

Classification: Biological Sciences: Neurobiology

# Shape perception reduces activity in human primary visual cortex

Scott O. Murray<sup>\*</sup>, Daniel Kersten<sup>†</sup>, Bruno A. Olshausen<sup>\*‡</sup>, Paul Schrater<sup>†§</sup>, and David L. Woods<sup>¶</sup>

<sup>\*</sup>Center for Neuroscience, <sup>†</sup>Department of Psychology, <sup>¶</sup>Department of Neurology, UC Davis, Davis, CA 95616

<sup>‡</sup>Department of Psychology, <sup>§</sup>Computer Sciences and Engineering, University of Minnesota, Minneapolis, MN 55455

<sup>¶</sup>Neurology Service (127E), VA-NCHCS, 150 Muir Rd., Martinez, CA 94553

Corresponding author:  
Scott Murray  
Center for Neuroscience  
UC Davis  
Davis, CA 95616  
530-757-8789  
530-757-8749  
somurray@ucdavis.edu

Text Pages: 17  
Figures: 3  
Supporting Figures: 4  
Supporting Videos: 2

	Character count
Total text (abst, text, legends, ref's)	= 33793
Fig. 1 (2-column 13.4 cm x 360)	= 4824
Fig. 2 (2-column 11.25 cm x 360)	= 4050
Fig. 3 (2-column 10 cm x 360)	= 3600
3 2-column figures(3*240)	= 720
<b>TOTAL</b>	<b>46987</b>

Abbreviations: fMRI, functional MRI; V1, primary visual cortex; MT+, motion-sensitive visual area; LOC, lateral occipital complex; SFM, structure-from-motion

## **ABSTRACT**

Visual perception involves the grouping of individual elements into coherent patterns that reduce the descriptive complexity of a visual scene. The physiological basis of this perceptual simplification remains poorly understood. We used functional MRI to measure activity in a higher object processing area, the lateral occipital complex (LOC), and in primary visual cortex (V1) in response to visual elements that were either grouped into objects or randomly arranged. We observed significant activity increases in LOC and concurrent reductions of activity in V1 when elements formed coherent shapes, suggesting that activity in early visual areas is reduced as a result of grouping processes performed in higher areas. These findings are consistent with predictive coding models of vision that postulate that inferences of high-level areas are subtracted from incoming sensory information in lower areas through cortical feedback.

## **INTRODUCTION**

One of the extraordinary capabilities of the human visual system is its ability to rapidly group elements in a complex visual scene, a process that can greatly simplify the description of an image. For example, a collection of parallel lines can be described as a single texture pattern without specifying the location, length, and orientation of each element within the pattern. Such grouping processes are reflected in the activities of neurons at various stages of the visual system. For example, the response of a neuron in V1 to a single visual element can be suppressed if the element in its receptive field shares the same orientation as surrounding elements or enhanced if orientations differ (1). These pattern context effects in V1 are thought to be mediated by both local connections (2) and by interactions with higher areas (3).

In natural scenes, elements are often grouped when they are perceived as belonging to the same object. This case is particularly interesting from a physiological perspective because object

shape is a feature that is represented only in higher stages of the visual system, so any influence of perceived shape on lower areas would require feedback processes. Although feedback is generally thought of as a process where activity in lower areas is enhanced by activity occurring in higher areas, recent work on probabilistic models has pointed to the importance of a phenomenon termed "explaining away" – a competition that occurs between alternative hypotheses when attempting to infer the probable cause of an event (4). When applied to models of visual perception, perceptual hypotheses are thought to compete via feedback connections from higher visual areas projecting their predictions about the stimulus to lower stages, where they are then subtracted from incoming data. According to such predictive coding models, the activity of neurons in lower stages will decrease when neurons in higher stages can "explain" a visual stimulus (5, 6). These models can be contrasted with traditional feature-detection models, which posit that visual features are extracted in a largely sequential, feed-forward manner with little influence from higher areas upon lower areas (e.g., 7, 8).

Recent neuroimaging studies have shown that the lateral occipital complex (LOC) is a higher visual area critical for object shape perception. The LOC was first identified by its increased response to images of objects versus scrambled versions of the same images and textures (9, 10). More recent studies have shown that this area increases in activity whenever individual features are grouped into an object or a coherent scene (11). Thus, the LOC may subserve high-level grouping of low-level image features. In the present study, we examined the effect of perceived shape on activity in V1 and in LOC in a series of fMRI experiments where visual elements were either perceived as coherent shapes or as random elements. We observed reduced activity in V1 and increased activity in LOC when elements were grouped into coherent

shapes, consistent with the view that higher visual areas “explain away” activity in lower areas through feedback processes.

