P r r r MRI r r
r

F r F r

D r K r

S O. M

(2008). . J a (7) 2, 2– //j /8/7/2/, 10.1167/8.7.2.

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 percep-

tual grouping? We 642(pa d6-64]TJ/6S1 191(fi16,96128)0TJ*(ay)84(e3ac)1/6(1)-1304(mo6Tf(9t)-4m0260T9iC5-g-17.94)4(lbicad[(pr)a)1979-258(i
processessualr-54(iox.)]TJ/T72n0(lowe)632re9(e7e(surf471al)-3632”ef)53J9(esciin)-2T72n(p)-8(exam46i)-9(o632x.)]TJ4Td-258(i

compared to when they were randomly assembled coil. BOLD signals were measured with an EPI (echo-planar imaging) sequence (TE: 30 ms; TR: 1000 ms; FOV: 22 $\times$ 22 cm^2; matrix: 64 $\times$ 64; flip angle: 60; slice thickness: 5 mm; gap: 0 mm; number of slices: 10; slice orientation: axial). The bottom slice was positioned at the bottom of the temporal lobes. A high-resolution 3D structural data set (3D MPRAGE; 1 $\times$ 1 $\times$ 1 mm^3 resolution) was collected in the same session before the functional scans. All four subjects participated in two fMRI sessions for the retinotopic mapping experiment and the main experiment, respectively.

MRI r r r

The anatomical volume for each subject in the retinotopic mapping session was transformed into the AC–PC space. The cortical surface was 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 (0.015 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

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 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.

A r r

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.

R r

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] 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]