RESEARCH ARTICLE

Occipital network for figure/ground organization

Lora T. Likova · Christopher W. Tyler

Abstract To study the cortical mechanism of figure/ground categorization in the human brain, we employed fMRI and the temporal-asynchrony paradigm. This paradigm is able to eliminate any differential activation for local stimulus features, and thus to identify only global perceptual interactions. Strong segmentation of the image into different spatial configurations was generated solely from temporal asynchronies between zones of homogeneous dynamic noise. The figure/ground configuration was a single geometric figure enclosed in a larger surround region. In a control condition, the figure/ground organization was eliminated by segmenting the noise field into many identical temporal-asynchrony stripes. The manipulation of the type of perceptual organization triggered dramatic reorganization in the cortical activation pattern. The figure/ground configuration generated suppression of the ground representation (limited to early retinotopic visual cortex, V1 and V2) and strong activation in the motion complex hMT+/V5+; conversely, both responses were abolished when the figure/ground organization was eliminated. These results suggest that figure/ground processing is mediated by top-down suppression of the ground representation in the earliest visual areas V1/V2 through a signal arising in the motion complex. We propose a model of a recurrent cortical architecture incorporating suppressive feedback that operates in a topographic manner, forming a figure/ground categorization network distinct from that for "pure" scene segmentation and thus underlying the perceptual organization of dynamic scenes into cognitively relevant components.

Keywords Perceptual organization · Figure/ground · Contextual interactions · Temporal asynchrony · Visual cortex · Suppression · Salience · Top-down feedback · V1 · V2 · hMT+

L. T. Likova (✉) · C. W. Tyler
The Smith-Kettlewell Eye Research Institute,
San Francisco, CA, USA
e-mail: lora@ski.org

Introduction

Global contextual processes such as figure/ground categorization have a fundamental importance in theories of shape processing, object recognition, and attention. The region of the visual scene categorized by the brain as a 'figure' tends to appear closer to the observer and usually represents an object. The contour 'belongs' to the figure and thus a shape is encoded for the figural region only, while the ground is 'shapeless' and perceived as extended behind the figure and its contour (Rubin 1921/2001; Wertheimer 1923/1938/1997, 1925). Thus, delineation of the figure region is a primary requirement for shape recognition. What neural architecture allows local and global information to be combined rapidly, to provide for figure/ground categorization? In terms of neural architecture, does the mechanism employ lateral interactions (e.g., Gerrits and Vendrik 1970; Grossberg and Mingolla 1985; Grossberg 1994; Baek and Sajda 2005; Pao et al. 1999; Zhaoping 2005), top-down feedback (e.g., Lamme and Roelfsema 2000; Lamme et al. 2002; Roelfsema et al. 2002; Craft et al. 2007) or more sophisticated combinations? In terms of its coding mechanism, do the figure/ground configurations use border-based coding (e.g., von der Heydt et al. 2003, 2005; Craft et al. 2007) or region-based coding (e.g., Lamme 1995; Zipser et al. 1996;

Lamme et al. 1998; Super et al. 2001; Roelfsema et al. 2002)?

There is much controversy even about the basic principles of the figure/ground mechanism, as may be seen from the above references. Addressing the discrepancies among the neurophysiological predictions, as well as among computational models for such global contextual processing, requires a large-scale method such as fMRI that has the ability to analyze the global pattern of activation and suppression across the cortex, and thus would be able to reveal the underlying cortical network. It also requires a paradigm that produces *no* differential activation for *local* stimulus features, that is, local bottom-up mechanisms, but is designed to identify only *global* interactions among stimulation regions.

A paradigm that satisfies these requirements employs image segmentation defined solely by *asynchrony* between transient refreshes of regions in a random-dot field. Because the noise is random, every dot in each display has a 50% chance of changing polarity with a noise refresh and all the dots are refreshed an equal number of times during the fMRI run (the same fixed dot update rate is applied in all stimuli), hence there is no net difference between the test- and the null-epochs of stimulation at any point in the visual field. Regions of the field that undergo synchronous noise replacement are perceptually integrated into unified zones (despite the fact that the noise is spatially uncorrelated from frame to frame), while the boundaries defined by asynchrony between the zones are seen as sharp segmentations (Fahle 1993). We term this phenomenon 'Structure-from-Asynchrony' (SfA). Kandil and Fahle (2001, 2003; see also Lee and Blake 1999; Blake and Lee 2005, for a review) have shown that accurate shape perception is possible with borders formed by an orientation version of such temporal asynchrony. When a region meets the Gestalt criteria for figure (e.g., smaller, surrounded) versus ground (larger, surrounding), the *SfA* is capable of supporting strong figure/ground percepts. Importantly, however, when these criteria are not met, the regional segregation mechanisms still operate but there is no perceptual categorization of the regions as figure versus ground. Thus, in a block design, with *SfA* stimuli in the *test epoch* contrasted against *synchronous* refreshes of the whole random dot field where no *SfA* is perceived (*null epoch*), the only stimulus difference lies in the synchrony or asynchrony between large areas of dots, making this paradigm of interest for studying *global* spatial interactions.

Comparison of the brain responses for two basic categories of spatiotemporal organization in *SfA* (either a figure/ground organization or a segmented structure of multiple equivalent stripes) can determine the cortical networks activated by the two levels of perceptual processing. The pattern of cortical responses for the multi-stripe configuration

can reveal the network processing for the temporal-asynchrony segmentation. This activation, in turn, forms the basis for comparison with the figure/ground configuration; to the extent that the cortical activation pattern is similar to that for the multi-stripe *SfA*, it can be attributed to segmentation processes per se, while a novel pattern of cortical activation will signify the operation of processes specific to figure/ground categorization as such, either facilitative or suppressive. The retinotopic location of the signal (along the borders, within the figure or within the ground region) will differentiate between the border-based and region-based coding schemes for the segmentation processing. Moreover, a recursive architecture of top-down feedback predicted from the neurophysiological experiments can be inferred on the basis of BOLD suppression in early areas together with activation solely in higher areas. This particular differential response pattern is therefore the signature of recursive feedback to be sought in comparison between the two experiments.

General methods

Stimuli

Classically, dynamic visual noise (DVN) stimuli are obtained by updating the noise synchronously at every point in the stimulus field with uncorrelated noise at a constant rate. In the fMRI paradigm, the main occipital responses to DVN are found in V1, while V3A and especially hMT/V5+ respond less to DVN than to coherent motion (Tyler et al. 2006).