## **MATERIALS AND METHODS**

**Experiment 1.** Drawings were presented of (a) random lines, (b) lines that formed two-dimensional (2D) shapes and (c) lines that formed three-dimensional (3D) shapes (Fig. 1A). The stimuli were white lines on a black background. The 2D shapes were generated by randomly selecting 4-7 vertices at a minimum distance from fixation and connecting the vertices. The random lines were created by breaking the 2D shapes at their intersections and shifting the lines in the display. The 3D shapes were the same as the 2D shapes with the addition of small extensions that added perceived depth. The lines in all three conditions were equated for their retinotopic distribution, with each stimulus condition having equivalent mean distance and variance from fixation (random: mean = 4.05 degrees from fixation, SD = .93 degrees; 2D: mean = 4.06, SD = .94; 3D: mean = 3.94, SD = .97). (Supporting Fig. 1).

There were 40 different shapes in each condition. Each shape was presented for 750 ms in blocks lasting 30 s. Subjects viewed each condition a total of 8 times over two imaging runs. To independently define the area of V1 processing the line drawings, a flickering (8 Hz) counterphase checkerboard ring with the same mean eccentricity as the line stimuli was presented 7 times in 20 s blocks alternating with a gray screen. The area we designated as LOC was defined as active voxels in a comparison between the 3D lines and random lines condition, and located immediately posterior to MT+.

The number of line segments, line orientation, retinotopic distribution, and overall luminance were equated across stimulus classes, nevertheless, two minor differences in stimulus properties remained: there were line terminations in the random line stimuli and more parallel

lines in the 3D shapes. To evaluate the possible contributions of these stimulus differences to V1 activations, a control experiment was performed. Portions of the line segments in the 3D shapes were eliminated (introducing line terminations) but in two different ways: (1) in the non-shape condition, the corners were deleted and the remaining line segments were shifted and/or rotated slightly (usually less than 15 degrees) to remove perceived shape, and (2) in the shape condition only the middle portions of lines were deleted introducing line terminations without significantly disrupting shape perception (Supporting Fig. 2). As in the initial experiment, the shape and non-shape stimuli were controlled to have the same mean distance and variance from fixation.

**Experiment 2.** A second experiment was performed using structure-from-motion (SFM) stimuli. Random-dot displays were presented under three conditions: (a) stationary dots, (b) projections of random-dot patterns onto moving three-dimensional geometric shapes (SFM) and (c) velocity-scrambled moving dots (Fig. 2A).

The stimuli contained 450 dots and subtended 10 degrees. The dots in the structure-from-motion (SFM) stimuli were projected onto rigid geometric shapes including cube, cylinder and ‘house-shaped’ figures. Dots were randomly selected from a uniform distribution on the object surface and kept fixed relative to the rotating object surface and orthographically projected onto the image plane. The dots were rotated on a randomly chosen 3D axis for 40 degrees in 1.5-degree increments. Each stimulus presentation lasted 890 ms followed by 110 ms blank-screen delay before the next stimulus presentation. The velocity-scrambled stimuli were created by using the same starting positions as the SFM stimuli, then randomly assigning each dot’s SFM velocity (speed and direction) to that of another dot. Thus, the velocity-scrambled stimuli had identical velocities as the SFM stimuli but lacked any perceived 3D structure. The stationary dot condition presented randomly chosen frames from the SFM stimuli.

As with the line drawing stimuli, conditions were presented in 30 s blocks and subjects viewed each condition a total of 8 times. In addition to V1 and LOC, responses of motion-sensitive area MT+ were examined. To independently define V1 and MT+ a separate scan was performed that alternated a stationary random dot field (450 dots) with a randomly moving dot field 6 times in 20 s blocks presented in the same locations as the SFM stimuli.

**Experiment 3.** A final experiment was performed using a constant stimulus that formed a changing bi-stable 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 (Fig. 3A). In two separate conditions, either the bars (“bars move”) or the diamond moved (“diamond moves”) back and forth in the horizontal direction. Regardless of condition, the stimulus was perceived as either a single diamond behind occluders (“diamond”), or as separately moving line segments (non-diamond) (12-14 and supporting videos 1 and 2). The two percepts alternated after stable intervals of several to tens of seconds and subjects indicated their perceptual state with a button press. The stimulus subtended 7 degrees visual angle. The diamond (and bars, in the “bars move” condition) moved at a constant horizontal speed of 0.8 deg/sec per second and reversed direction every 1.3 seconds. A total of four 300 s imaging scans were performed – two for each condition.

