

# Perceptual grouping and inverse fMRI activity patterns in human visual cortex

**Fang Fang**

Department of Psychology and Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, P.R. China



**Daniel Kersten**

Department of Psychology, University of Minnesota, Minneapolis, MN, USA



**Scott O. Murray**

Department of Psychology, University of Washington, Seattle, WA, USA



We used functional magnetic resonance imaging (fMRI) to measure activity in human visual cortex, including a higher object processing area, the lateral occipital complex (LOC), and primary visual cortex (V1), in response to a perceptually bistable stimulus whose elements were perceived as either grouped into a shape or randomly arranged. We found activity increases in the LOC and simultaneous reductions of activity in V1 when the elements were perceived as a coherent shape. Consistent with a number of inferential models of visual processing, our results suggest that feedback from higher visual areas to lower visual areas serves to reduce activity during perceptual grouping. The implications of these findings with respect to these models are discussed.

Keywords: V1, lateral occipital complex, perceptual grouping, feedback

Citation: Fang, F., Kersten, D., & Murray, S. O. (2008). Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *Journal of Vision*, 8(7):2, 2–9, <http://journalofvision.org/8/7/2/>, doi:10.1167/8.7.2.

## Introduction

A defining characteristic of human visual perception is the ability to assemble complex visual features—sometimes spatially separated and partially occluded—into coherent, unified representations of objects and surfaces. Grouping processes can vastly simplify the description of a visual scene because multiple features can be assigned to a single “cause.” For example, multiple lines of the same orientation can be described as a single texture without needing to specify each element within the pattern.

What are the neural mechanisms that underlie perceptual grouping? We have suggested (Murray, Schrater, & Kersten, 2004) that perceptual grouping involves increases in activity in higher visual areas that code for spatial patterns (e.g., objects, surfaces, and textures) along with decreases in activity in lower visual areas that code for local, individual elements of the pattern (e.g., local orientation or direction of motion). We propose that this inverse relationship in neural activity between higher and lower visual areas reflects an “efficient code” of visual information. As higher visual areas converge on a single, global hypothesis for the individual elements in a visual scene, lower visual areas no longer need to represent the individual elements.

A variety of computational models propose mechanisms that include interactions between high-level and low-level representations of image feature that result in inverse activity patterns. Predictive coding models (Mumford, 1992; Rao & Ballard, 1999), for example, are one class of models that suggest that feedback may operate to reduce activity in lower areas. Predictive coding models usually posit a subtractive comparison between hypotheses generated in higher areas and incoming sensory input in lower areas. In these models, reduced activity occurs when the predictions of higher-level areas match incoming sensory information. Recent fMRI and MEG studies have provided evidence supporting these models (Furl, van Rijsbergen, Treves, Friston, & Dolan, 2007; Harrison, Stephan, Rees, & Friston, 2007; Summerfield et al., 2006, but see also Dumoulin & Hess, 2006).

Consistent with feedback models, in an initial fMRI study, we observed signal increases in the lateral occipital complex (LOC)—a higher visual area well known for its involvement in shape and object processing (Grill-Spector, Kourtzi, & Kanwisher, 2001)—and concurrent signal reductions in primary visual cortex (V1) when visual elements were assembled into coherent shapes (Murray, Kersten, Olshausen, Schrater, & Woods, 2002). This inverse activity pattern between the LOC and V1 was observed both for images of lines and for images of moving dots that were arranged to form shapes (“grouped”)

compared to when they were randomly assembled (“ungrouped”). Although we had performed a number of control studies showing that the results were not due to various stimulus differences between the grouped and the ungrouped stimulus conditions, completely ruling out such differences is difficult.

One class of stimuli that control for image differences are image patterns with bistable perceptual properties. These stimuli provide a powerful tool for investigating the neural mechanisms underlying perceptual grouping because they have constant image features that spontaneously switch between being perceived as grouped into a single object and being perceived as ungrouped visual features. In the current experiment, we measured fMRI activity as subjects viewed a perceptually bistable stimulus with four moving line segments that were either perceived as a rigid diamond translating horizontally or perceived as ungrouped line segments. Earlier measurements (Murray et al., 2002) had shown that activity in V1 decreases when the line segments are perceptually grouped into a rigid diamond. The current study represents a significant advance in methodology, together with additional analyses of extrastriate areas. Across all of our subjects, we observed significant inverse activity patterns in the LOC and V1: Activity increases in the LOC during perceptual grouping were accompanied by activity decreases in V1.

## Methods

### Subjects

A total of four healthy subjects (3 male and 1 female) participated in the experiments, all of whom had extensive experience as subjects in psychophysical and fMRI experiments. They were right-handed, reported normal or corrected-to-normal vision, and had no known neurological or visual disorders. Ages ranged from 26 to 32. They gave written, informed consent in accordance with the procedures and the protocols approved by the human subjects review committee of the University of Minnesota.

### Stimuli and designs

The main experiment was performed using a constant image sequence that formed a changing bistable percept with either grouped or ungrouped line segments. The stimulus was a line drawing of a diamond whose four corners were occluded by three vertical bars of the same color as the background (Figure 1A). The diamond moved at a constant horizontal speed of 1.3 deg/s and reversed direction every 1 s. A similar stimulus was used by Lorenceau and Shiffrar (1992). The line segments were

2.6° in length, and their centers were at 2.8°–3.9° eccentricity. The stimulus could be perceived either as a rigid diamond moving horizontally behind occluders (diamond, Figure 1B) or as individual line segments moving vertically (non-diamond, Figure 1C). The two percepts alternated and subjects indicated their perceptual state with a button press. A total of 5–6 400 s scans were performed for each subject.

Retinotopic visual areas were defined by a standard method developed by Engel, Glover, and Wandell (1997) and Sereno et al. (1995). Two block-design scans were used to define the regions of interest (ROI). In one scan, a 10-Hz counterphase-flickering stimulus (Figure 2) was



Figure 2. A flickering checkerboard stimulus used for defining the subregion of V1 corresponding to the areas covered by the moving line segments.

coil. BOLD signals were acquired with an EPI (echo-planar imaging) sequence (TE: 30 ms; TR: 1000 ms; FOV:  $22 \times 22 \text{ cm}^2$ ; matrix:  $64 \times 64$ ; flip angle: 60; slice thickness: 5 mm; gap: 0.5 mm; number of slices: 10; slice orientation: axial). The slice was positioned at the bottom of the temporal pole. A high-resolution 3D structural data set (3D T1-weighted GE;  $1 \times 1 \times 1 \text{ mm}^3$  resolution) was collected in the same session before the functional scans. All five subjects participated in two fMRI sessions for the retinotopic mapping experiment and the main experiment, respectively.

## MRI data processing and analysis

The anatomical volumes from each subject in the retinotopic mapping session were transformed into the AC-PC space. The cortical surfaces were extracted and then inflated using BrainVoyager 2000. Functional volumes in all the sessions for each subject were preprocessed, which included 3D motion correction using SPM99, linear trend removal, and high-pass (0.01 Hz) (Smith et al., 1999) filtering using BrainVoyager 2000. The images were then aligned to the anatomical volume in the retinotopic mapping session and transformed into the AC-PC space.

The first 10 s of BOLD signals were discarded to minimize transient magnetic-saturation effects.

A GLM (general linear model) procedure was used for ROI analysis. The ROIs in V1, V2, and V3 were defined as areas that responded more strongly to the flickering checkerboard stimulus (Figure 2) than blank interval ( $p < 0.01$ , corrected) and confined by the retinotopic boundaries defined by the retinotopic mapping experiment. LOC was defined as areas in the occipital cortex that respond more strongly to object images than scrambled object images ( $p < 0.01$ , corrected). With this contrast, some temporal object areas (TOA) were also found to be sensitive to object images (Fang & He, 2005; Grill-Spector, 2003), which have been suggested to be at a higher position than the LOC in the hierarchy of visual system (Grill-Spector & Malach, 2001). TOA was defined as areas of activation in response to objects that were in more anterior and ventral locations than the LOC.

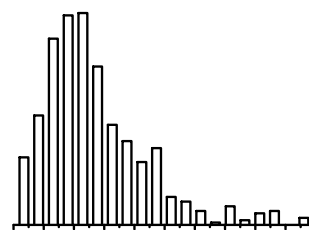
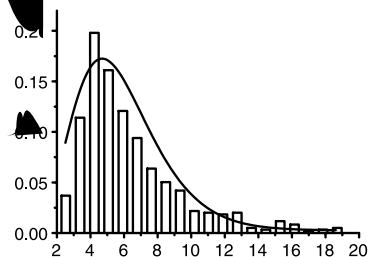
The event-related BOLD signals were calculated separately for each subject and each ROI. For each event-related scan, the time course of MR signal intensity was extracted by averaging the data from all the voxels within the pre-defined ROI, transformed into percent signal changes, and event-related averaged according to subject's response. Finally, the event-related signals were averaged across scans.

## Eye movement recording

Eye movements were recorded at 60 Hz with an iView X RED eye tracker (SensoMotoric Instruments GmbH, Teltow, Germany) in a psychophysics lab for two subjects when they viewed the same stimuli as those in the magnet.

## Results

We measured cortical activities in both earlier and higher visual areas while subjects viewed a bistable stimulus that was either perceived as a moving diamond



behind occluders or as individual moving line segments. Importantly, only the percept and not the stimulus features changed during this experiment. The perceptual states switched every several seconds. The mean durations of diamond and non-diamond percepts were 6.3 s and 7.3 s, respectively. [Figure 3](#) shows the frequency histograms of durations for the diamond and the non-diamond percepts. We find that the data are well fit with a Gamma distribution, which has been well documented in another bistable visual phenomenon—binocular rivalry (Kovács, Papathomas, Yang, & Fehér, 1996). This suggests that these two types of bistable perceptions have similar temporal dynamics.

Activities in V1 and LOC showed significant changes during perceptual switches but in inverse patterns ([Figure 4](#)).

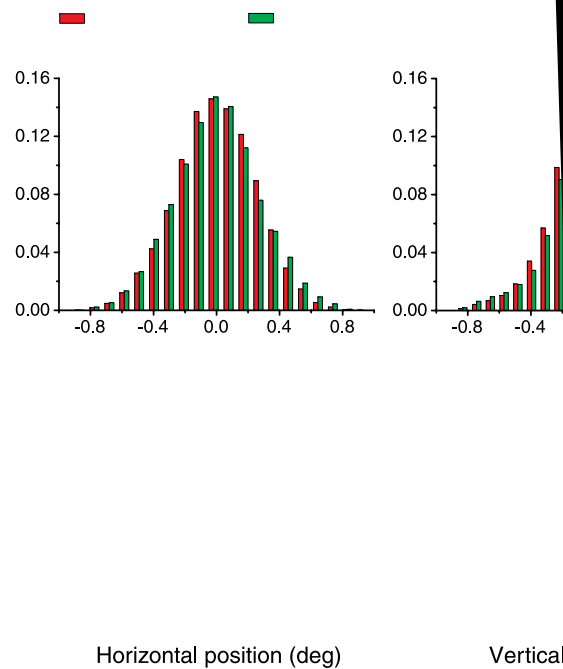
Activity in V1 was significantly reduced when a coherently moving diamond was perceived and significantly increased

with a smaller amplitude. V3 did not show a significant modulation associated with perceptual grouping.

Eye movements are a possible confound. We monitored two subjects' eye positions. [Figure 6](#) shows the frequency histograms of horizontal and vertical eye positions during the diamond and non-diamond percepts. The data show that their eye movements were small and there was no significant difference in the distribution of eye position between the diamond percept and the non-diamond percept. Also, further statistical analyses confirmed that both horizontal and vertical mean eye positions did not significantly deviate from the fixation point during both the diamond percept and the non-diamond percept. These results suggest that it is unlikely that our results could be significantly confounded by eye movements.

## Discussion

Our results demonstrate a significant and reliable pattern of fMRI activity associated with perceptual grouping: when moving line segments were perceived as a single, translating object, activity increased in the LOC and decreased in V1 compared to when the same line segments were perceived as ungrouped. The LOC activity pattern is expected because this cortical region is known to be shape-selective (e.g., Kourtzi & Kanwisher, 2001). The V1 activity pattern is consistent with our earlier finding (Murray et al., 2002). Taken together, these results suggest



Although our earlier study (Murray et al., 2002) included a condition using a similar bistable “translating diamond,” the current study represents a significant advance in methodology and analysis. Here we used an independently defined, retinotopically specific localizer for V1. Thus, we are confident that the modulations in the fMRI signal that we observed occurred in the retinotopic representation of the stimulus and not in immediately adjacent retinotopic regions (e.g., artifacts due to “blood-flow steal”). In addition, due to limited slice-selection, our previous report using the translating diamond only made measurements in V1. Here, using an independent localizer for the LOC, we show significant changes that inversely reflect the pattern of activity observed in V1. Finally, the current study employs event-related averaging that characterizes the temporal dynamics and the magnitudes of V1 and LOC changes in more detail than our previous report.

In addition to the LOC and V1, we analyzed the fMRI signal in V2, V3, and TOA, none of which showed the kinds of signal changes observed in the LOC and V1. The small modulation observed in TOA likely has a straightforward explanation—this region is considered to be at a relatively high level in the visual hierarchy and simple geometric shapes (e.g., the diamond) are unlikely to evoke much activity in this region. While V2 had a similar pattern of activity as V1, its amplitude was significantly reduced. V3 essentially had no change in signal in response to perceptual transitions. These observations are important as they point to a potentially unique computational role for V1 in perceptual grouping.

Given the convincing empirical demonstration of inverse activity patterns in V1 and the LOC, the current

findings raise important questions about the interpretation of the fMRI signal in V1 when the line is bistable. First, what implications do these findings have on the interpretation of the fMRI signal representing perceptual processes related to multiple object grouping (Logothetis & Wandell, 1996)? Second, what implications have been reported for the LOC (Trinath, & Oeltermann, 2000), our observations may be a manifestation of perceptual processes in addition to V1. Other, more direct, methods are needed to answer this question.

Second, do the anti-correlations between V1 and the LOC necessarily mean that the LOC is necessary for V1? More specifically, are the changes in V1 necessarily *caused* by the LOC? Although correlations are not sufficient to establish a relationship, if the LOC is necessary for the grouped elements, then changes in perception should be related to changes in the relationship between the LOC and V1. A more conclusive way to address this question is to remove feedback connections between the LOC and V1.

Third, and the most important, what are the perceptual reductions in V1 when the line is bistable? Although the modulations observed in V1 are known for shape



perception, making a similar argument for V1 is more difficult. V1 has traditionally been thought to maintain a veridical representation of retinal information. Consequently, a stimulus that has physically constant features—as with the translating diamond—is not generally expected to change V1 activity. We consider several alternative accounts of the potential functional significance of the V1 signal changes.

On one end of the spectrum of possibilities, the changes in V1 might not be functionally significant. For example, fMRI measurements of V1 have shown reliable signal changes associated with spatial attention. Is it possible that the changes we observed simply reflect incidental shifts in spatial attention that occur during perceptual transitions? This explanation would require that subjects directed their spatial attention away from the line segments when they perceived the diamond, relative to the non-diamond condition. There is no reason to believe that these shifts occurred. In fact, our subjects claimed that they needed to focus their attention on the line segments in order to perceive the diamond. However, future studies that explicitly manipulate spatial attention and its effect on perceptual grouping and the fMRI signal are warranted.

Along similar lines, the argument could be made that the differences in V1 and LOC activity might simply reflect attention to the features (“diamond” vs. “ungrouped line segments”) that result from the different perceptual states. For example, when subjects perceived ungrouped line segments they might have attended to this feature of the stimulus, consequently leading to more activity in V1 because it is presumably specialized for processing this feature. In contrast, when subjects perceived the diamond they might have attended to its overall shape leading to more activity in the LOC because of its specialization in shape processing. On one hand, attention to features is part of the process. During the perception of the diamond, subjects are certainly “attending to the diamond-ness” and separating the role of attention—which is directly tied to perceptual awareness—would be very difficult in our experimental setup. However, there is empirical evidence which renders a simple feature-based attention explanation unlikely. First, we observed notably diminished (V2) and abolished (V3) modulation of the fMRI signal in other early visual areas. There is no *a priori* reason to believe that these areas are any less specialized for the features of the “non-diamond” than V1. Second, Buracas, Fine, and Boynton (2005) compared fMRI responses in early visual cortex as subjects switched attention between different features (contrast vs. speed) of a moving grating. They found no modulation of the fMRI signal in any early visual area (V1, V2, V3, and MT) as a function of feature-based attention when, in theory, it might be expected. For example, early visual cortex is highly sensitive to contrast but attending to that feature did not modulate the fMRI signal. However, given the differences in underlying features in the Buracas et al. study (contrast and speed)

compared to our study (grouping of line segments) to fully address the potential contribution of feature-based attention will require future direct empirical tests. Such an experiment might alternate attention between local versus global elements of simple shapes (such as the diamond) and measure activity in both lower and higher visual areas.

An alternative interpretation of the decrease in V1 activity is that it might not have a direct functional significance but reveal a general metabolic efficiency constraint placed on neural processing. Spiking activity is metabolically expensive (Lennie, 2003) and there may be a general strategy to minimize neural activity whenever possible. For example, if one cortical area can represent the visual stimulus, another area should not. In our case, when the line segments form a representation that can be maintained in the LOC, V1 may participate less in the representation simply to minimize overall activity. Although sparseness constraints have been shown to have important theoretical implications related to the emergence of receptive field properties *within* a cortical area (Olshausen & Field, 1996), the implications of extending this principle to *between* areas are less clear.

Finally, the reductions in V1 activity observed during perceptual grouping may reveal important functional mechanisms of visual information processing. One such mechanism, mentioned in the [Introduction](#) section, is predictive coding (Mumford, 1992; Rao & Ballard, 1999). Predictive coding models posit that higher areas are actively attempting to “explain” activity patterns in lower areas via feedback projections. Because most predictive coding models include a subtractive comparison between the hypotheses formed in higher areas and the incoming sensory input represented in lower areas, the overall effect of feedback may be to reduce activity in lower areas. Specifically, reduced activity in lower visual areas would occur whenever the predictions of higher-level areas match incoming sensory information. In the case of the translating diamond, when the LOC maintains a representation of a grouped shape, this “expectation” or “understanding” of the image features is sent back to V1 and removed, resulting in less activity. When the LOC is unable to form such an understanding (i.e., when they are perceived as ungrouped), these feedback processes are not occurring and there is consequently more activity in V1.

In summary, although our results are consistent with a number of theoretical interpretations, they demonstrate that perceptual grouping involves activity modulations at multiple stages of the visual hierarchy. The two areas considered in detail here—the LOC and the V1—correspond to areas that are known to represent global shape and local visual features, respectively. Importantly, the activity patterns in these areas are inversely related and suggest that perceptual grouping involves both increases and decreases in activity in the human visual system.

## Acknowledgments

This work is supported by NIH grant R01 EY015261-01. The 3-T scanner at the University of Minnesota, Center for Magnetic Resonance Research, was supported by NCRR P41 008079 and P30 NS057091 and by the MIND Institute.

Commercial relationships: none.

Corresponding author: Scott O. Murray.

Email: somurray@u.washington.edu.

Address: Department of Psychology, Box 351525, University of Washington, Seattle WA 98195, USA.

## References

- Buracas, G. T., Fine, I., & Boynton, G. M. (2005). The relationship between task performance and functional magnetic resonance imaging response. *Journal of Neuroscience*, *25*, 3023–3031. [[PubMed](#)] [[Article](#)]
- Dumoulin, S. O., & Hess, R. F. (2006). Modulation of V1 activity by shape: Image-statistics or shape-based perception? *Journal of Neurophysiology*, *95*, 3654–3664. [[PubMed](#)] [[Article](#)]
- Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, *7*, 181–192. [[PubMed](#)] [[Article](#)]
- Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature Neuroscience*, *8*, 1380–1385. [[PubMed](#)]
- Furl, N., van Rijsbergen, N. J., Treves, A., Friston, K. J., & Dolan, R. J. (2007). Experience-dependent coding of facial expression in superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 13485–13489. [[PubMed](#)] [[Article](#)]



- functional magnetic resonance imaging. *Science*, 268, 889–893. [[PubMed](#)]
- Smith, A. M., Lewis, B. K., Ruttimann, U. E., Ye, F. Q., Sinnwell, T. M., Yang, Y., et al. (1999). Investigation of low frequency drift in fMRI signal. *Neuroimage*, 9, 526–533. [[PubMed](#)]
- Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, 314, 1311–1314. [[PubMed](#)]