For generating *SfA*, all regions throughout a field of random dots were replaced by uncorrelated random dots at a constant rate of 500 ms/frame, in both the test and the null epochs of the fMRI paradigm. However, in the test epochs (Figs. 1a, 2a), the 500 ms/frame replacements in specified zones of the DVN field were offset by a fixed delay (250 ms) relative to the rest of the field. This temporal asynchrony between regions resulted in perceived spatial segmentation during the test epoch despite the identical replacement rate throughout the field. For example, if the 500 ms/frame noise replacements in the central horizontal zone (see Fig. 1a) had an asynchrony relative to the noise replacements in the upper and lower surrounding regions then virtual borders (dashed lines in Fig. 1a) were perceived segmenting the field into a central figural zone and its surroundings. The virtual borders between the zones appeared as sharp and vivid as motion or luminance borders. Because the noise updates were uncorrelated, however, no coherent apparent motion was perceived in these transient fields. The block fMRI design consisted of the SfA noise field during the test epoch and a uniform,

synchronously updated noise field during the null epoch. The full field size was 30×40 deg and the visual noise elements were 'featureless' pixel-size dots.

Scanning procedure

The fMRI responses were collected with 2D spiral acquisition from a custom occipital surface coil in a GE Signa 3T scanner at the Lucas Center at Stanford University. There were 23 coronal slices at 3 s TR, with TE of 30 ms and flip angle 70°, providing $2.5 \times 2.5 \times 3$ mm voxels throughout the occipital cortex. A high-resolution anatomical (T1-weighted) volume scan of the entire brain was also obtained for each observer (voxel size $= 0.94 \times 0.94 \times 1.2$ mm). The stimuli were rear-projected onto a translucent screen inside the bore of the scanner by means of an LCD projector controlled by a Macintosh computer. The observer's head was stabilized by chin and forehead tape, with the eyes viewing the front of the projection screen via a 45° mirror. Observers were required to maintain fixation on a red 4×4 pixel fixation point at the center of the stimulus. The test and the null stimuli were alternated in 18 s cycles (9 s test-epochs alternating with 9 s null-epochs), for a total of 12 cycles per experimental condition, alternating between conditions during the scan. Each scanning session included four repeats for each of the two conditions.

The experiments incorporated an *attention control* (a psychophysical task designed to equate the attention between the test and null epochs without suppressing figure/ground perception), allowing us to map the figure/ground response per se. As figure/ground modulation is a global process that depends on the level of awareness in the perception of the stimuli (Lamme et al. 1998; Super et al. 2001), an attention task was chosen that does not eliminate the global interactions and does not suppress perceptual awareness. Thus, we used a global dimming task, where the observers were asked to detect instantaneous changes in the luminance of the stimulus at an average rate of 1 per 20 frames. The dimming was adjusted during the scans in a 3-up/1-down staircase with steps of 0.1 log units based on whether the observer responded to each presentation with a button press within 2 s. The staircase asymptoted at the level of 79% correct detection of the dimming events.

Data analysis and visualization

Gray (cortex) and white (nerve fiber) matter were segmented from the T1 scan using the Stanford Vistasoft software. The cortex was specified as a manifold extending 3 mm above the segmented gray/white matter boundary, which was flattened from a locus centered on the occipital pole to a radius of 70–80 mm (depending on the size of the occipital lobe) to form a flatmap for full visualization of the relevant fMRI activity throughout the occipital lobe.

The horizontal axis in the flatmaps (Figs. 1e, f, 2c) is oriented from medial (V1 periphery) toward left and right lateral (hMT+/V5+) directions; the top of the flatmaps is dorsal, the bottom–ventral. (Gyri are coded in lighter gray than the sulci). The boundaries of the retinotopic areas V1, V2d, V2v, V3d, V3v, and hV4 were established as described in Sereno et al. (1995) and Engel et al. (1997). Retinotopic areas V3A, V3B and V7 were specified in accordance with Press et al. (2001) and Tyler et al. (2005). Locations of the projections of the stimulus borders in retinotopic regions used the procedures described in Tyler et al. (2005) and Schira et al. (2007). Briefly, traveling wave stimuli—standard cyclic ring and wedge scans were used to define the eccentricity and meridian phase maps, respectively, throughout the occipital lobes. The retinotopic area borders were drawn by following the lines of phase reversal in the meridian maps, supplemented by the eccentricity information for some of the higher maps. The hMT+(V5+) motion complex was identified by a standard motion localizer using an expanding and contracting motion vector field of white dots on a black background, alternating with static dots. The projections of the figure borders were determined by taking into account the cortical magnification factor and specifying points where the corresponding eccentricities and meridians intersected; the borders were then drawn through these points.

To determine the differential signal for the test-stimulus relative to the null-stimulus epochs, the blood-oxygen-level-dependent (BOLD) fMRI signal was analyzed by extracting the Fourier fundamental of the time series at every voxel at the stimulus alternation rate of 1/18 Hz. (Note that the BOLD signal for blocks alternating at this rate closely approximates a sinusoid.) The initial response transient to stimulus onset was excluded by beginning visual stimulation 9 s before the analysis was initiated. A statistical correction for multiple occurrences was applied to the criterion for significant response, in terms of the amplitude of the Fourier fundamental. A coherence level of 0.47 provided a significance level of $P < 0.00005$ in each voxel, or a corrected level of $P < 0.05$ per 1,000 voxels to specify the presence of a significant response. Once a seed location had been established, adjacent voxels were included in the ROI down to a coherence level of 0.3 ($P < 0.01$ in each voxel).

Observers

The four observers, one female and three male, ranged in age from 28 to 59. They were all experienced with fMRI recording, two being naïve to the purpose of the study and

two being the authors. All were corrected to normal acuity for the 20 cm viewing distance by custom refractive lenses.

Experiment 1: Figure–ground configuration

To study figure/ground organization, the configuration of the *test*-stimulus was manipulated so as to satisfy Gestalt principle of categorizing the stimulus regions into figure versus ground. For this arrangement, a central horizontal region defined by two asynchrony borders was centered between two flanking surround regions (Fig. 1a). This configuration was perceived as a central thin bar of noise transients (figure; $8 \times 40°$) against larger background

fields of noise transients, as was reported by all the observers in the study. The *null*-epoch stimulus was a full field of synchronous transients (Fig. 1b) perceived as a uniform with no structure or segmentation. Note that the asynchronous *borders* in Fig. 1a have an identifiable cortical signature in terms of their projected locations in the early topographic areas (marked by white dashed lines in the case of V1 in Fig. 1c). The central bar *figure* forms a cortical projection region that is wider in the foveal representation and narrows toward the peripheral representation as a result of the cortical magnification function. The projection of the *ground* lies on the other side of the same borders extending dorsally and ventrally toward the V1 periphery.
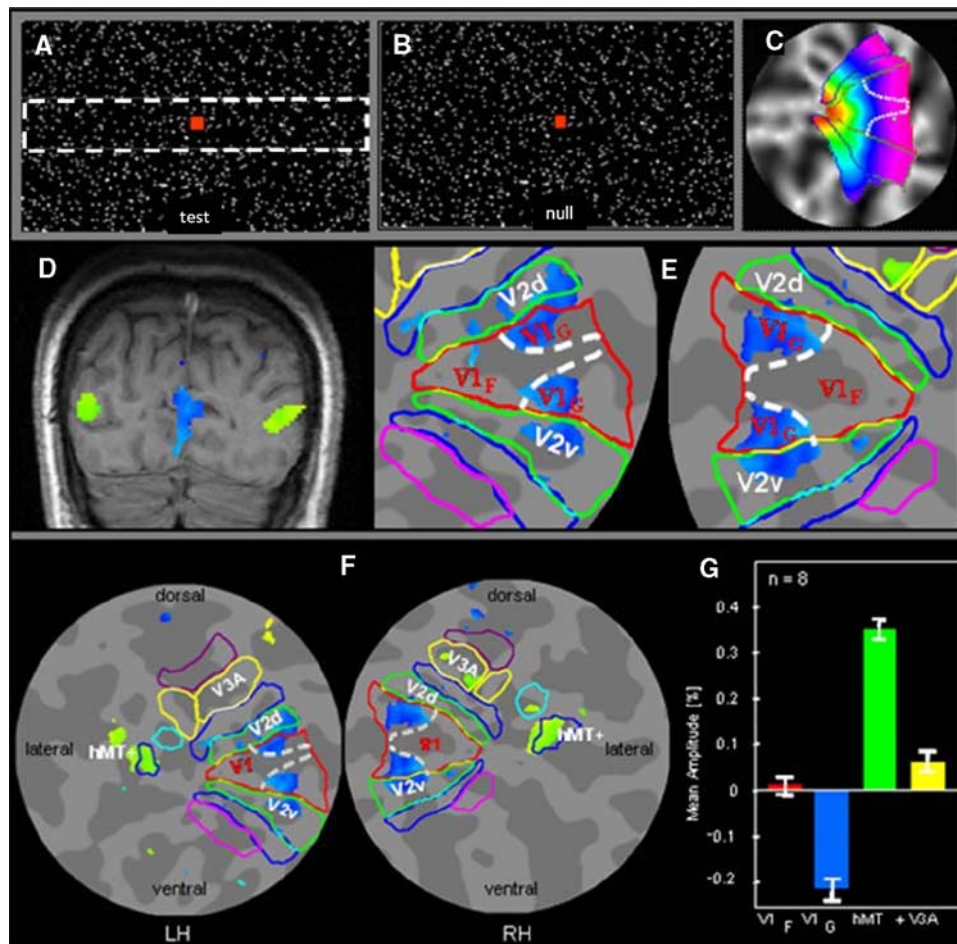

Fig. 1 **a** Figure–ground configuration achieved solely by updating the *random dots in a horizontal bar* in *asynchrony* to its surroundings. The dot-update asynchrony produces virtual borders which appear as sharp as motion or luminance borders (the *dashed lines* are used to delineate the *perceived* borders). **b** When the updates were *synchronized* across the whole stimulus field, however, no structure was generated. **c** The border projections (*white dashed lines*) in V1 localized on a cortex atlas of the retinotopic areas. **d** Coronal cortical slice showing activation (*greenish-yellow coloration*) and suppression (*bluish coloration*) in response to the **a** versus **b** stimulus contrast. **e** *Left and right hemisphere* flat cortical maps showing area V1 (*red outline*) and the retinotopic re-

gions around it. Note there is no significant signal in either the figure projection (*V1$_F$*) or along the border projections (*white dashed lines*) in retinotopic areas V1–3. All the signals are negative (bluish) and are largely restricted only to the retinotopic projections of the ground-regions in V1 (*V1$_G$*, for the lower and upper ground zones) and in V2 as well. **f** Full occipital flatmaps of the fMRI signals, showing that the only strong activation is in the hMT+ complex. Minor additional activation is seen in retinotopic V3A. **g** Average responses over eight hemispheres (±sem) for four cortical regions defined on the flatmaps: V1 projections of the figure (*V1$_F$*) and of the ground (*V1$_G$*), and areas hMT+, V3A

Note that it is solely the Gestalt figural properties, such as relative size and enclosedness, that determine which region is figure and which is ground. The essence of figure/ground organization is to assign an increased salience to the figure relative to the ground regions. In contrast to neurophysiological studies (e.g., Lamme 1995), to date no fMRI study has reported a response that is retinotopically matched with an arbitrary figure shape, in the absence of differentiating local cues.

The main hypotheses as to how figure/ground differentiation could be represented in the fMRI signal are the following:

(a) by retinotopic **enhancement** along the *borders* of the figure representation,
(b) by retinotopic **enhancement** throughout the region categorized as *figure*,
(c) by retinotopic **suppression** throughout the region categorized as *ground*,
(d) by both *figure* **enhancement** and *ground* **suppression**, or
(e) by **non-retinotopic activation** restricted to *object-specific* cortex.

The experiment was designed to test which of these hypothetical mechanisms operates in the human brain.

Results

The colored patches in Fig. 1d–f denote cortical regions of significant BOLD signal (p < 0.01, corrected for multiple applications within each flatmap; see Supplementary Fig. 1 for more individual examples). The white dashed lines within the V1 areas (Fig. 1c, e, f) indicate the projected shape of the horizontal *SfA* borders shown in Fig. 1a. It can be seen in Fig. 1e that there was no significant BOLD activation (greenish-yellow phases) specific to the locus either of the border projection or of the figure projection (V1$_F$) in the early retinotopic areas V1–V3. This result rules out Hypotheses a, b and d listed above. The effect of the figure/ground stimulation in early visual cortex was to generate a negative signal (blue coloration) in the peripheral regions of V1 (and V2). Inspection of the negative signals on the flatmap reveals that the suppression pattern is well structured: it starts from the borders of the figure and expands throughout the projection of the upper and lower representations of the flanking ground regions. In addition, there was only one area of strong positive activation—area hMT+ (Fig. 1d, f), together with a weak signal in a small part of V3A/B.

This response pattern was evaluated quantitatively in eight hemispheres of four observers. The negative response in the regions defined as the ground projection in V1 (i.e., V1$_G$ regions) was statistically significant at $P < 0.01$ across the eight hemispheres studied (Fig. 1g, blue bar), with no

significant positive (or negative) response in the V1 projection of the figure area V1$_F$ (Fig. 1g, red bar), thus supporting Hypothesis (c). Conversely, the signal in hMT+ showed significant and robust positive activation across the eight hemispheres (Fig. 1g, green bar), while V3A/B was significantly but weakly activated across the sample, averaging less than 20% of the mean hMT+ activation.

The fact that the only substantial activation was in hMT+ suggests that this cortical area might be the core site of figure/ground processing, and consequently, the origin of top-down specification of the retinotopic ground suppression in V1 and V2. Since the local transients were equally present throughout all stimulus regions in both test and null epochs, they could not have given rise to any differential fMRI signal. Instead, it seems that the hMT+ activation must to be specific to the global figure/ground organization.
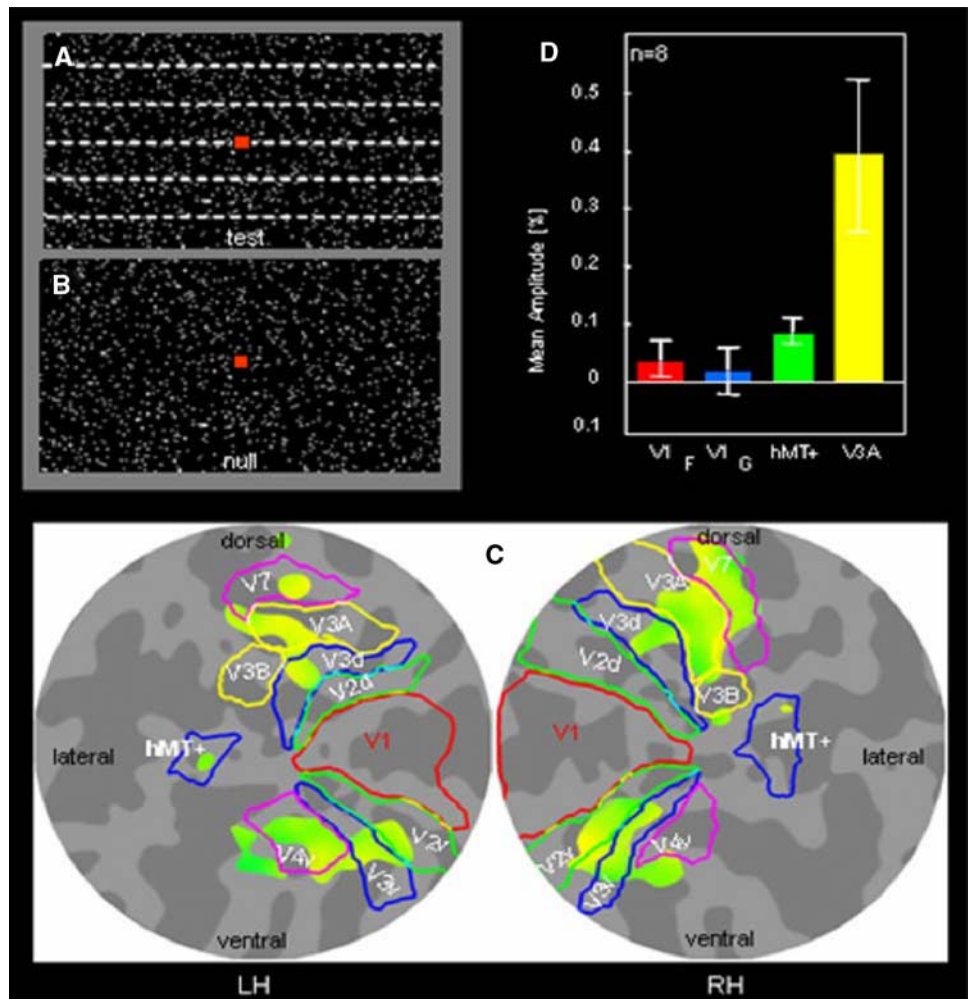
Experiment 2: Eliminated figure/ground: multi-bar configuration.

To determine whether the coupling of retinotopic suppression in the ground projection in V1/V2 and hMT+ activation reflects the process of figure/ground categorization, we looked for a test-stimulus configuration that (1) eliminates the global figure/ground organization, but (2) preserves and even increases the segmentation "work". Thus, in a second experiment the number of temporal asynchrony borders was increased so as to segment the random-dot field into multiple regions of *equivalent* horizontal stripes (Fig. 2a), such that none had the privileged status of a Gestalt "figure". There were now six stripes of equal width spanning the field, with the central border aligned with the fixation point. This manipulation should *eliminate* the responses specific to the *figure/ground* configuration, while conversely it should *increase* the strength of the responses for both *virtual border processing* and *regional segmentation*. Thus, the second experiment provides a test for the double dissociation between mechanisms of figure/ground categorization versus those of asynchrony border specification and segmentation. All observers reported that there was no tendency for one bar to stand out from the equally spaced array as a figure during the runs, as during the preceding lab testing of the stimuli. The null-stimulus (Fig. 2b) was the same as in Experiment 1 (Fig. 1b), and the same four observes participated.

Results

The seemingly simple configural change from the test-stimulus for Experiment 1 (Fig. 1a) to the test-stimulus for Experiment 2 (Fig. 2a) resulted in a dramatic reorganization of the activation pattern throughout the occipital

Fig. 2 a *Test* stimulus configuration for the control experiment, in which the addition of more equivalent SfA bars and borders eliminates the percept of a distinct figure against a background while providing stronger stimulation for border- and segmentation coding mechanisms. **b** The *null* stimulus was the field of synchronously updated random dots as in Experiment 1, perceived as uniform surface. **c** Elimination of the figure/ground organization in the asynchrony structure resulted in a radically different pattern of cortical activation. That result confirmed the predictions described in the text: both the activation of hMT+ and the suppression of V1/V2 disappeared, while, in contrast, the V3A signal was enhanced. In addition, there was strong activation in higher dorsal (V3, V7) and higher ventral retinotopic areas (V2v, V3v, V4) in all observers. **d** Average responses over eight hemispheres (±sem) for the same four cortical regions explored in Exp. 1

cortex. Figure 2c provides a typical example of the activation pattern on a cortical flatmap centered on the occipital pole (see Supplementary Fig. 2 for more individual examples). Both the activation pattern (Fig. 2c) and the bar graph of the average signal over the eight hemispheres (Fig. 2d) supported the predictions of the figure/ground elimination hypothesis:

> (a) drastic reduction or elimination of both the hMT + activation and the retinotopic V1/V2 ground suppression, but
> (b) strongly increased activation in clusters of higher extrastriate areas (V3, V3A/B, V7 dorsally and V2v, V3v, hV4, VMO ventrally, designated in the figures as V3A+ and V4+, respectively).

Thus the reorganization of the activation pattern was consistent with the predictions of perceptual reorganization, and conversely, inconsistent with a preserved figure/ground organization. If there had been any tendency for the perceptual system to pick one of the central bars of the equal array as figure, the prediction would be that both the hMT+ activation

and some tendency for peripheral suppression in V1/V2 should be preserved. Since both are completely eliminated by the new test configuration, such a prediction is not supported.

An alternative hypothesis is that each stripe alternates between figure and ground over time. In this case, the prediction from the results of Experiment 1 would be that the inhibition was the average of the relative figure activation and the surround suppression, that is, half the surround inhibition from Fig. 1, since there was no significant figure activation. In fact, this prediction is the lower bound on the amount of inhibition, because the presence of "figure" stripes on either side of each "ground" stripe should be expected to roughly double the amount of inhibition relative to that for the single stripe in Experiment 1. In either case, since there is no significant inhibition at all in either V1 or V2, this prediction is completely unsupported by the results (see Fig. 2d). In addition, the motion complex hMT+, while strongly activated in Experiment 1, did not show any significant activation here. We conclude that this alternative hypothesis is not supported by the data.

Analysis: a recurrent figure/ground model

The nature of the temporal-asynchrony paradigm of alternation between *SfA* (*test-stimuli*) and the non-*SfA* (*null-stimuli*) in a block design allowed us to equate the *local* features (such as update frequency, local temporal structure, dot density, transients, luminance, contrast, etc.) between the test and null epochs, and consequently to leave only *global* perceptual features as a basis for the BOLD signal. These global features include (1) perceived borders, (2) perceived segmentation into regions of synchronous updates, and (3) figure/ground relationships between the *SfA* regions. Experiments 1 and 2 were designed to dissociate the responses to figure/ground from those to borders and segmentation per se.

A core result is that under the *figure/ground* condition (Experiment 1) there was a pronounced negative signal in the earliest visual areas V1 and V2. Moreover, this signal was not in the retinotopic projections of the asynchrony borders, or of the figure region, but it was in the retinotopic projections of the regions *surrounding* the figure (Experiment 1), and it was eliminated when the figure/ground organization was eliminated (Experiment 2). This result implies that the negative signal must reflect processing specific to the figure/ground organization. Since the negative signal was the primary signal in the retinotopic areas, its interpretation needs to be considered before a model of the figure/ground processing can be proposed.

The origin of the negative BOLD response has not been fully understood and it has generated much theoretical interest. Recent studies have provided evidence of a neural component to the inhibitory rebound. Shmuel et al. (2006, 2002) have studied the origin of the negative BOLD activation in cortical regions adjacent to those receiving direct projections from stimulated retina and shown that it cannot be attributed to blood dynamics but must be a result of a reduction in the neuronal signal in those regions. Such studies imply that it is reasonable to attribute negative BOLD signals to inhibitory suppression of neural activity, especially in the present situation when (1) there is no difference in the local stimulation between the test- and the null-epochs, (2) and the only sites of activation are remote from the zone of reduced BOLD response. What is the source and the function of that extended retinotopic suppression? Although it is possible (in principle) that local center/surround receptive fields in V1/V2 could respond in the region of the perceived contours of the present study, they do not incorporate a mechanism that could generate the large *ground* suppression reported here. The large area of suppression suggests that lateral long-range interactions are also an unlikely candidate mechanism, because of the speed requirements to the

processing. This extended suppression must derive from a mechanism able to categorize all regions of the cortical representation on the basis of the global Gestalt relations between them and that can send or withhold the inhibitory signal back to the requisite locations of V1 and V2 accordingly. As will be detailed in the General Discussion, there are both neurophysiological and anatomical data suggesting that most likely a top-down feedback from a higher area, that is able to implement the global Gestalt principles, underlies the figure/ground process (Mumford 1992; Lamme 1995; Lamme et al. 1998; Hupé et al. 2001; Craft et al. 2007).

The only plausible source of the retinotopic ground suppression in V1/V2 observed in our fMRI data is area hMT+, because hMT+ was the sole region of occipital cortex that showed substantial BOLD activation in the figure/ground conditions (Experiment 1), and that activation was eliminated by eliminating the figure/ground organization (Experiment 2). It makes sense that hMT+ would be involved in 'structuring' our perception of dynamic images since it is equipped for processing transient-based signals; as the homolog of the MT/MST complex in monkey, it has a plenitude of large-scale receptive fields available for the purpose, and it is well established anatomically that there are strong reciprocal connections between MT and V1/V2. The results in Experiment 1 show how such feedback-based categorization can utilize the topographic nature of the early visual representations. What is achieved by suppressing the ground is an increased salience of the figure relative to the ground. In principle, the relative figure salience could be increased by facilitating the signal in the figure representation. Interestingly enough, however, from a computational perspective, it is a significantly more effective strategy to suppress the irrelevant information rather than to facilitate the relevant information.

The only other occipital area activated at the same criterion level as hMT+ in Experiment 1 was a small part of V3A (Fig. 1 and similarly Supplementary Fig. 1). We addressed the role of this activation by increasing the number of *SfA* contours in Experiment 2, so as to eliminate the figure/ground interpretation, while increasing the processing requirements for segmentation. The dramatic reorganization of the activation pattern from the figure/ground network of Experiment 1 (hMT+, V1/V2) to higher dorsal and ventral areas of the retinotopic hierarchy (Fig. 2c) implies that the activation in these higher areas may have a direct role in the processing of the *SfA* contours or in the segmentation of the regions they define. Note that the pattern of responses also excludes V3A as a source of the retinotopic suppression, since its activation *increases* when the suppression *disappears*.

In summary, the pattern of these results suggests a novel recurrent model for the organization of figure/
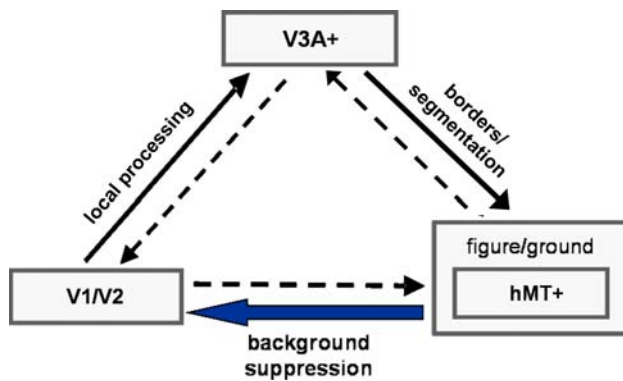
Fig. 3 Recurrent model of interactions among hMT+, V1/V2 and higher dorsal and ventral extrastriate areas. The local signals travel from V1 to V3A+ and V4+ (note that here the "+" suffix means "and beyond or around" the specified area), where the asynchrony border linkage and areal segmentation may take place. Local transients (in the temporal asynchrony paradigm) also travel to hMT+ but have little effect on the differential BOLD response there because any transient and motion signals are balanced. The segmentation signals from the higher extrastriate areas are sent on to hMT+, which is proposed as the main site involved in categorizing the image regions as figure versus ground and sending a suppressive feedback signals toward the retinotopic ground projections in V1/V2. This diagram illustrates the proposed flow of information during an iterative process of categorizing the image regions. *Dashed arrows* indicate known neural connections that may be involved in refining the response properties in an iterative manner, but are not needed to account for the present data. The basic functions performed in each area are specified

ground processing (Fig. 3). In this model, hMT+ implements the Gestalt principles for the SfA figure/ground categorization (such as relative size and enclosedness in our displays), and suppresses the signal in the V1/V2 retinotopic projections of the ground, while higher-retinotopic areas (V3A+, hV4+) are involved in the asynchronous borders and the segmentation.

General discussion

The mechanisms of perceptual organization are basic and of high importance in particular for theories of shape, object recognition and attention. The image region categorized by the brain as a figure 'owns' the contour, appears closer to the observer, has a specified shape, and has increased salience; while the ground has no contour, appears farther away, is 'shapeless' and is perceived as extending behind the figure and the contour (Rubin 1921/2001). The figure/ground categorization thus is considered to precede and govern attention (Palmer 1999; Craft et al. 2007).

In spite of the long-standing interest in figure/ground, its neural mechanism is still not fully resolved. The reported experiments demonstrate, that if the relations in a multi-element image correspond to Gestalt principles for figure/

ground categorization, the early visual representation of the 'less important' *ground* is suppressed, thus increasing the relative salience of the 'more important' *figure* region. The response configuration implies that the suppression is generated by a top-down feedback from hMT+, rather than by lateral interactions in the retinotopic areas, offering the first concrete fMRI evidence of the specific network of figure/ground categorization in the human brain.

How does this novel finding comport with other forms of evidence? Is the link between V1/V2 and hMT+ a plausible mechanism for contextual influence in general? Anatomically, both feedforward and feedback connections between V1/V2 and MT are well-established by retrograde tracer techniques (Tigges et al. 1981; Maunsell and van Essen 1983; Ungerleider and Desimone 1986; Krubitzer and Kaas 1990; Rockland and Van Hoesen 1994; Angelucci et al. 2002). MT is one of the primary output pathways for layer six neurons of these early retinotopic regions (Rockland and Van Hoesen 1994; Rockland and Knutson 2000). Neurophysiologically, a role for MT/MST in figure/ground assignment for structure-from-motion has been suggested for the past two decades (Allman et al. 1985; Olavarria et al. 1992; Eifuku and Wurtz 1999; Born 2000; Born and Bradley 2005). Rapid feedback to V1 from areas with large receptive fields, such as V2 or MT, was proposed in many cases (Knierim and Van Essen 1992; Lamme 1995; Zipser et al. 1996; Lee et al. 1998; Nothdurft et al. 1999, 2000; Hupé et al. 2001; Craft et al. 2007). The connections between V1 and MT were found to conduct at the high speed of ~10 m/s (Gilbert and Wiesel 1983, 1989; Born and Tootell 1991; DeAngelis et al. 1994; Grinvald et al. 1994; Fitzpatrick 2000; Anderson et al. 2001; Girard et al. 2001; Bair et al. 2002). In a center-surround interaction paradigm, Bair et al. (2002) showed that the temporal characteristics of V1 surround suppression suggest two suppression mechanisms operating on different spatial scales, with the one mediated by feedback being the most plausible to subserve global interactions (with MT as a likely candidate). In a figure/ground paradigm, evidence for causal MT–V1 coupling was provided by MT cooling (Hupé et al. 1998), which affected neural responses in V1/2/3. Thus, both anatomical and neurophysiological studies show that the neural circuit between V1/V2 and MT is well suited to subserve rapid top-down feedback.

Figure/ground contextual modulation, however, seems to require awareness and can be eliminated under deep anesthesia (Lamme et al. 1998). This implies (1) that different mechanisms underlie local and global processing in V1 and (2) that higher-level mechanisms are critical for figure/ground assignment. To account for the neurophysiological evidence on global contextual effects in V1, a mechanism with a feedback from (non-established) extra-striate areas was proposed for both coding schemes, that is, for the regional figure/

ground determination (Mumford 1992; Lamme 1995; Lamme et al. 1998), and for the border ownership signals in V2 as well (Zhou et al. 2000; Craft et al. 2007).

However, evaluating the neurophysiological predictions requires a large-scale method, such as fMRI, to provide the ability to analyze the global pattern of activation and suppression across the cortex, and to reveal the full activated network. Thus, our fMRI data are consistent (1) with the hypothesis of involvement of top-down feedback, (2) with the nature of that feedback being suppressive and (3) with region-based coding for figure/ground that effectively utilizes the retinotopic nature of the visual input.

How hMT+ performs the figure/ground categorization, however, remains to be determined, although some suggestions could be made. Regional segregation and integration constitute a dual process and different models focus more on one or on the other. It is well known psychophysically that the visual system is able to "capture" slight regularities in a case of integrating the trajectory of a single coherently moving dot in a field of many randomly moving dots (Verghese and McKee 2002). It is inherent to the motion-processing function of hMT+ to extract spatiotemporal correlations across the visual field. It is also likely that hMT+ is looking for coherence on a *global* scale and is well equipped to segment by integration over regions of similar synchronous changes, based on its 'expertise' in spatiotemporal integration. Once the image is segmented into regions some mechanism should be able to compare and analyze the Gestalt relations between them, including their figure/ground relations. MT neurons appear to be involved in the perceptual decision process, for instance in the case of perceptually bistable three-dimensional figures (Dodd et al. 2001).

The results reported here suggest that the retinotopic nature of the early areas is effectively utilized in the salience mapping of the visual field (Lu and Sperling 2001). Such salience mapping is underlain by a global mechanism (and thus, being independent of local cues), and has been proposed to form the input to subsequent processing of the stimulus field. The feedback suppression described herein would play an essential role in the organization of the salience map because suppressing the ground would reduce the salience of the less relevant regions, and would tend to leave only the figure regions in the competition for the currently most-salient structure in the field, that will subsequently govern attention. Interestingly, the concept of suppressing the information from the irrelevant surrounding regions, rather than figural enhancement, has been implemented in a computational model of Tsotsos et al. (1995) and Tsotsos (1997). Their simulations showed the computational efficiency of such suppression early in the processing sequence to avoid wasting later processing capacity. Moreover, recent behavioral studies (Peterson

and Salvagio under review; Salvagio et al. 2008) add behavioral evidence for a form of ground suppression and emphasize the importance of suppression in figure–ground perception.

Conclusions

Natural visual scenes typically contain multiple objects that need to be parsed as separate from their background. We have focused on revealing the cortical network for the perceptual organization mechanism of figure/ground, which increases the relative salience of a target object by the strategic mechanism of suppressing its irrelevant surrounding. The data of this study suggest the involvement of a recurrent brain architecture incorporating a suppressive top-down feedback from hMT+ to V1/V2 that is organized in a topographic manner. The pattern of the BOLD signals through V1 and V2 implies that these retinotopic maps play a critical role in determining the salience map of the visual scene forming the input to subsequent recognition and attention processing. The proposed recursive architecture has the advantages of computational economy (because irrelevant responses are inhibited), efficiency (because the process of salience re-weighting pre-filters the signals for the winner-take-all selection of the figure), topographical organization (because it operates in strongly retinotopic regions, such as V1 and V2), ubiquity (because the re-weighting is set at the input visual areas, and thus infuses all subsequent cortical processing) and flexibility (in the interplay of the figure/ground salience weighting by both bottom-up feature processing and top-down suppressive modulation). Thus, the present results offer a substantial new component for models of the complex operations underlying human object recognition.

References

Allman J, Miezin F, McGuinness E (1985) Direction- and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). Perception 14:105–126

Anderson JS, Lampl I, Gillespie DC, Ferster D (2001) Membrane potential and conductance changes underlying length tuning of cells in cat primary visual cortex. J Neurosci 21:2104–2112

Angelucci A, Levitt JB, Walton EJ, Hupe JM, Bullier J, Lund JS (2002) Circuits for local and global signal integration in primary visual cortex. J Neurosci 22:8633–8646

Baek K, Sajda P (2005) Inferring figure–ground using a recurrent integrate-and-fire neural circuit. IEEE Trans Neural Syst Rehabil Eng 13:125–130

Bair W, Cavanaugh JR, Smith MA, Movshon JA (2002) The timing of response onset and offset in macaque visual neurons. J Neurosci 22:3189–3205

Blake R, Lee SH (2005) The role of temporal structure in human vision. Behav Cogn Neurosci Rev 4:21–42

Born RT, Bradley DC (2005) Structure and function of visual area MT. Annu Rev Neurosci 28:157–189

Born RT, Tootell RBH (1991) Single-unit and 2-deoxyglucose studies of side inhibition in macaque striate cortex. Proc Natl Acad Sci USA 88:7071–7075

Born RT (2000) Center–surround interactions in the middle temporal visual area of the owl monkey. J Neurophysiol 84:2658–2669

Craft E, Schutze H, Niebur E, von der Heydt R (2007) A neural model of figure–ground organization. J Neurophysiol 97:4310–4326

DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. J Neurophysiol 71:347–374

Dodd JV, Krug K, Cumming BG, Parker AJ (2001) Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area MT. J Neurosci 21(13):4809–4821

Eifuku S, Wurtz RH (1999) Response to motion in extrastriate area MSTl: disparity sensitivity. J Neurophysiol 82:2462–2475

Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cerebral Cortex 7:181–192

Fahle M (1993) Figure–ground discrimination from temporal information. Proc Roy Soc B Biol Sci 254:199–203

Fitzpatrick D (2000) Seeing beyond the receptive field in primary visual cortex. Curr Opin Neurobiol 10:438–443

Gerrits H, Vendrik A (1970) Simultaneous contrast, filling-in process and information processing in man's visual system. Exp Brain Res 11:411–430

Gilbert CD, Wiesel TN (1983) Clustered intrinsic connections in cat visual cortex. J Neurosci 3:1116–1133

Gilbert CD, Wiesel TN (1989) Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. J Neurosci 9:2432–2442

Girard P, Hupé JM, Bullier J (2001) Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. J Neurophysiol 85:1328–1331

Grinvald A, Lieke EE, Frostig RD, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. J Neurosci 14:2545–2568

Grossberg S, Mingolla E (1985) Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. Psychol Rev 92:173–211

Grossberg S (1994) 3-D vision and figure–ground separation by visual cortex. Percept Psychophys 55:48–120

Hupé J-M, James AC, Girard P, Lomber SG, Payne BR, Bullier J (2001) Feedback connections act on the early part of the responses in monkey visual cortex. J Neurophysiol 85:134–145

Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394:784–787

Kandil FI, Fahle M (2001) Purely temporal figure–ground segregation. Eur J Neurosci 13:2004–2008

Kandil FI, Fahle M (2003) Mechanisms of time-based figure–ground segregation. Eur J Neurosci 18:2874–2882

Knierim JJ, Van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol 67:961–980

Krubitzer LA, Kaas JH (1990) Cortical connections of MT in four species of primates: areal, modular, and retinotopic patterns. Vis Neurosci 5:165–204

Lamme VA (1995) The neurophysiology of figure–ground segregation in primary visual cortex. J Neurosci 15:1605–1615

Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feed-forward and recurrent processing. Trends Neurosci 23:571–579

Lamme VA, Zipser K, Spekreijse H (1998) Figure–ground activity in primary visual cortex is suppressed by anesthesia. Proc Natl Acad Sci USA 95:3263–3268

Lamme VA, Zipser K, Spekreijse H (2002) Masking interrupts figure–ground signals in V1. J Cogn Neurosci 14:1044–1053

Lee SH, Blake R (1999) Visual form created solely from temporal structure. Science 284:1165–1168

Lee TS, Mumford D, Romero R, Lamme VA (1998) The role of the primary visual cortex in higher level vision. Vis Res 38:2429–2454

Lu ZL, Sperling G (2001) Three-systems theory of human visual motion perception: review and update. J Opt Soc Am A 18:2331–2370

Maunsell JH, van Essen DC (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. J Neurosci 3:2563–2586

Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. Biol Cybern 66:241–251

Nothdurft H-C, Gallant JL, Van Essen DC (1999) Response modulation by texture surround in primate area V1: correlates of "pop-out" under anesthesia. Vis Neurosci 16:15–34

Nothdurft H-C, Gallant JL, Van Essen DC (2000) Response profiles to texture border patterns in area V1. Vis Neurosci 17:421–436

Olavarria JF, DeYoe EA, Knierim JJ, Fox JM, van Essen DC (1992) Neural responses to visual texture patterns in middle temporal area of the macaque monkey. J Neurophysiol 68:164–181

Palmer SE (1999) Vision science: from photons to phenomenology. Bradford Books/MIT Press, Cambridge, MA

Pao HK, Geiger D, Rubin N (1999) Measuring convexity for figure/ground separation. In: Proceedings of 7th international conference on computer vision, Kerkyra, Greece

Peterson MA, Salvagio E (under review) Context enhances effectiveness of convexity as a figural cue: evidence for biased competition in figure-ground perception

Press WA, Brewer AA, Dougherty RF, Wade AR, Wandell BA (2001) Visual areas and spatial summation in human visual cortex. Vis Res 41:1321–1332

Rockland KS, Knutson T (2000) Feedback connections from area MT of the squirrel monkey to areas V1 and V2. J Comp Neurol 425:345–368

Rockland KS, Van Hoesen GW (1994) Direct temporal-occipital feedback connections to striate cortex (V1) in the macaque monkey. Cereb Cortex 4:300–313

Roelfsema PR, Lamme VA, Spekreijse H, Bosch H (2002) Figure–ground segregation in a recurrent network architecture. J Cogn Neurosci 14:525–537

Rubin E (1921/2001) Visuell wahrgenommene Figuren. Copenhagen: Gyldendalske; reproduced In: Yantis S (ed) Visual perception: essential readings, vol 12. Psychology Press, Philadelphia, pp 225–229

Salvagio E, Mojica AJ, Peterson MA (2008) Context effects in figure-ground perception: the role of biased competition, suppression and long-range connections [Abstract]. J Vis 8(6):1007, 1007a. http://journalofvision.org/8/6/1007

Schira MM, Wade AR, Tyler CW (2007) Two-dimensional mapping of the central and parafoveal visual field to human visual cortex. J Neurophysiol 97:4284–4295

Shmuel A, Yacoub E, Pfeuffer J, Van de Moortele PF, Adriany G, Hu X, Ugurbil K (2002) Sustained negative BOLD, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. Neuron 36:1195–1210

Shmuel A, Augath M, Oeltermann A, Logothetis NK (2006) Negative functional MRI response correlates with decreases in neuronal activity in monkey visual area V1. Nat Neurosci 9:569–577

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889–893

Super H, Spekreijse H, Lamme VA (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat Neurosci 4:304–310

Tigges J, Tigges M, Anschel S, Cross NA, Letbetter WD, McBride RL (1981) Areal and laminar distribution of neurons interconnecting the central visual cortical areas 17, 18, 19, and MT in squirrel monkey (Saimiri). J Comp Neurol 202:539–560

Tsotsos JK, Culhane SM, Kei Wai WY, Lai Y, Davis N, Nuflo F (1995) Modeling visual attention via selective tuning. Artif Intell 78:507–545

Tsotsos JK (1997) Limited capacity of any realizable perceptual system is a sufficient reason for attentive behavior. Conscious Cogn 6:429–436

Tyler CW, Likova LT, Kontsevich LL, Chen CC, Schira MM, Wade AR (2005) Extended concepts of occipital retinotopy. Curr Med Imaging Rev 1:319–329

Tyler CW, Likova LT, Kontsevich LL, Wade AR (2006) The specificity of cortical area KO to depth structure. NeuroImage 30:228–238

Ungerleider LG, Desimone R (1986) Cortical connections of visual area MT in the macaque. J Comp Neurol 248:190–222

Verghese P, McKee SP (2002) Predicting future motion. J Vis 2:413–423

von der Heydt R, Zhou H, Friedman HS (2003) Neural coding of border ownership: Implications for the theory of figure–ground perception. In: Behrmann M, Kimchi R, Olson CR (eds) Perceptual organization in vision: behavioral and neural perspectives. Lawrence Erlbaum Associates, Mahwah, pp 281–304

von der Heydt R, Macuda TJ, Qiu FT (2005) Border-ownership dependent tilt aftereffect. J Opt Soc Am (A) 22:2222–2229

Wertheimer M (1923/1938/1997) Laws of organization in perceptual forms. First published as Untersuchungen zur Lehre von der Gestalt II. Psycol Forschung 4:301–350. Trans in Ellis W (1938) A source book of Gestalt psychology, Routledge & Kegan Paul, London, pp 71–88. Reprint (1997) The Gestalt Journal Press, New York

Wertheimer M (1925) Gestalt theory. In: Ellis W (ed) Trans (1938 reprinted 1997). A source book of Gestalt psychology Routledge & Kegan Paul, London, pp 1–11. Reprint (1997) The Gestalt Journal Press, New York

Zhaoping L (2005) Border ownership from intracortical interactions in visual area V2. Neuron 47:143–153

Zhou H, Friedman HS, von der Heydt R (2000) Coding of border ownership in monkey visual cortex. J Neurosci 20:6594–6611

Zipser K, Lamme VA, Schiller PH (1996) Contextual modulation in primary visual cortex. J Neurosci 16:7376–7389