**Subjects and Procedure.** In all experiments stimuli were presented using an LCD projector onto a rear projection screen at the feet of the subjects. Subjects viewed the screen with an angled mirror positioned on the head-coil. In experiments 1 and 2, data were collected from a total of six subjects. Three of the six subjects were scanned in a control experiment using the stimuli in which portions of the line segments in the 3D shapes were eliminated (Supporting Fig. 2). In experiments 1 and 2 subjects were asked to maintain fixation and were required to detect a

brief, 100 ms, 40% dimming of the entire stimulus. The task was designed to equally engage attention in all conditions. An additional two subjects participated in experiment 3. In experiment 3, subjects were allowed to freely view the stimulus and eye movement patterns were measured in response to the same visual stimuli outside of the magnet.

**MRI Scans and Analysis.** For experiments 1 and 2 subjects were scanned on a 1.5 T Marconi magnet, using a standard head coil. A spin-echo sequence (FOV = 240 mm, matrix = 128 x 128, TR = 2000 ms, TE = 40 ms, flip angle = 90) was used to visualize the BOLD (blood oxygen level-dependent) contrast changes that accompany brain activity. Images were acquired in 20 contiguous 5-mm horizontal slices. SPM99 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)) was used to generate statistical maps using a “boxcar” response model smoothed with a hemodynamic response function. The functional data were motion corrected and spatially smoothed with a 4mm Gaussian window. For the average percent change data presented (Fig.’s 1C & 2C), average values of each region-of-interest for each 30 s epoch were included in paired two-tailed t-tests. Percent change was calculated by first linearly detrending each voxel’s timecourse, then subtracting and dividing by the mean across all conditions. Statistical maps (p-values at each voxel location) were projected onto ‘inflated’ and flattened cortices using FreeSurfer (15, 16) derived from a high-resolution anatomical scan from each subject (FOV = 240 mm, matrix = 256 x 212 x 256, TE = 4.47 ms, TR = 15 ms, flip angle = 35°). The boundaries between the retinotopic areas were defined with counterphase flickering narrow wedges at the horizontal and vertical meridian. In all of the analyses, an ROI was defined as a contiguous region of voxels surpassing a minimum statistical threshold of  $p < 10^{-4}$ .

In experiment 3 subjects were scanned using a Siemens Vision 1.5 T scanner using an EPI sequence, TR/TE = 1000/60 ms, 64x64 matrix at FOV = 22x22 cm. Seven one centimeter

thick oblique coronal slices were scanned, covering a region centered to ensure coverage of primary visual cortex. The restricted number of slices allowed greater temporal resolution, however, higher visual cortical areas were not imaged.

In experiment 3 task related BOLD responses were detected using correlation analysis. The time series for each voxel was first filtered to remove drift and high frequency noise and normalized to the mean of the time series. We represented the button responses as a series of delta functions indicating the subject's report of a transition from diamond to non-diamond percept. This was then convolved with an ideal BOLD impulse response function to generate a template to identify pixels whose activity correlated with the template:  $h(t) = t^\alpha e^{-t/\beta}$ ,  $\alpha = 8.6$ ,  $\beta = 0.55$  (17). V1 voxels were defined as those with a significant ( $p < .001$ ) cross-correlation value and located in the calcarine sulcus. The position of V1 was confirmed in one of the two subjects using standard retinotopic mapping stimuli (18, 19).

In order to obtain a quantitative measure of how well V1 activity predicts the modeled response transient, a simple decision rule was used to determine a Receiver Operating Characteristic (ROC) curve for each scan. The rule tested whether a fixed number of contiguous time points of BOLD activity were all above a criterion level. The number of contiguous measurements was pre-determined and equal to the width at the half-height of the simulated response from the perceptual measurements. A "hit" means the decision rule correctly predicted the perceptual response, and "false alarm" means that it did not. An ROC curve shows hit versus false alarm rates for a range of criterion levels. The z-score of the area under the ROC curve provides a standard measure of sensitivity called  $d'$  (20), where chance performance corresponds to a diagonal ROC curve for which  $d' = 0$ . For each subject scan 3 was used to select a subset of



V1 voxels that were good predictors of the perceptual transient and then compared the time series of activity in these locations to the reported percept transitions on the three other scans.

## RESULTS

**Experiment 1.** Activity in the lateral occipital complex (LOC) increased in response to the shape stimuli, with the 2D shapes activating the LOC significantly more than the random lines, and the 3D shapes activating LOC more than the 2D shapes consistent with previous findings (9,10, 21-23). By contrast, activity in V1 showed an opposite pattern – significant reductions in response to the 2D shapes compared to the random lines and significantly less activity for 3D than 2D shapes (Fig. 1). The reduction of activity was localized on the flattened visual cortices of 6 subjects (12 hemispheres) using retinotopic-mapping techniques (18, 19). (Supporting Fig. 3 shows the relative location of the ROIs and the location of the “cuts” to flatten the cortex). A region of interest was independently defined in V1 using a flickering counterphase ring stimulus with the same mean eccentricity as the line drawing stimuli. This region showed significantly reduced activity during object perception (Fig. 1B). The area of V1 representing the fovea (where no line stimuli occurred) did not show processing-related alterations in activity. Possibly because the stimuli used to map the horizontal and vertical meridian were very narrow wedges, we could not reliably differentiate intermediate retinotopic areas (V2, V3, V4, etc.) to perform an ROI analysis. However, inspecting the flattened statistical maps revealed little reliable activity differences in intermediate retinotopic areas. Several subjects did show reduced activity for 3D shapes in what is likely V2 and V3 (e.g., see Fig. 1B), however, only V1 showed reduced activity in all subjects. There was no significant difference in the luminance-change detection task, designed to equally engage attention in all conditions (Hit rate: Random = 90%, 2D = 89%, 3D = 91%).

Although the number of line segments, line orientation, retinotopic distribution, and overall luminance were equated across stimulus classes, there were line terminations in the random line stimuli and more parallel lines in 3D shapes. To evaluate the possible contributions of these stimulus differences to V1 activations, a control experiment was performed with three subjects using the stimuli with the eliminated line segments (Supporting Fig. 2). All three subjects showed significant ( $p < .01$ ) reductions of activity in V1 and significant ( $p < .01$ ) increases in LOC in the shape versus non-shape control comparisons.

**Experiment 2.** To evaluate the generality of the findings a second experiment was performed using 3D structure-from-motion (SFM) shapes. Activity was compared to velocity-scrambled and stationary stimuli. The three regions-of-interest (V1, MT+, and LOC) showed significant increases in activity in response to motion stimuli (velocity-scrambled or SFM) compared to the stationary stimuli, with MT+ showing the largest signal increase. In response to the SFM shapes, activity increased significantly in LOC in comparison with the velocity-scrambled control, consistent with other studies showing that LOC integrates multiple shape cues (21, 22). By contrast, both V1 and MT+ showed significant reductions in activity during object perception (Fig. 2C). As before, the retinotopic position of the stimuli was defined in independent scans, and the reduction in V1 activity was restricted to the retinotopic position of the stimuli (Fig. 2B). There was a significant difference in the luminance-change detection task between the velocity-scrambled condition compared to the SFM and stationary conditions (Hit rate: velocity-scrambled = 86%, SFM = 94%, stationary-dots = 96%).

**Experiment 3.** Though many low-level stimulus features were controlled in the first two experiments, it is still possible that some stimulus differences could have contributed to the reductions in V1 activity. For example, in experiment 2, the SFM stimuli had higher motion

coherence (percentage of dots locally moving in the same direction) as well as higher opponent motion (opposite local motion directions). While it is not clear that these differences would cause activity decreases in V1, it was nevertheless important to use a stimulus with non-changing features. A final experiment was therefore performed using a constant stimulus that formed a changing bi-stable percept with either grouped or ungrouped line segments. Experiment 3 measured V1 activity while subjects viewed a bi-stable stimulus that was either perceived as a moving rigid diamond behind occluders or as individual moving line segments. Importantly only the percept, and not the stimulus features, changed during this experiment.

The percept changed, on average, every 8 seconds (SD = 3.5) for subject 1 and every 7.7 seconds (SD=3.3) for subject 2. The stimulus was perceived as a coherent diamond for approximately 50 percent of the scan. Significant changes in activity were observed in V1 in accordance with the percepts of the subjects. Consistent with the results from experiments 1 and 2, activity in V1 was significantly reduced when a coherently moving diamond was perceived, and significantly increased when visual elements were not perceived as part of a shape. The restricted number of slices acquired in this experiment allowed greater temporal resolution, however, higher visual cortical areas were not imaged. No significant difference in the pattern of eye movements was observed between the coherent “diamond” percept and the “no diamond” percepts (Supporting Fig. 4).

To compare cortical activity with the perceptual response, the perceptual transient corresponding to “no diamond” was convolved with a model transient BOLD impulse response function (17). Fig. 3D shows the time course of V1 activity (thin red lines) superimposed on the smoothed perceptual transient (thick gray lines). It can be seen by inspection that V1 activity is correlated with the perceptual response – when subjects’ perception changed from a coherent

shape to individual line segments V1 activity increased. Fig. 3E shows the results from the ROC analysis. For each scan,  $d'$  is greater than zero. The average  $d'$  and standard errors (excluding scan 3 which was used to select voxels), for the two subjects were:  $d'=1.9$ , s.e. = .82 and  $d'=1.15$ , s.e. = .26. Thus, decreased activity in V1 is an accurate predictor of perceived grouping on the time scale of the perceptual behavior.

## **DISCUSSION**

Our results show that grouping individual elements into shapes increased activity in higher object-processing areas and concurrently reduced activity in primary visual cortex. While shapes and non-shapes differed in some low-level stimulus features in experiments 1 and 2, the control experiment showed that minor stimulus differences could not plausibly account for the findings. Moreover, the results of Experiment 3 clearly demonstrate that the effects in V1 are not due to stimulus differences but to perceptual interpretations of the stimuli.

Our results help to explain differences in V1 activity obtained incidentally in previous experiments. For example, a recent study examining the effects of progressive image scrambling on LOC activity also showed significant changes in V1 activity (24). Specifically, it was observed that progressive decreases in LOC activity were accompanied by progressive increases in V1 activity, as images were changed from ordered to scrambled. The interpretation was that V1 has greater sensitivity to local image features. However, an alternative explanation, consistent with the results of our current study, is that the changes in V1 reflect the level of perceived grouping – as higher visual areas are able to group local image features into coherent objects the need for lower areas to signal their presence is reduced. Our results from experiment 2 are also consistent with a previous SFM study that showed reduced activity in V1 for dots that

were perceived as belonging to a spherical surface compared to random dot motion (25).

However, unlike our study no reliable differences in MT+ activity were observed.

One possible alternative explanation for our finding of reduced V1 activity is that attention is enhanced in V1 for non-grouped stimulus elements. Any such effect would appear to be automatic since in experiments 1 and 2 subjects' were engaged in a demanding luminance detection task in all conditions. Therefore, any attentional differences would have to be caused by an involuntary mechanism, whereas previous studies showing attentional modulations required voluntary switches in attention between different features or spatial locations (e.g., 26, 27). Further, there is no evidence that V1 is specialized for processing ungrouped versus grouped line segments so there is no reason to believe it would be specifically enhanced by attention. Also, to explain the results from experiment 2 as an attentional confound, one would have to argue that V1 is specialized for processing random motion over structured motion, for which there is no evidence. And finally, the lack of any difference in eye movements in experiment 3 also suggests there are no differences in the focus of overt spatial attention between grouped and ungrouped elements.

The observed decreases in V1 activity in the current experiments may at first appear inconsistent with the results of previous contextual modulation studies. For example, in previous studies contextual effects signaling the presence of figure-from-ground resulted in *increased* V1 activity (28). However, in these previous studies figure perception involved the ungrouping of a line or texture patch from the background or surrounding elements (1,28). These observations have led some to suggest that context effects – in particular increases in activity in V1 – are an index of perceptual saliency (29). When line segments are grouped with the background or

combined into a pattern (i.e., have reduced saliency), V1 activity is reduced (1) as in the current study.

Feedback projections are a prominent anatomical feature of the primate visual system (30) and recent evidence suggests they play a critical role in visual perception (31). Feedback from higher areas appears to be the primary contributor to the context effects observed in the current experiments. In the present study, the context – global shape – is a feature that can only be represented in higher stages of the visual system. Hence, our results are consistent with other evidence pointing to the importance of feedback as a mechanism for context effects in lower areas (3).

Feedback is frequently thought of as a process where activity in lower areas is enhanced by activity occurring in higher areas. However, there are reasons to think that feedback may also reduce activity in lower areas. First, recent computational models of predictive coding (5, 6) posit a subtractive comparison between the predictions of hypotheses generated in higher areas and incoming data in lower areas, with the residual from the subtraction operation passed along as neural activity. Thus, reduced activity would occur when the predictions of higher-level areas match incoming sensory information. Second, higher areas may serve to disambiguate activity patterns in lower areas – reducing irrelevant activity and enhancing activity appropriate for the percept. The net effect of such a process may be reduced activity. Both accounts suggest that activity will be reduced proportional to the degree to which higher areas can account for the stimulus. In fact, the difference in V1 activity in experiment 1 between 2D and 3D shapes suggest that it is not grouping per se that leads to reduced activity, but the degree to which neurons in higher visual areas are tuned for the stimulus. Based on the results of the current study and others (23), it appears that LOC neurons are tuned for processing three-dimensional

surface structure as reflected by both the increase observed in LOC and the decrease observed in V1 for 3D versus 2D shapes.

Reduced activity may afford several behavioral advantages. For example, by selectively decreasing activity in V1 for grouped elements, ungrouped or novel elements may be more readily detected. Importantly, our results demonstrate that neuronal activity, even in V1, does not simply represent the signaling of features in a visual scene but is strongly influenced by high-level perceptions of object shape.

We thank K. Ahlo, C. Anderson, S. David, L. Disbrow, J. Gallant, J. Johnson, J. Mazer, E. Mercado, E. Wojciulik, and E. Yund for discussion and comments on the manuscript and K. Ugurbil, L. Shen, E. Yacoub and X. Hu and F. Fang for their contributions. Portions of this work were reported earlier at *Human Brain Mapping*, *ARVO*, and *Soc. Neurosci.* conferences (Shen, Kersten, & Ugurbil, 1999; Kersten, Shen, Ugurbil, Schrater, 1999; Murray, Olshausen, Woods, 2001). Supported by NSF SBR-9631682, NIH P41 RR08079, pre-doctoral NRSA MH-12791 (S.O.M.), NSF SBR-9631682 (D.K.), NIH MH-57921 (B.A.O.), NIH MH-41544 and VA Research Service (D.L.W.).

1. Knierim, J.J. & Van Essen, D.C. (1992). *J. Neurophysiol.* 67, 164-181.
2. Das, A. & Gilbert, C.D. (1999). *Nature* 399, 655-661.
3. Hupe J.M., James A.C., Payne B.R., Lomber S.G., Girard P., & Bullier J. (1998). *Nature* 394, 784-787.
4. Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers, San Francisco).
5. Mumford, D. (1992). *Biol. Cybern.* 66, 241-251.
6. Rao R.P. & Ballard D.H. (1999). *Nat. Neurosci.* 2, 79-87.
7. Fukushima, K. (1980). *Biol. Cybern.* 36, 193-202.
8. Mel, B. (1997). *Neural Comp.*, 9, 777-804.
9. Malach, R., Reppas, J.B., Benson, R.R., Kwong, K.K., Jiang, H., Kennedy, W.A., Ledden, P.J., Brady, T.J., Rosen, B.R., & Tootell, R.B. (1995). *Proc. Natl. Acad. Sci. USA* 92, 8135-8139.
10. Kanwisher, N., Chun, M.M., McDermott, J., & Ledden, P.J. (1996). *Brain Res. Cognitive Brain Res.* 5, 55-67.
11. Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). *Vision Res.* 41, 1409-1422.
12. Shiffrar, M., & Pavel, M. (1991). *J. Exp. Psych.: Human Percep.Perf.* 17, 749-761.
13. Lorenceau, J., & Shiffrar, M. (1992). *Vision Research*, 32, 263-373.
14. McDermott, J., Weiss, Y., & Adelson, E. H. (2001). *Perception*, 30, 905-23.
15. Dale, A.M., Fischl, B., & Sereno, M.I. (1999). *NeuroImage* 9, 179-194.
16. Fischl, B., M.I. Sereno, M.I., & Dale, A.M. (1999). *NeuroImage* 9, 195-207.
17. Friston, K. J., Jezzard, P., & Turner, R. (1994). *Human Brain Mapping*, 1, 153-171.



18. Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., & Tootell, R.B. (1995). *Science* 268, 889-893.
19. Engel, S.A., Glover, G.H., & Wandell, B.A. (1997). *Cereb. Cortex* 2, 181-192.
20. Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. (Wiley, New York).
21. Grill-Spector, K., Kushnir, T., Edelman, S., Itzhak, Y., & Malach, R. (1998). *Neuron* 21, 191-202.
22. Kourtzi, Z. & Kanwisher, N. (2000). *J. Neurosci.* 20, 3310-3318.
23. Moore, C. & Engel, S.A. (2001). *Neuron* 29, 277-286.
24. Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). *Cereb. Cortex* 11, 287-297.
25. Paradis A.L., Cornilleau-Peres V., Droulez J., Van de Moortele P.F., Lobel E., Berthoz A., Le Bihan D., & Poline J.B. (2000). *Cereb. Cortex* 10, 772-783.
26. Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., Petersen, S.E. (1990). *Science* 248, 1556-1559.
27. Gandhi, S.P., Heeger, D.J., Boynton, G.M. (1999). *Proc. Natl. Acad. Sci. U.S.A.* 1999 6, 3314-9
28. Lamme, V.A.F. (1995). *J. Neurosci.* 15, 1605 1615.
29. Lamme, V.A.F. & Spekreijse, H. (2000). In: *The New Cognitive Neurosciences* (MIT Press, Cambridge, MA).
30. Felleman, D.J. & Van Essen, D.C. (1991). *Cereb. Cortex* 1, 1-47.
31. Pascual-Leone, A. & Walsh, V. (2001). *Science* 292, 510-512.

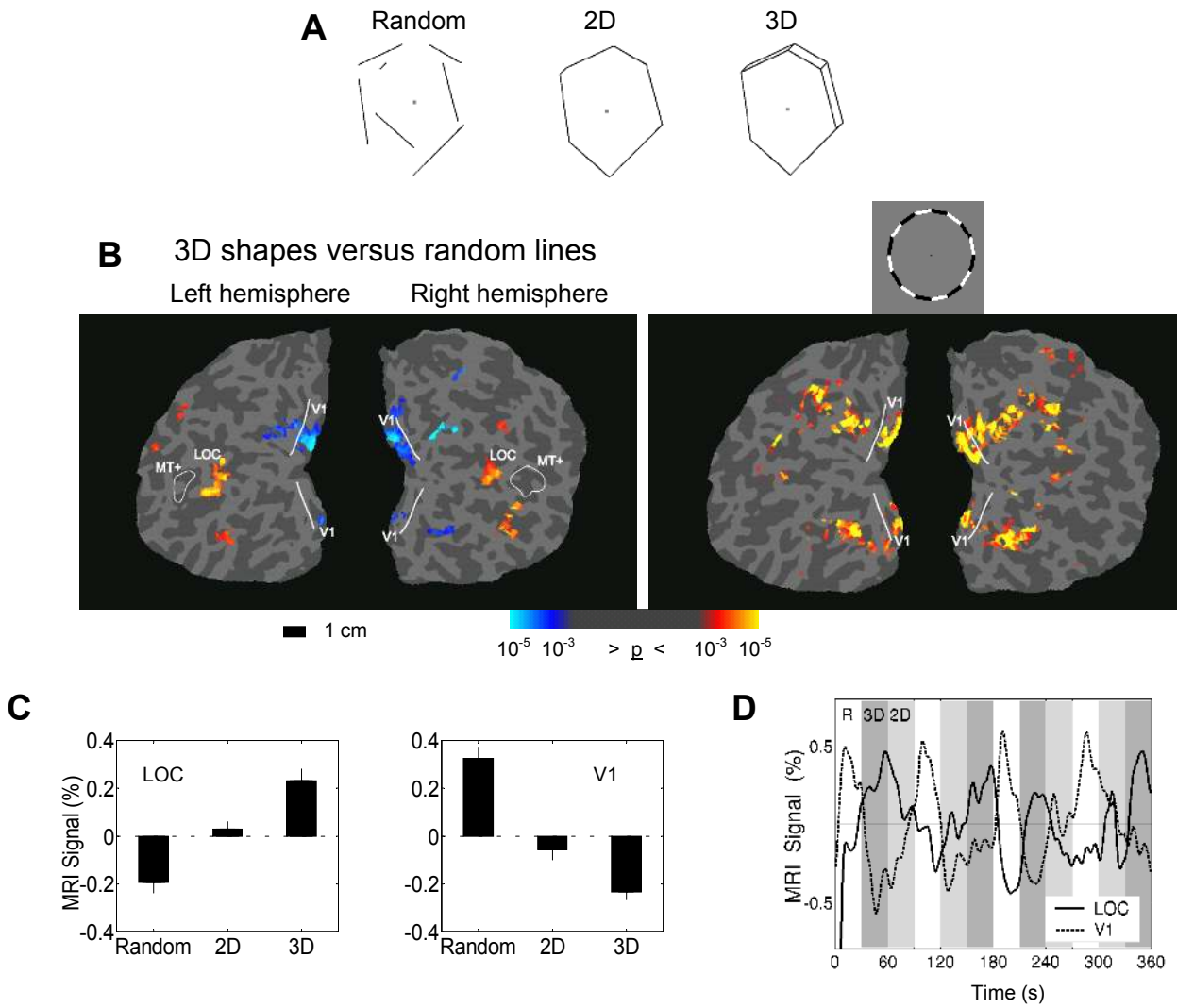


FIGURE 1

Fig. 1. Experiment 1. (A) Examples of the three different stimulus conditions. (B) Left, Areas of increased (red/yellow) and decreased (blue) activity comparing 3D figures to random lines for a representative subject on a flattened representation of occipital cortex. Right, A flickering ring stimulus matching the mean eccentricity of the line drawings was used to independently locate the portion of V1 where the line drawing stimuli occurred. The reduced activity for the 3D figures in V1 is restricted to the cortical area representing the stimuli. The solid line indicates the representation of the vertical meridian, marking the boundary of V1. The location of MT+ defined by random dot motion is included as a reference. Supporting Fig. 3 shows the relative location of the ROIs and the location of the ‘cuts’ to flatten the cortex. (C) The average percent signal change from the mean for the three conditions averaged over 6 subjects. All pair-wise comparisons are significant,  $p < .001$ . Error bars are SEM. (D). The average time-course of the MRI signal in LOC (solid line) and V1 (dashed line). Percent signal change is from the mean activation across all three conditions. Periods corresponding to the three conditions, random (R – white), three-dimensional (3D – dark gray) and two-dimensional (2D – light gray) are shown. The dissociation between LOC and V1 is clearly evident – as activity increases in LOC, activity in V1 declines.

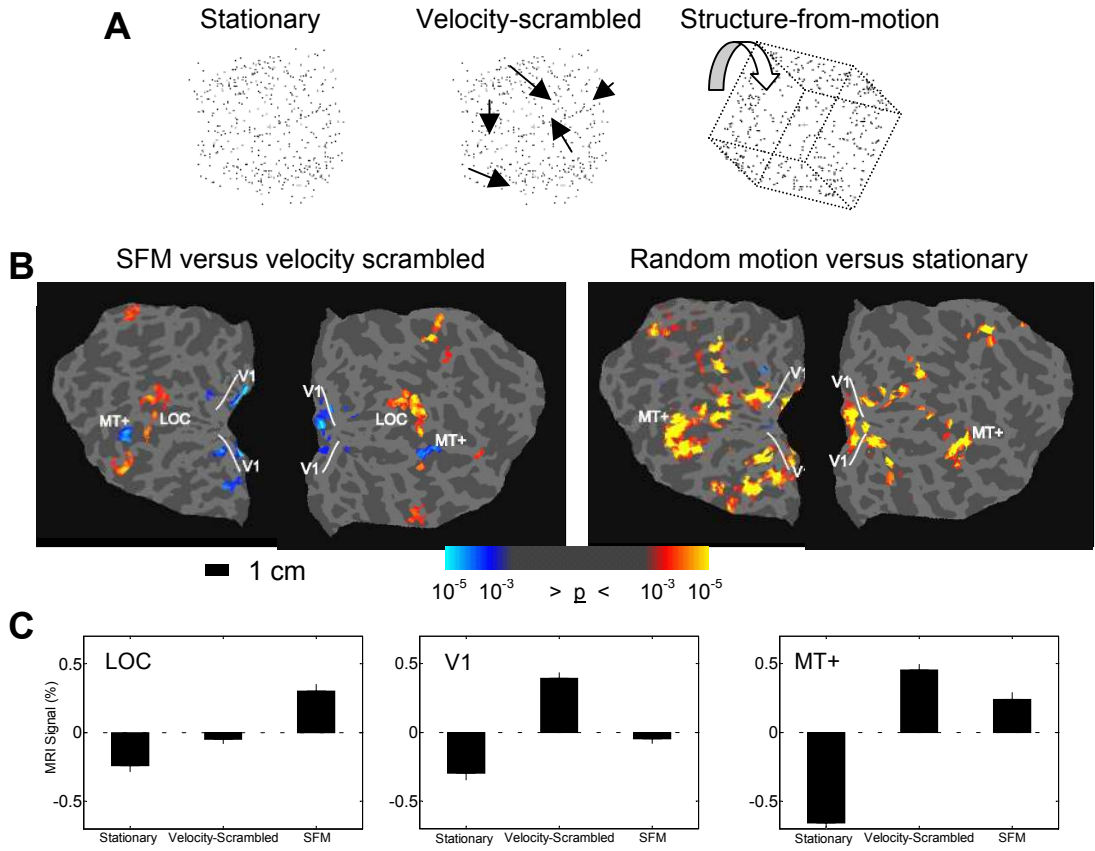


FIGURE 2

Fig. 2. Experiment 2. (A) The three types of random dot stimuli used in Experiment 2. (B) Left, Areas of increased (red/yellow) and decreased (blue) activity comparing SFM and the velocity-scrambled control stimuli. Right, V1 and MT+ were defined in a separate scan using a randomly moving dot field compared to stationary dots. The area of decreased activity in V1 to the SFM stimuli was restricted to the region independently defined by the random motion stimulus. (C) The percent signal change averaged over six subjects in LOC, V1, and MT+. