the dark gray region). Thus, the task of vision is not only to detect the occluding contours, but also to assign them correctly to the occluding objects.

It might be thought that this perceptual interpretation is only possible because the image contains familiar

shapes of objects. Yet psychologists in the early twentieth century argued that mechanisms of figure-ground organization exist that work automatically, and independently of the observer’s knowledge and expectation (Koffka, 1935; Rubin, 1921; Rubin, 2001; Wertheimer, 1923; Wertheimer, 2001; for a review, see Spillmann and Ehrenstein, 2003). Indeed, figure-ground perception can be manipulated experimentally by providing specific cues that define the depth relationships explicitly, for example, by means of stereograms. Under these conditions, the perception of form and recognition of objects is dramatically affected when the depth ordering between regions is altered (Nakayama et al., 1989). This indicates that assignment of border ownership precedes the recognition process.

Single-cell recordings show that stereoscopic cues contribute to the cortical representation of contours in many ways. Some of the neurons in area V2 that signal location and orientation of luminance contours respond also to disparity-defined contours created by “random-dot stereograms” (RDS) and represent the depth ordering of surfaces (von der Heydt et al., 2000). Binocular disparity influences the representation of contours in V1 and V2 (Bakin et al., 2000; Heider et al., 2002; Sugita, 1999) and affects motion signals in area MT (Duncan et al., 2000) in ways that parallel perceptual figure-ground organization. Illusory contour signals depend on occlusion cues that might also be used for assigning figure and ground (Baumann et al., 1997; von der Heydt et al., 1993). Thus, depth cues profoundly influence the neural visual representation at early cortical levels.

The phenomenon of figure-ground organization in the absence of specific depth cues is still a mystery. Why is the white square in Figure 1B generally perceived as an object *in front of* a dark background rather than a window in a dark screen, or simply a lightly pigmented patch of surface surrounded by a darker pigmented region? The borders between light and dark are interpreted as the edges of an occluding object. Apparently, the system assigns border ownership despite the absence of depth cues, by using criteria such as compact shape or the global configuration of contours (closure, “surroundedness”), or perhaps by identifying familiar shapes (in this case a square). Without implying a specific theory, we refer to this phenomenon as Gestalt-based figure-ground organization.

Neural correlates of Gestalt-based figure-ground organization were recently discovered at early levels in the visual cortex (Lamme, 1995; Lee et al., 1998; Zhou et al., 2000; Zipser et al., 1996; but see Rossi et al., 2001). Lamme and colleagues found enhancement of texture-evoked activity in figure regions compared to the ground region in neurons of V1. Zhou et al. found that neural edge responses were selective for the side of the figure to which the edge “belonged” (see below). This phenomenon was more pronounced in V2 and V4 than in V1. Remarkable about these findings is that neurons at these early levels integrate the image context far beyond the classical receptive field (for a review, see Albright and Stoner, 2002).

*Correspondence: von.der.heydt@jhu.edu

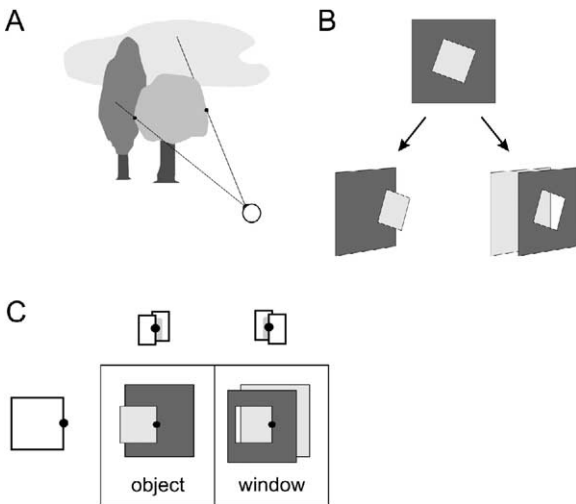


Figure 1. The Problem of Interpreting 2D Images in Terms of Objects in a 3D World

Images are composed of regions that correspond to objects in space (A). The boundaries of these regions are generally the contours of objects that occlude more distant parts of the scene (occluding contours). To interpret images successfully, the visual system has to detect these contours and link them to the occluding regions. (B) The light textured region is generally perceived as a tilted square on a dark background, and the light-dark border as the contour of the square. But the display is ambiguous: the square could be a window. (C) The concept of border ownership. The interpretation of a 2D display depends on how the contrast borders are assigned (top). Consider the border marked by a black dot: if the border is assigned left, the square is an object in front of a dark background; if the border is assigned right, the square becomes a piece of background that is seen through a window. Given flat displays without depth cues, the visual system assumes the object interpretation.

The selectivity for side-of-figure of neurons might be just a random asymmetry of receptive fields. If it indeed reflects the process of figure-ground segregation as hypothesized (Zhou et al., 2000), then these neurons should also respond to stereoscopically defined 3D edges and be selective for depth order. For example, a neuron with a preference for figure-to-the-left (Figure 1C; black dot indicates receptive field) should respond to edges in which the surface to the left of the receptive field is nearer than the surface to the right, because this is so for objects in 3D space, but the neuron should not respond to edges of the opposite depth order, because a left-far edge can only occur if the figure is a window. Zhou et al. presented two examples of cells in which the preferred side-of-figure in fact coincided with the “near” side of the preferred depth order. Finding this in two cells could have been a coincidence. The question of whether the visual cortex systematically combines stereoscopic cues with Gestalt-based criteria and how it does this remained open. Is there a statistical association between both kinds of cues, and if so, how strong is it? Are Gestalt cues comparable to “real” depth cues such as binocular disparity? How do neurons respond if Gestalt cues contradict the binocular information?

In the present study, we have investigated the interplay between stereoscopic cues and Gestalt cues in

the visual cortex quantitatively. The results show that there is a robust tendency to combine these different sources of information according to the rule that a compact shape corresponds to an object in 3D space. Experiments with combinations of cues show that Gestalt factors influence the border ownership signal even when explicit depth information is available.

Results

Two main experiments were performed. The aim of experiment 1 was to determine if side-of-figure preference and stereoscopic edge preference are combined in a systematic way in single neurons. The two hypothetical mechanisms were tested separately: side-of-figure selectivity was determined with contrast-defined figures, which do not provide depth cues, and stereo-edge selectivity was determined with RDS, which define depth but are devoid of contrast-defined form. In experiment 2, depth and Gestalt cues were combined, and synergistic and antagonistic combinations were tested to see how the cues interact.

Additional experiments were performed on a subset of the neurons to establish size invariance of the Gestalt effect, and position invariance of 3D edge selectivity. We will begin by discussing these results in the following two sections, because they serve well to explain the basic findings of side-of-figure selectivity and stereo edge selectivity. We will then present the results of the main experiments and some controls.

Side-of-Figure Selectivity

A fraction of the orientation-selective neurons in macaque area V2 signal not only the location and orientation of luminance and color edges, but also the location of the figure to which an edge “belongs” (Zhou et al., 2000). Figure 2A illustrates a V2 neuron that responds more strongly to the bottom edge of a light square than to the top edge of a dark square, although the edge in the receptive field is the same. Note that the left and right displays in Figure 2A are indistinguishable over the entire region occupied by the two squares (dashed line in Figure 2B) and that information about the side of the figure can only come from outside that region. Thus, despite its small receptive field (black ellipse), the neuron apparently processes a large image context. As can be seen in Figure 2B, the size of the square determines the distance over which context signals need to be integrated to determine the location of the figure. Cells were tested with two sizes of squares, 3° and 8° visual angle, and two contrast polarities, and the side-of-figure effect was quantified by the response modulation index, taking the preferred side for the 3° figure as reference (see next section for definition of the index). This index is plotted in Figure 2C for all cells in which the effect for the 3° figure was significant ($p < 0.05$, analysis of variance [ANOVA]). The points corresponding to the same neuron are connected by lines. It can be seen that most cells (27 of 33) showed same side preference for the 8° figure as for the 3° figure. Zhou et al. found consistent side selectivity for figures that spanned up to 20° of visual angle. This range of context integration is huge compared to the small size of the

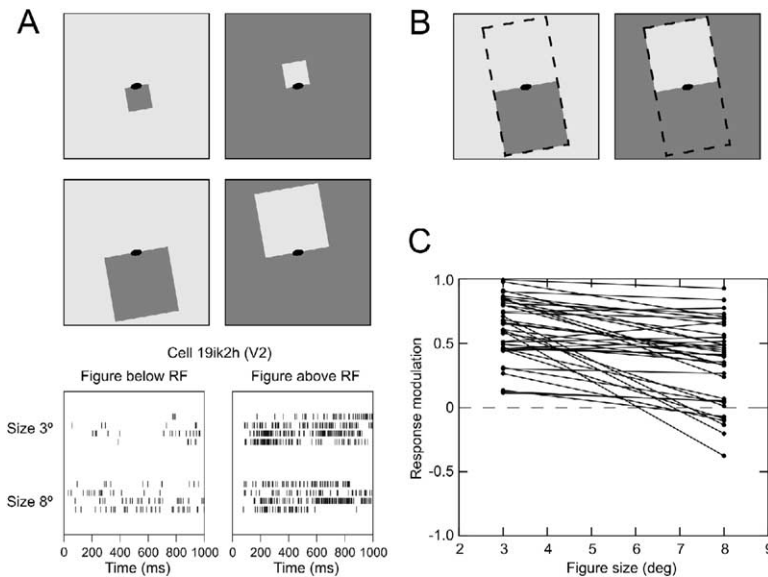


Figure 2. Side-of-Figure Selectivity

(A) Responses of a V2 neuron to the same local contrast border forming either the top edge of a dark square, or the bottom edge of a light square. Squares of two sizes were tested (3° and 8° visual angle). Displays with the reversed contrast were also tested but are not illustrated. Ellipses show size of minimum response field. Despite the same local stimulation—the juxtaposed displays are indistinguishable over the regions delineated with dashed lines in (B)—the firing rate is higher for figure above than for figure below. In (C), the response modulation index for preferred versus nonpreferred side is plotted as a function of square size for 33 V2 neurons that were side-of-figure selective for a 3° square ($p < 0.05$, ANOVA). Lines connect points corresponding to the same neuron. It can be seen that most of the neurons have a positive modulation index also for the 8° square, indicating mechanisms of global form processing. The finding of side-of-figure selectivity in neurons suggests the existence of cortical mechanisms that use Gestalt

rules to determine which region might be an object and which might be background, such as compact shape, closed contour, and the fact that the square is surrounded by a region of uniform color (Rubin, 1921). The plot in (C) also shows that smaller squares tended to produce stronger side-of-figure modulation than larger squares, corresponding to the Gestalt rule that smaller regions have a stronger tendency to be perceived as figure than larger regions.

“classical receptive field” of V2 neurons, which is only 0.6° on average for the median eccentricity of receptive fields in our sample (Gattass et al., 1981).

Stereoscopic Edge Selectivity

Many neurons in V2 are sensitive to binocular disparity (Poggio et al., 1985), and some respond to stereoscopically defined 3D edges (von der Heydt et al., 2000). The majority of these cells are selective for the orientation of the edge and also for the depth order, that is, which surface is in front and which is in back. Figure 3 illustrates this selectivity for three V2 neurons. Disparity-defined edges were created by RDS. The disparity of one surface was set to the preferred disparity of the neuron (or zero if there was no clear tuning), and the other surface was placed behind it at a distance corresponding to 10 or 24 arc min disparity (depending on the eccentricity of the receptive field). The edges were tested in four orientations, as illustrated at the top of Figure 3. (For the purpose of illustration, the preferred orientation was assumed to be vertical; hatching indicates the nearer of the two surfaces.) To control for effects of stimulus position, each edge was presented at various positions relative to the receptive field, as indicated by the scales. The bar graphs below show the responses as a function of position.

It can be seen that, at the preferred orientation, each neuron responds vigorously to one depth order, but hardly at all to the opposite depth order. For example, the cell in Figure 3A responds to a vertical edge whose right surface is in front, but not at all if the left surface is in front (although the edge is at the same depth in both configurations!). The other two cells have the opposite preference. Note that the preference for one or the other depth order does not depend on the exact position of the edge in the receptive field; at any posi-

tion, the responses to the nonpreferred depth order are much smaller than the maximum response. Also, edges orthogonal to the preferred orientation (horizontal in the figure) produce only weak, erratic responses. Thus, cells in V2 can signal orientation and depth order of 3D edges. Generally, these cells respond to contrast edges as well as to disparity-defined edges and show similar orientation tuning for both (von der Heydt et al., 2000).

Convergence of Gestalt Processing and Stereoscopic Mechanisms in Single Cells

The stereoscopic selectivity of neurons provides a key to understanding the meaning of their signals. If neurons are selective for the depth order of stereoscopic edges, we know that they are involved in the representation of the 3D layout of surfaces, and hence border ownership coding. While contrast-defined displays are generally ambiguous (Figure 1), there is no such ambiguity in RDS, because the depth relations are defined by the binocular disparities; the nearer surface owns the border (Nakayama et al., 1989). Thus, the RDS can be considered as the “gold standard” for border ownership assignment. If the side-of-figure-selective neurons are involved in border ownership coding, they should also be selective for the depth order of edges in RDS. We may not expect to see this in every case, because stereopsis is obviously not indispensable for the perception of border ownership. However, if neurons combining side-of-figure with depth order selectivity exist in significant numbers, and if the depth order preference, in the population, is biased toward the object interpretation (Figure 1C), this would be strong evidence for mechanisms that implement Gestalt rules to infer border ownership.

In experiment 1, we examined the relationship between preferred side-of-figure and preferred depth or-

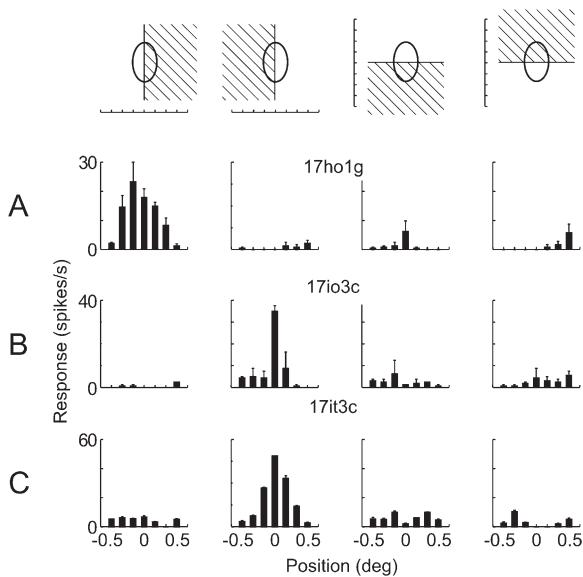


Figure 3. Neural Selectivity for Stereoscopic Edges

Neurons were tested with random-dot stereograms (RDS) portraying a square floating in front of a background plane. An edge of the square was presented in the receptive field (ellipse) at four orientations, as illustrated schematically at the top, where hatching indicates the nearer surface (only one edge of the square is illustrated, because the results of the main experiments showed that the responses depended on the edge in the receptive field, while the global shape had no influence). The preferred orientation is depicted as vertical. Seven positions in steps of 1/6 of a degree of visual angle were tested for each orientation, as indicated by the scales. Bar graphs represent means and standard errors of the responses of typical 3D edge-selective cells of area V2. It can be seen that the neurons responded selectively for only one depth order at the preferred orientation (vertical), either to a far-near step (A) or to near-far steps (B and C). Edges orthogonal to the preferred orientation (horizontal) produced only weak responses. The graphs show that preference for one depth order or the other does not depend on the position of the edge relative to the receptive field.

der of single neurons. Figure 4 illustrates this experiment for a neuron recorded in area V2. The responses to the contrast-defined figures (Figures 4A–4D) show

that the neuron is activated more strongly when the square is located to the left of the receptive field (responses shown in Figures 4A and 4C are stronger than those in Figures 4B and 4D). The test with RDS (Figures 4E–4H) shows that the neuron responds vigorously to the step when the left-hand surface is nearer than the right-hand surface (Figures 4E and 4F), but hardly at all to the reverse step (Figures 4G and 4H). Thus, the neuron associates “figure-left” with “left surface in front,” which is consistent with an interpretation of the contrast-defined square as an object in front of a background. Note also that, in the case of the RDS, the responses are determined by the depth order of the surfaces in the receptive field but are independent of the location of the global shape. Whether the edge was the right-hand edge of a square surface (Figure 4E) or the left-hand edge of a window (Figure 4F) made no difference.

Figure 5 illustrates the results from four other V2 neurons in this experiment. The averaged firing rate is plotted as a function of time after stimulus onset. The plots labeled “Contrast” show the responses to edges of contrast figures: solid line for preferred side, dashed line for nonpreferred side (averaged over both contrast polarities). The plots labeled “RDS” show the responses to 3D steps, and solid lines correspond to steps in which the surface on the preferred figure side was near (the “object” case in Figure 6), whereas dashed lines correspond to steps in which the surface on the preferred figure side was far (the “window” case). It can be seen that, in neurons in Figures 4A–4C, the 3D step that was consistent with the object interpretation evoked the stronger response, while for the neuron in Figure 4D, the 3D step corresponding to the window interpretation was more effective. In each case, the differentiation of side-of-figure and depth order occurred soon after the onset of responses.

This experiment was performed in 251 orientation-selective neurons, 77 from area V1 and 174 from area V2. Figure 6 shows how these neurons combined side-of-figure and 3D step selectivity. The modulation index for side-of-figure,

$$I_{side} = (R_{preferred} - R_{nonpref}) / (R_{preferred} + R_{nonpref}),$$

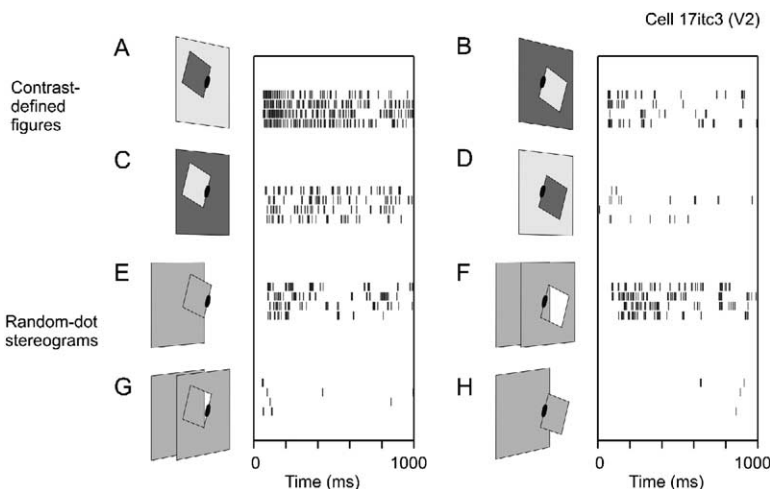


Figure 4. Convergence of Gestalt Mechanisms and Stereoscopic Mechanisms in a Single Neuron

(A–D) Responses to left and right sides of contrast-defined figures. For either contrast polarity of the local edge, figure location left of the receptive field (A and C) produces stronger responses than figure location right of the receptive field (B and D).

(E–H) Responses of the same neuron to 3D step edges produced by random-dot stereograms. The neuron responds more strongly when the surface to the left of the receptive field is in front (E and F) rather than in back (G and H). This combination of side-of-figure preference and 3D step preference is consistent with an object interpretation of the contrast figure (see Figure 1C). Cell recorded in area V2.

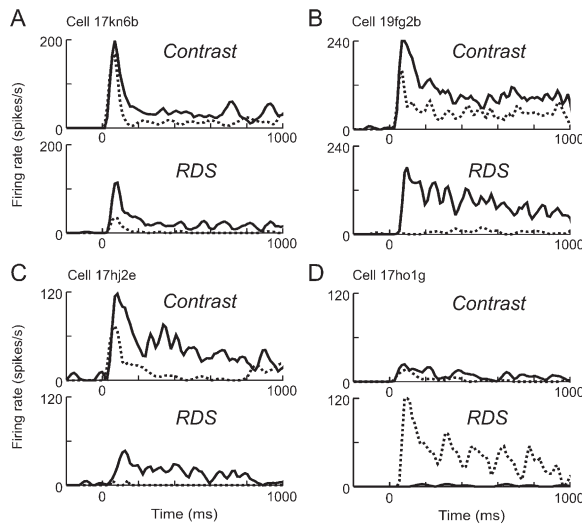


Figure 5. The Responses of Four Other V2 Neurons in the Same Experiment

The graphs show the smoothed mean firing rates as a function of time after stimulus onset. For contrast-defined figures (*Contrast*), solid and dashed lines show the responses for preferred and non-preferred side-of-figure, respectively. For random-dot stereograms (*RDS*), solid lines show the responses to 3D edges with the *near* surface on the preferred side-of-figure, and dashed lines show the responses to edges with the *far* surface on that side. In the neurons shown in (A)–(C), RDS edges with the *near* surface on the preferred figure side produced the greater responses (object interpretation of the figure), while for the neuron shown in (D), RDS edges with the *far* surface on the preferred figure side was optimal (window interpretation of the figure). This neuron (D) is shown here despite its weak responses to contrast borders, because it is the best example we could find for a neuron representing the window interpretation.

where R is mean firing rate, is plotted on the vertical axis, while the horizontal axis shows the corresponding modulation index for depth order:

$$I_{depth} = (R_{pref-near} - R_{pref-far}) / (R_{pref-near} + R_{pref-far}),$$

where *pref-near* and *pref-far* signify the edges whose surface on the preferred side is near, and far, respectively (“preferred side” for the contrast figure). This index is >0 if side-of-figure and step-edge preferences are consistent with an object interpretation of the figure, and <0 if they are consistent with a window interpretation. (The side-of-figure modulation index is always positive, because preferred side was defined as the side associated with the greater response.) Filled symbols indicate cells that were selective for both side-of-figure and depth order ($p < 0.05$ in each case, ANOVA). It can be seen that in the V2 sample (Figure 6, top) cells on the “object” side are more frequent and tend to have higher modulation indices for side-of-figure than cells on the “window” side. Of the 174 neurons tested in area V2, 35% were selective for side-of-figure, 40% were selective for depth order, and 21% were both. Of the latter, 81% (30/37) represented the object interpretation. In area V1 (Figure 6, bottom), only two of 77 neurons tested selective for both side-of-figure and depth order,

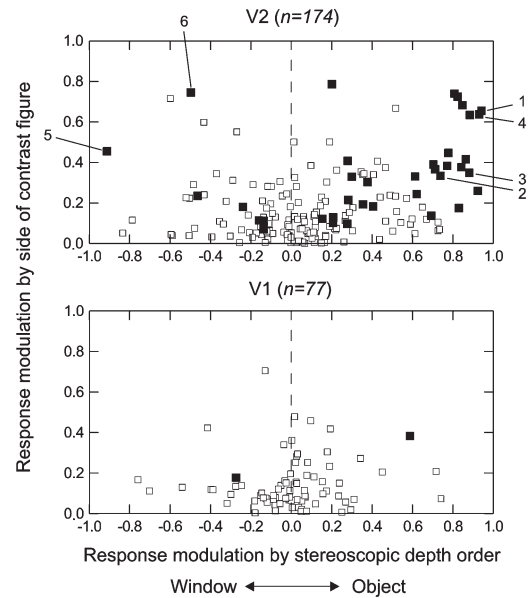


Figure 6. Gestalt-Based and Stereoscopic Figure-Ground Mechanisms in Neurons of Areas V2 and V1

The modulation index for side-of-figure is plotted on the vertical axis, and the modulation index for depth order is plotted on the horizontal axis. Each symbol represents a neuron. A depth order index >0 indicates that 3D edge preference and side-of-figure preference were combined according to the object interpretation of the figure, and an index <0 indicates the window interpretation. Filled symbols indicate neurons with significant response differences in both tests ($p < 0.05$ for each, ANOVA; note that the p value refers to differences in number of spikes; in the modulation indices, these differences are normalized). These neurons were almost exclusively found in V2 and generally represented the object interpretation of the figure.

significantly less than in V2 ($p < 0.0001$, Fisher’s exact test).

To quantify the degree of object preference in the population of neurons, we calculated the *object bias* of the population response, defined as the mean of the index I_{side} with each neuron weighted by its index I_{depth} . I_{depth} indicates which way, and how strongly, a neuron signals figure and ground when unambiguous depth information is provided. Thus, we take the RDS as the standard test that tells us how to read the neural signals. The *object bias* thus calculated would be zero if there was no association between side-of-figure and depth order preference, positive (between 0 and 1) if there was a bias toward object interpretation, and negative if there was a bias toward window interpretation. All cells tested were included in this analysis. For the V2 data of Figure 6, we obtained an *object bias* of $+0.42$ ($t = 24.2$; $df = 173$; $p < 0.0001$). For V1, it was not significantly different from zero ($t = -0.1$; $n = 77$; $p = 0.93$). Note that the side-of-figure modulation index was calculated from the responses to contrast-defined figures without depth cues, and the *object bias* was obtained from this index by pooling neurons according to their 3D edge selectivity (which is their signature of coding 3D layout). Thus, the fact that the *object bias* for V2 is positive means that contrast-defined figures

without specific depth information are represented in V2 as if they were objects in 3D space.

Besides the neurons that combined selectivity for side-of-figure and depth order (filled symbols), Figure 6 shows that there were also neurons that were selective for side-of-figure, but not for stereoscopic depth order, and others that were selective for depth order, but not side-of-figure. This indicates that two different mechanisms provide inputs to these neurons and sometimes converge onto a single neuron. The predominance of the object interpretation shows that the two mechanisms are not combined at random, but according to the rule that the region of the figure corresponds to an object in 3D space. The convergence seems to occur mainly in V2.

The symbols corresponding to the examples in the previous figures are labeled with numbers in Figure 6; number 1 represents the cell of Figure 4, and numbers 2–5 represent the cells of Figures 5A–5D. It was easy to find examples of cells with strong modulation in both dimensions on the object side, but on the window side only two of the seven cells with both effects had larger modulation indices. Cell number 5 was the best example of this kind. This cell responded vigorously to stereoscopic edges and was completely selective for depth order (Figure 5D), and this was confirmed by recording responses for various edge positions relative to the receptive field (Figure 3A). The contrast edge and bar responses were weak (Figure 5D). Nevertheless, the side-of-figure preference was confirmed by several repetitions, and for different sizes of the square. Cell number 6 of Figure 6 barely responded to RDS, but its depth order preference was confirmed with displays of drifting, dense random-dot patterns. Such displays generate strong depth stratification in perception (cf. Kaplan, 1969; Yonas et al., 1987) and were found to evoke depth order-selective responses in V2 cells similar to those from RDS (von der Heydt et al., 2003). In cell 6, such displays again produced responses according to the window interpretation. Thus, the window combination of side preference and edge selectivity might be more than a variation produced by chance; representing the alternative interpretation might have functional significance. However, the general weakness of response modulation in the few “selective” cells on the window side underscores the predominance of the object-type wiring in neurons of area V2.

The modulation index plotted in Figure 6 indicates the relative change of responses, but not their absolute strength. To show that our analysis is based on robust responses, we have listed in Table 1, for contrast edges and for RDS edges, the means and medians of the response strengths (mean firing rate for the preferred of the four stimulus conditions illustrated in Figure 4). For comparison, the statistics are listed for cells classified as “selective in both tests” (represented by filled dots in Figure 6) and for other cells. The average response strengths were in the range between 30 and 47 spikes/s for contrast edges, and about half of that for RDS. The V2 data show that the responses of the “selective” cells were actually stronger than those of the other cells on average, for contrast edges as well as for RDS.

Experiment 1 consisted of two tests, one with contrast-defined figures, and the other with stereoscopic

Table 1. Comparison of Response Strengths between Cells that Were Selective for Side-of-Figure as Well as Depth Order versus Other Cells

	n	Means		Medians	
		Contrast-defined	RDS	Contrast-defined	RDS
V1					
Selective	2	30.6	11.5	30.6	11.5
Others	75	30.3	17.8	26.3	10.7
V2					
Selective	37	47.3	21.5	44.1	18.3
Others	137	33.2	18.4	23.1	10.2

Mean firing rates for preferred side or depth order (spikes/s).

figures. Each involved two factors, and only the effects of side-of-figure and depth order are represented in Figure 6. In the contrast figure test, the second factor was edge contrast polarity (Figures 4A–4D). The effect of this factor was significant in 42% of the V2 cells. Similar to previous results (Zhou et al., 2000), the effect of contrast polarity was found in about half of the side-of-figure-selective cells, and interaction was found in one-fifth. The most frequent type of interaction was multiplicative behavior, with a strong side-of-figure difference for the preferred contrast polarity, but little difference for the other polarity, because responses were close to zero.

In the stereogram test, the second factor was the location of the disparity-defined figure (Figures 4E–4H). This factor was rarely significant (9%, compared to 35% for contrast-defined figures), and interaction between side of disparity-defined figure and local depth order was also rare. The example shown in Figure 4 is typical. Thus, RDS responses depended on the depth order of the edge in the receptive field, but not on the location of the global shape. We conclude that disparity-defined (“cyclopean”; Julesz, 1971) figures have a weaker Gestalt effect than contrast-defined figures. The selectivity for stereoscopic depth order is produced mainly by local mechanisms.

Contradictory versus Coherent Cues for Objects: Do Gestalt Cues Modulate Stereoscopic Responses?

In the above experiment, side-of-figure preference and stereoscopic selectivity were examined in separate tests. The contrast-defined figures had no stereoscopic cues, while the stereoscopic figures had no contrast borders that would define the shape of the figure. Natural stimuli generally provide global shape information as well as stereoscopic cues. The stereoscopic information tends to “disambiguate” perception. For example, the tilted square in Figure 1B could be perceived as an object or as a window. Although the object interpretation usually dominates, perception may flip back and forth between the two interpretations. However, when texture is added to the display and the square region is given a “near” disparity relative to the dark region, an object is invariably perceived. But when the same region is given a “far” disparity, a window is perceived. In the latter case, disparity overrides the Gestalt influence. This observation suggests that the Gestalt

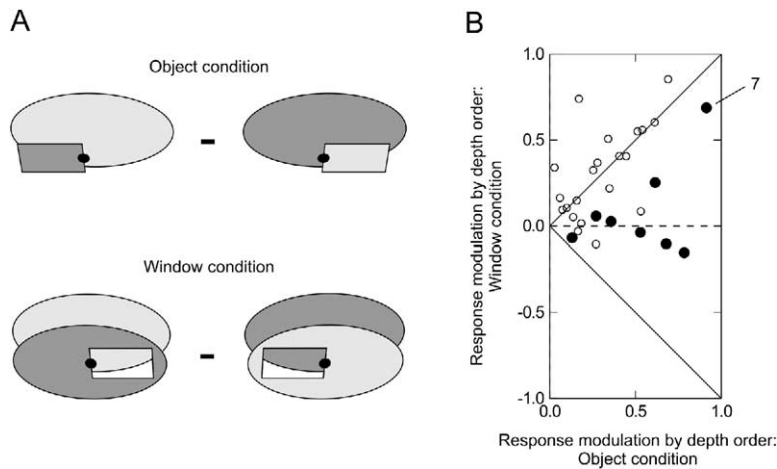


Figure 7. Interaction of Gestalt Factors and Stereoscopic Depth

Figures were defined by luminance contrast *and* disparity. (A) Schematic illustration of 3D stimuli and receptive field position (the random-dot texture is not illustrated; in the case of window stimuli, the border of the background is shown for illustration but was not visible in the experiment). In the absence of depth information, the squares would be perceived as figures and the circular surrounds as ground according to the Gestalt rule that smaller, enclosed regions tend to be interpreted as objects (Rubin, 1921). In the top displays, the stereoscopic information supports this interpretation, because the disparity indicates that the square region is in front of the surrounding region. In the bottom displays, the stereoscopic information contradicts the object interpretation, because the

disparity makes the square region appear farther away than the surrounding region, and the edges therefore cannot be the edges of the square. We compared the neuronal responses to the edges marked by black dots between object and window displays. Note that the corresponding edges are locally identical; only the global context was different. For each condition, a depth order modulation index was calculated by subtracting the response to the two stimuli, as indicated by the minus sign, and dividing the result by the sum of the two. Thus, for both conditions the responses to far-near edges are subtracted from the responses to near-far edges. (B) Scatter diagram of the indices obtained for the two conditions. Each symbol represents the responses of a single cell. Filled dots indicate cells with significant effect of side-of-figure (Gestalt factors). Data points near the 45° diagonal represent cells in which the responses depended on local depth order alone (stereoscopic cue), and data points near the -45° diagonal would indicate that responses were dominated by side-of-figure (Gestalt factors). It can be seen that the influence of the Gestalt factors was to reduce the depth order modulation in the window condition compared to the object condition (data points are below the 45° diagonal). However, the Gestalt effect never dominated (no data points on the -45° diagonal).

influence may be easily obliterated by unambiguous depth cues. How are the different cues combined in single neurons? Are the Gestalt cues weaker than conventional cues such as stereoscopic disparity? Can they influence the responses when pitted against disparity?

In experiment 2, we studied displays in which figures were defined by luminance contrast *and* disparity. As before, a contrast square was presented left or right of the receptive field, but the light and dark regions were also textured with a random-dot pattern (RDS contrast = 0.3). The neural selectivity for depth order was determined with object and window displays, as shown schematically in Figure 7A (which does not show the random-dot texture). The same 3D edge was presented in the receptive field in two conditions: one in which the global shape supports the object interpretation, and the other in which the global shape was located on the “wrong” side, that is, the Gestalt cue contradicts the depth cue. For each condition, the depth order modulation index was calculated. The index for object displays is plotted on the horizontal axis, and the index for window displays is plotted on the vertical axis. The former was taken as the reference; if it was negative, the signs of both indices were reversed. Responses were recorded for the two contrast polarities of the local edge and averaged (only one polarity is illustrated).

Neurons whose responses were determined solely by the local 3D edge would tend to produce the same depth order modulation index for object and window displays, because, in both cases, the index subtracts responses to far-near edges from responses to near-far edges. Such cells would therefore be represented by data points clustering about the 45° line. However, neurons that were dominated by side-of-figure would show

inverted modulation indices, because for the horizontal axis, figure-right was subtracted from figure-left, whereas for the vertical axes, figure-left was subtracted from figure-right. Thus, neurons that are dominated by side-of-figure would be represented near the -45° line.

The cue interaction experiment was performed in 29 stereo edge-selective cells (9 of V1 and 20 of V2), and the results are plotted in Figure 7B. Filled dots indicate neurons in which the main effect of side-of-figure was significant ($p < 0.05$, three-way ANOVA with factors depth order, side-of-figure, and contrast polarity). The plot shows that these cells are represented below the 45° diagonal; they had a lower modulation index in the window condition than in the object condition. Thus, the “wrong” localization of the figure reduced or abolished the depth order signal (the fact that most of these cells cluster about the horizontal axis suggests that the window displays are represented with no clear depth at all in those cells). This shows that Gestalt factors influenced the responses even in the presence of effective stereoscopic cues. However, in none of the cells did the Gestalt cue fully reverse the modulation (no dots on the -45° line).

The interaction of cues is further illustrated by an example in Figure 8 (recordings from the cell labeled “7” in Figure 7). As before, the figures were defined by *luminance contrast and disparity*, but in this case, the contrast of the random-dot texture was varied, thereby varying the strength of the stereoscopic cue. The insets illustrate the four configurations; Figures 8A and 8C represent object conditions, and Figures 8B and 8D represent window conditions; in Figures 8A and 8B, the square shape is located on the left of the receptive field, in Figures 8C and 8D, it is located on the right.

The bar graphs at the bottom of Figure 8 show the

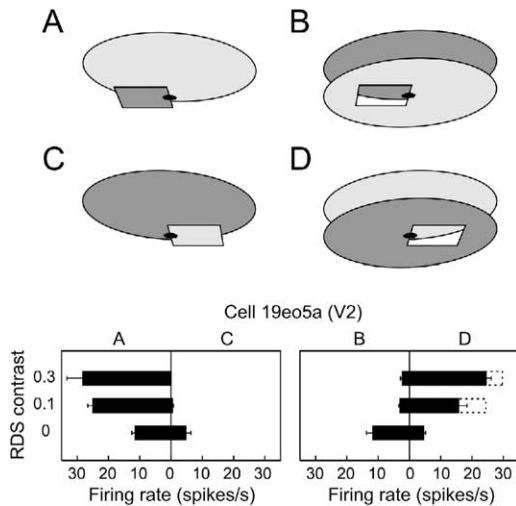


Figure 8. Interaction of Gestalt Factors and Stereoscopic Depth
 Figures were defined by luminance contrast *and* disparity, as in the previous experiment (Figure 7), but the contrast of the random-dot texture was varied to show the transition to the no disparity condition (RDS contrast = 0). The stimulus insets show the figure-left conditions at the top (A and B) and the figure-right conditions below (C and D). (A) and (C) are object conditions, and (B) and (D) are window conditions. The bar graphs show the responses of a V2 neuron with left border ownership preference (mean firing rates and SEM). Bars pointing to left and right of zero represent responses to figure-left and figure-right conditions, respectively (letters above the graph refer to stimulus insets). With the disparity cue (RDS contrast = 0.1 and 0.3) the neuron responds whenever the surface to the left of the receptive field is in front (A and D), but hardly at all when the surface to the right is in front (B and C). But when the disparity cue is removed (RDS contrast = 0) the responses for the window displays reverse; the neuron now responds to (A) and (B) better than to (C) and (D), which means that it signals border ownership according to side-of-figure (the Gestalt factor). Even with disparity present (RDS contrast = 0.1 and 0.3), the “right” responses for window displays are weaker than the “left” responses for object displays (dashed lines in [D] show the size of the corresponding responses [A] for comparison), indicating that a Gestalt effect is present even when unambiguous depth information is available.

responses of the neuron for these four conditions at three different contrast levels of the random-dot texture (RDS contrast). Bars extending left and right of the zero line correspond to left and right location of the square. It can be seen that, with stereoscopic cues (RDS contrast = 0.1 and 0.3), responses to Figure 8A are stronger than responses to Figure 8C, and responses to Figure 8D are stronger than responses to Figure 8B. Thus, the neuron responds according to stereoscopic depth order. However, in the no texture condition (RDS contrast = 0), the responses to the window displays flip to the left; Figure 8B now produces stronger responses than Figure 8D. This corresponds to a change in perception of border ownership—without the stereoscopic cues, the squares in Figures 8B and 8D are no longer perceived as windows, but as objects, according to Gestalt cues. Border ownership flips from right to left in Figure 8B, and from left to right in Figure 8D. Note that, even with the disparity cue, the responses for Figure 8D were slightly weaker than the responses for Figure

8A (dashed lines in plot in Figure 8D are copies of the bars from Figure 8A). This shows the attenuation of stereoscopic signals by the Gestalt factor that was demonstrated in Figure 7.

Controls

We considered errors in centering the edge of the test figure in the receptive field and deviations of direction of gaze as possible confounds. For the *side-of-figure* test, position errors can probably be neglected, because we compare responses between two conditions in which the displays are identical over a region that is larger than the “minimum response field” of the cells. Thus, random position errors would produce similar variations of response in both cases and thus cancel. Systematic deviations of fixation according to figure location were ruled out by eye movement recordings. For the stereoscopic test, *depth order* selectivity was verified by recording position-response curves (Figure 3) for part of the cells of our sample, specifically for 18 of the 37 V2 neurons classified as selective for side-of-figure and depth order (filled symbols in Figure 6). Changes in convergence of the eyes would not be detected by our eye movement recordings, which were only for one eye. To see if the stereograms caused changes of convergence, we analyzed the responses of disparity-selective cells in the presence of background disparities (see Experimental Procedures). This analysis indicated that convergence was maintained accurately.

Discussion

The phenomenon of figure-ground organization played a key role in the formulation of the Gestalt theory, which conjectured that central processes such as attention and recognition access visual image information not directly, but through an intermediate, structured representation (Rubin, 1921; Wertheimer, 1923). Later studies have demonstrated that changes in perceived depth stratification dramatically affect perception of form, recognition of objects, and selective visual attention (Driver and Baylis, 1996; He and Nakayama, 1992; Nakayama et al., 1989; Rensink and Enns, 1998). Both older and recent studies pointed out that the internal assignment of border ownership seems to be the key to understanding these results. Based on single-cell recordings in macaques, Zhou et al. (2000) suggested that border ownership is encoded in the contrast edge responses of neurons in the visual cortex.

The present results show that the visual cortex processes global configuration together with binocular information to relate contrast borders to object contours and assign border ownership. There are two key observations. First, neurons that are side-of-figure selective for edges of 2D figures are often (61%) selective for depth order of 3D edges. Second, the side of the figure that produces the stronger response is also usually the “near” side of the 3D step for which the neuron is selective (Figure 6). Thus, the system assigns the contrast borders of 2D figures as if they were objects in 3D space. For contrast-defined figures that provide no stereo cues, the configuration of contours determines the

border ownership signal according to Gestalt rules. When contrast borders are missing, as in RDS, the depth order determines the signal. In general, both kinds of information contribute to the border ownership signal, but if stereo depth is in conflict with Gestalt rules (according to which enclosed, compact image regions should be interpreted as objects), the influence of the stereoscopic input is reduced or abolished (Figure 7). These results support the hypothesis of border ownership coding (Zhou et al., 2000). Side-of-figure selectivity by itself might be dismissed as a random asymmetry of receptive fields (spatial heterogeneity of nonclassical surround has been observed in V1; Freeman et al., 2001; Jones et al., 2001; Levitt and Lund, 2002), but the linkage between stereoscopic selectivity and 2D contextual influence is unequivocal evidence for border ownership coding.

The possibility that the side-of-figure effect is an artifact of displacements of the receptive field due to residual eye movements can be ruled out because responses are compared between stimulus conditions that are identical in and around the minimum response field. That selectivity for depth order was genuine, and not due to eccentric positioning, was demonstrated by recording position-response profiles for figures in RDS in about half of the neurons of the main sample. If anything, positioning errors would have produced depth order preferences at random in different neurons, but Figure 6 shows that depth order preference was correlated with side-of-figure preference. Stimulus-induced changes in fixation were ruled out by eye movement recordings and by analysis of the disparity tuning of neurons, which indicated that convergence of the eyes was unaffected by the stimulus. Also, the effects of positioning errors and eye movements would be more noticeable in V1 than in V2 because of the smaller size of receptive fields in V1, but the observed depth order selectivity was more pronounced in V2.

Cells that were selective for side-of-figure and depth order (filled symbols in Figure 6) responded with higher mean firing rates than other cells (Table 1). One possible explanation for this is that border ownership modulation produces enhancement of responses for the preferred condition. However, there might be other reasons. The most effective spatial pattern generally varies from cell to cell, some responding best to edges, others to gratings, bars, or other patterns. These variations are probably related to the different functions of cortical cells in the visual process, for example, contour versus surface representation. Thus, border ownership-selective cells might be more responsive to edges than other cells because they are involved in contour representation.

The fact that only a fraction of cells was found to be selective for side-of-figure or depth order (combined, these were 54% of the cells tested) is not surprising considering that only a fraction of the contrast borders in natural images are occluding contours (contrast borders are also produced by surface pigmentation, bending of a surface, shadows, etc.). Accordingly, border ownership assignment is only one of several tasks performed in the visual cortex. Also, in microelectrode recording experiments, as described here, signals are selected randomly from the neural network and, therefore,

presumably, reflect various stages of processing and thus various levels of neural selectivity.

The Origin of the Gestalt Influence

The influence of global configuration is still mysterious. Our results show that the range of this influence extends far beyond the limits of the classical receptive fields, which might be taken as indicating a process of central origin. However, several observations argue against this possibility.

One is the early differentiation of the responses for the two sides of figure (Figures 4 and 5), which seems to exclude central loops such as IT cortex as the mechanism of figure-ground differentiation, as we have discussed earlier (Zhou et al., 2000).

Another observation is that the side-of-figure preference of each single neuron is fixed in relation to its receptive field. Another neuron with the same location and orientation of receptive field may have the opposite preference. This means that the identification of the figure area is probably not due to an influence of top-down attention. How can attention signals, which should be able to gate the activity for a figure in either location, produce different effects for the two locations? And, if attention is directed to the figure in one location, how can it simultaneously enhance activity in one cell but suppress it in the other? It seems that, for the top-down signal to produce opposite effects in different neurons, there must be lower-level mechanisms that differentiate the cells. A similar argument can be made regarding back-propagation of signals from a shape recognition stage such as the inferior temporal cortex. It is unlikely that such influences would be side specific to the individual receptive fields.

It is important also to remember that our findings reflect the activity in the visual cortex when the animal was engaged in a demanding fixation task (depth matching at stereoscopic threshold). This probably means that the animal tried, as much as possible, to ignore the stimuli to which the neurons responded. Recent experiments with multiple figures and operational control of attention confirmed that border ownership in V2 is generated independently of attention (although many cells also show an attention effect) (von der Heydt et al., 2004).

The present results, showing that side-of-figure selectivity is “wired up” with stereoscopic selectivity in a specific way in single neurons, support the conclusion that the preference of neurons for one or the other side is hard-wired and not under central control. Stereoscopic selectivity originates early in the visual cortex and, therefore, probably is hard-wired. Because the object bias illustrated in Figure 1 is an invariable property of images of a 3D world, the side-of-figure preference of neurons and its link to depth order preference should also be invariant.

Exactly how the lower cortical areas would integrate information from distant parts of the visual field remains to be determined. Because image information is laid out retinotopically in area V2 (Gattass et al., 1981; Van Essen and Zeki, 1978), the representations of the figure boundaries are widely distributed in the cortex. Thus, for the processing to occur within V2, one would have

to assume fast horizontal propagation of signals to explain the rapid emergence of border ownership signals. Given the large size of V2, the conduction velocity of intracortical fibers might be too slow. Another possibility is that the integration occurs via recurrent signals from nearby areas, such as V3 or V4, which would travel through the much faster fibers of the white matter (Bullier, 2001; Hupe et al., 2001).

Neural Coding of Figure-Ground Organization

Figure-ground organization is a complex phenomenon that involves depth stratification as well as grouping of elementary features into larger units (“figures”). Border ownership coding provides a key to understanding a broad range of observations. Our results suggest that the coding of border ownership is surprisingly simple: each segment of contrast border is represented by two groups of orientation-selective neurons, one for each side of ownership, whose differential activity encodes the border assignment, similarly as motion is encoded by cells with opposite direction preference, or light and dark by on- and off-center ganglion cells. We assume that the strength of the neural border ownership signal is related to the probability of perceiving one of two adjacent regions as occluding the other. Thus, neural border ownership assignment is not an all-or-none process. For example, side-of-figure signals in V2 decrease with increasing figure size (Figure 2C). Correspondingly, smaller regions have a higher probability to be perceived as foreground than larger regions (Rubin, 1921). We do not imply that V2 is “the site of perception” of figure and ground. Such an interpretation would be incompatible with the graded nature of border ownership signals and the observation of neurons representing alternative interpretations in parallel (Figure 6).

Coding border ownership in orientation-selective cells is an effective way of representing the overlay structure of scenes, because the signals of these cells form the basis of shape representation for subsequent stages of processing. The assignment of border ownership directly specifies which contour elements are to be processed for each shape and which are not. Indeed, border ownership assignment affects shape recognition (Driver and Baylis, 1996; Nakayama et al., 1989) and shape-specific visual search (He and Nakayama, 1992; Rensink and Enns, 1998). The figure-ground dependence of motion signals in MT (and of motion perception) indicates that MT mechanisms compute the direction of motion of a surface from features at the borders of the surface, selecting the features according to border ownership (Duncan et al., 2000; Shimojo et al., 1989).

The finding of a convergence of stereoscopic and Gestalt-based mechanisms provides interesting clues about how the visual cortex might represent surfaces. Our results show that 3D edge-selective neurons not only detect disparity edges (von der Heydt et al., 2000), but in many cases also assign border ownership. Thus, these neurons do not represent isolated 3D features, but edges with reference to an adjacent region. Other neurons represent brightness and color borders, again with a pointer to an adjacent region (Zhou et al., 2000). Taken together, these instances of “Gestalt” influence

are evidence for mechanisms that link diverse feature signals to larger entities. We argue that linking contour features to regions is a fundamental operation in coding 3D surfaces.

The existence of 3D surface representations has been suggested by many studies (for a review, see Nakayama et al., 1995). For example, stereograms can produce illusory surfaces (Gregory and Harris, 1974; Idesawa, 1991), depth order influences the perceived color of surfaces (Nakayama et al., 1989; Nakayama et al., 1990), and visual attention is deployed according to perceptual surfaces (He and Nakayama, 1995). Correlates of depth stratification and illusory surface formation have been demonstrated in neuronal responses (Bakin et al., 2000). The convergence of stereoscopic and contrast information in border ownership-selective neurons might provide a basis for a general explanation of these phenomena. Theoretical studies (E. Craft, H. Schuetze, E. Niebur, and R.v.d.H., submitted; Schuetze et al., 2003) show that the neural mechanisms of Gestalt-based border ownership assignment can be modeled by relatively simple “grouping” circuits. They suggest that these circuits might serve top-down attention mechanisms to access the “grouped” information by polling the various neurons representing the borders of a figure. This way the features that define a surface, such as 3D shape and color, can be selected as a whole for further processing. Thus, the demonstration of a link between stereoscopic and Gestalt-based mechanisms for assignment of contrast borders is a step toward understanding the coding of visual information at this intermediate stage and its role in the vision process.

Experimental Procedures

Single neurons were recorded from areas V1 and V2 of the visual cortex in alert, behaving macaques (*Macaca mulatta*). Three small posts for head fixation and two recording chambers over the left and right visual cortex were attached to the skull with bone cement and surgical screws. The surgery was done under aseptic conditions under pentobarbital anesthesia induced with ketamine, and buprenorphine was used for postoperative analgesia. All animal procedures conformed to National Institutes of Health and USDA guidelines as verified by the Animal Care and Use Committee of the Johns Hopkins University.

Recording

Single-neuron activity was recorded extracellularly with glass-insulated Pt-Ir or Quartz-insulated Pt-W microelectrodes inserted through small (3–5 mm) trephinations. Area V1 was recorded right under the dura, and V2 was recorded either in the posterior bank of the lunata sulcus, after passing through V1 and the white matter, or in the lip of the postlunate gyrus. The two areas were distinguished by their retinotopic organization and by histological reconstruction of the recording sites as described previously (Zhou et al., 2000).

Control of Fixation

Eye movements were recorded for one eye using a video-based infrared pupil tracking system with a camera mounted on the axis of fixation via a 45° beam splitter. A fixation task was used that required the subjects to align a dot to a short line stereoscopically to within a disparity near the stereoscopic threshold. To facilitate fixation in the presence of random-dot texture, the fixation target was presented on a black circular background of 20 arc min diameter. The criterion disparity was set so low (generally 0.5–0.67 arc min) that the adjustment took 1–2 s, during which fixation was steady. Lateral movements during fixation were generally small (SD

0.15°–0.2°), and data from trials during which the fixation deviated from the target by more than 1° were discarded. Performance in the depth matching task reached the limits of stereoscopic acuity. This indicates that the eyes converged accurately on the target, because stereoscopic acuity is highest only for targets on the horopter and falls off steeply with distance from the horopter (Blake-more, 1970).

To see if the depth of fixation was altered by the disparity of the random-dot patterns, we estimated the stimulus-induced vergence movements from the responses of V2 neurons with sharp disparity tuning. Using two-surface RDS stimuli, we recorded the disparity tuning functions for various disparities of the texture surrounding the fixation target. We then modeled the effect of vergence movements on the neuronal responses under the assumption that the disparity around the fixation target would induce a proportional deviation of vergence and determined the gain factor of vergence induction that maximized the cross-correlation between the different tuning functions recorded from each neuron (R.v.d.H. and F.T.Q., unpublished data). For inducing disparities varying between –10 and +10 arc min, the mean estimated gain of vergence induction was 0.03 (SD 0.04, range –0.01 to 0.12; $n = 7$). Thus, the estimated gain was very low. There are two possible explanations for this. Either stimulus-induced vergence eye movements are virtually absent, or they occur but are compensated by neural mechanisms. It has been shown that the responses of V2 cells to a stimulus in the receptive field can be influenced by the disparity of the surrounding region, causing, in some cases, the disparity tuning to shift in the direction of the surround disparity (Thomas et al., 2002). In the extreme, cells may signal the “relative disparity” (the difference between center and surround disparities) rather than the “absolute disparity” of the stimulus in the receptive field. However, to produce gain factors as low as a few percent, as in our estimates, neuronal mechanisms would have to compensate for 97% of the vergence. Cells with nearly complete disparity differencing are rare, though, and the majority of V2 cells show no effect of the disparate surround (Thomas et al., 2002). Thus, the possibility that the cells we tested were all of the differencing kind is extremely unlikely. We conclude that the more likely explanation for the small estimates of the gain of vergence induction is that the animal was able to maintain stable vergence even under conditions in which the disparity of the surrounding texture varied. Based on our estimate, an induced vergence change of $0.03 \times 24 \text{ arc min} = 0.72 \text{ arc min}$ would be expected for the largest surround disparity used. Changes this small can probably be neglected.

Visual Stimuli and Test Procedures

Stimuli were generated on a Silicon Graphics O2 workstation using the antialiasing feature of the Open Inventor software and presented on a Barco CCID 121 FS color monitor with a 72 Hz refresh rate. Stereoscopic pairs were presented side by side and superimposed optically at 40 cm viewing distance. The optical system could be switched between magnifications of 0.74 and 1.56 arc min/pixel, providing fields of $8^\circ \times 12^\circ$ and $17^\circ \times 26^\circ$ visual angle, respectively. Stationary bars were used to determine the color preference, and bars and drifting gratings to map the “minimum response field” (the minimum region outside which the stimulus does not evoke a response; Barlow et al., 1967) of each cell. Orientation and disparity tuning curves were recorded using moving bars. The bars were presented on a neutral background of 16 cd/m^2 luminance. Subsequently, an edge of a square figure (3° – 8°) was centered on the minimum response field at the preferred orientation. For contrast figures, the preferred color and gray (16 cd/m^2) were used for figure and surround (Zhou et al., 2000). The preferred color could be chromatic or achromatic, and white was used in the absence of color selectivity. In general there was a luminance contrast between figure and surround. Stereoscopic figures were generated by means of dynamic RDS (Julesz, 1960) using randomly positioned white dots (53 cd/m^2) on gray (16 cd/m^2) with 2 or 6 arc min dot size, 14% coverage, and a pattern renewal frequency of 8 Hz. Stereoscopic (cyclopean) squares and square windows with edges corresponding exactly to the edges of the contrast figures were generated. The preferred disparity (or zero, if the disparity tuning was flat) was used for the “near” plane, while the “far” plane was

placed at a distance of 10 or 24 arc min disparity behind the fixation target. In the experiment of Figure 7, the luminance modulation of the random-dot pattern (Michelson contrast = 0.3) was applied to the colors of figure and surround.

Experimental Design and Data Analysis

For the experiment of Figure 2, four displays as shown in Figure 2A plus four displays with reversed contrast were tested. For the experiment of Figure 3, cyclopean squares, 3° on a side, were generated using dynamic RDS. The depth of the square was set to the optimum disparity for the cell under study, or zero, if there was no clear disparity tuning. Each of the four sides of the square was placed in the receptive field at seven positions spaced 0.167° , in random order. For the experiment of Figures 4–6, four displays of contrast-defined figures, as shown in Figures 4A–4D, and four RDS portraying cyclopean squares and windows at the same positions as the contrast squares were tested. In the experiment of Figures 7 and 8, the figures were defined by both contrast and disparity, and stimuli consisted of four displays as illustrated in Figure 7, plus four displays with reversed contrast. In each experiment, all stimuli were presented four times in random order.

Analysis was based on the spike counts during 800 ms after stimulus onset. Cells with contrast edge responses <4 spikes/s were excluded, because we felt that our stimuli were not adequate to drive these cells (23%). Selectivity was assessed by ANOVA performed on the square-root transformed spike counts, using a significance criterion of $p < 0.05$. For the experiment of Figure 2, a three-factor ANOVA was performed (factors: side-of-figure, edge contrast polarity, and size). For the experiment of Figures 4–6, two separate two-factor ANOVAs were performed, one for the contrast figure data (factors: side-of-figure, edge contrast polarity), and one for the RDS data (factors: depth order, side-of-figure). To calculate the modulation index for side-of-figure (Figure 6, vertical axis), the responses to the two contrast polarities were averaged; to calculate the modulation index for depth order (Figure 6, horizontal axis) the responses to squares and windows were averaged. No subtraction was made for spontaneous activity (which would have exaggerated the modulation indices). The data from the experiment of Figure 7 were analyzed by three-way ANOVA (factors: depth order, side-of-figure, and edge contrast polarity).

Acknowledgments

We thank Ofelia Garalde for technical assistance and Todd J. Macuda for his participation in some of the experiments. This work was supported by National Institutes of Health grants EY-02966 and NS-38034.

Received: November 11, 2004

Revised: April 12, 2005

Accepted: May 24, 2005

Published: July 6, 2005

References

- Albright, T.D., and Stoner, G.R. (2002). Contextual influences on visual processing. *Annu. Rev. Neurosci.* 25, 339–379.
- Bakin, J.S., Nakayama, K., and Gilbert, C.D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *J. Neurosci.* 20, 8188–8198.
- Barlow, H.B., Blakemore, C., and Pettigrew, J.D. (1967). The neural mechanism of binocular depth discrimination. *J. Physiol.* 193, 327–342.
- Baumann, R., van der Zwan, R., and Peterhans, E. (1997). Figure-ground segregation at contours: a neural mechanism in the visual cortex of the alert monkey. *Eur. J. Neurosci.* 9, 1290–1303.
- Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *J. Physiol.* 211, 599–622.
- Bullier, J. (2001). Integrated model of visual processing. *Brain Res. Brain Res. Rev.* 36, 96–107.
- Driver, J., and Baylis, G.C. (1996). Edge-assignment and figure-

- ground segmentation in short-term visual matching. *Cognit. Psychol.* 31, 248–306.
- Duncan, R.O., Albright, T.D., and Stoner, G.R. (2000). Occlusion and the interpretation of visual motion: perceptual and neuronal effects of context. *J. Neurosci.* 20, 5885–5897.
- Freeman, R.D., Ohzawa, I., and Walker, G. (2001). Beyond the classical receptive field in the visual cortex. *Prog. Brain Res.* 134, 157–170.
- Gattass, R., Gross, C.G., and Sandell, J.H. (1981). Visual topography of V2 in the macaque. *J. Comp. Neurol.* 201, 519–539.
- Gregory, R.L., and Harris, J.P. (1974). Illusory contours and stereo depth. *Percept. Psychophys.* 15, 411–416.
- He, Z.J., and Nakayama, K. (1992). Surfaces versus features in visual search. *Nature* 359, 231–233.
- He, Z.J., and Nakayama, K. (1995). Visual attention to surfaces in three-dimensional space. *Proc. Natl. Acad. Sci. USA* 9, 11155–11159.
- Heider, B., Spillmann, L., and Peterhans, E. (2002). Stereoscopic illusory contours—cortical neuron responses and human perception. *J. Cogn. Neurosci.* 14, 1018–1029.
- Hupe, J.M., James, A.C., Girard, P., Lomber, S.G., Payne, B.R., and Bullier, J. (2001). Feedback connections act on the early part of the responses in monkey visual cortex. *J. Neurophysiol.* 85, 134–145.
- Idehara, M. (1991). Perception of 3-D illusory surface with binocular viewing. *Jpn. J. Appl. Phys.* 30, 751–754.
- Jones, H.E., Grieve, K.L., Wang, W., and Sillito, A.M. (2001). Surround suppression in primate V1. *J. Neurophysiol.* 86, 2011–2028.
- Julesz, B. (1960). Binocular depth perception of computer-generated patterns. *Bell System Technical Journal* 39, 1125–1161.
- Julesz, B. (1971). *Foundations of Cyclopean Perception* (Chicago: University of Chicago Press).
- Kaplan, G.A. (1969). Kinetic disruption of optical texture: the perception of depth at an edge. *Percept. Psychophys.* 4, 193–198.
- Koffka, K. (1935). *Principles of Gestalt Psychology* (New York: Harcourt, Brace and World).
- Lamme, V.A.F. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *J. Neurosci.* 15, 1605–1615.
- Lee, T.S., Mumford, D., Romero, R., and Lamme, V.A.F. (1998). The role of the primary visual cortex in higher level vision. *Vision Res.* 38, 2429–2454.
- Levitt, J.B., and Lund, J.S. (2002). The spatial extent over which neurons in macaque striate cortex pool visual signals. *Vis. Neurosci.* 19, 439–452.
- Nakayama, K., Shimojo, S., and Silverman, G.H. (1989). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception* 18, 55–68.
- Nakayama, K., Shimojo, S., and Ramachandran, V.S. (1990). Transparency: relation to depth, subjective contours, luminance and neon color spreading. *Perception* 19, 497–513.
- Nakayama, K., He, Z.J., and Shimojo, S. (1995). Visual surface representation: a critical link between lower-level and higher-level vision. In *Invitation to Cognitive Science*, S.M. Kosslyn and D.N. Osherson, eds. (Cambridge, MA: MIT Press), pp. 1–70.
- Poggio, G.F., Motter, B.C., Squatrito, S., and Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res.* 25, 397–406.
- Rensink, R.A., and Enns, J.T. (1998). Early completion of occluded objects. *Vision Res.* 38, 2489–2505.
- Rossi, A.F., Desimone, R., and Ungerleider, L.G. (2001). Contextual modulation in primary visual cortex of macaques. *J. Neurosci.* 21, 1698–1709.
- Rubin, E. (1921). *Visuell wahrgenommene Figuren* (Copenhagen: Gyldendal).
- Rubin, E. (2001). Figure and ground. In *Visual Perception: Essential Readings*, S. Yantis, ed. (Philadelphia: Psychology Press), pp. 225–229.
- Schuetze, H., Niebur, E., and von der Heydt, R. (2003). Modeling cortical mechanisms of border ownership coding. *J. Vis.* 3, 114.
- Shimojo, S., Silverman, G.H., and Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Res.* 29, 619–626.
- Spillmann, L., and Ehrenstein, W.H. (2003). Gestalt factors in the visual neurosciences. In *The Visual Neurosciences*, L.M. Chalupa and J.S. Werner, eds. (Cambridge, MA: MIT Press), pp.1573–1589.
- Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature* 401, 269–272.
- Thomas, O.M., Cumming, B.G., and Parker, A.J. (2002). A specialization for relative disparity in V2. *Nat. Neurosci.* 5, 472–478.
- Van Essen, D.C., and Zeki, S.M. (1978). The topographic organization of rhesus monkey prestriate cortex. *J. Physiol.* 277, 193–226.
- von der Heydt, R., Heitger, F., and Peterhans, E. (1993). Perception of occluding contours: Neural mechanisms and a computational model. *Biomed. Res. Suppl.* 14, 1–6.
- von der Heydt, R., Zhou, H., and Friedman, H.S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Res.* 40, 1955–1967.
- von der Heydt, R., Qiu, F.T., and He, Z.J. (2003). Neural mechanisms in border ownership assignment: motion parallax and gestalt cues. *J. Vis.* 3, 666.
- von der Heydt, R., Sugihara, T., and Qiu, F.T. (2004). Border ownership and attentional modulation in neurons of the visual cortex. *Perception* 33, 46.
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. *Psychol. Forsch.* 4, 301–350.
- Wertheimer, M. (2001). Laws of organization in perceptual forms. In *Visual Perception: Essential Readings*, S. Yantis, ed. (Philadelphia: Psychology Press), pp. 216–224.
- Yonas, A., Craton, L.G., and Thompson, W.B. (1987). Relative motion: Kinetic information for the order of depth at an edge. *Percept. Psychophys.* 41, 53–59.
- Zhou, H., Friedman, H.S., and von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *J. Neurosci.* 20, 6594–6611.
- Zipser, K., Lamme, V.A.F., and Schiller, P.H. (1996). Contextual modulation in primary visual cortex. *J. Neurosci.* 16, 7376–7389.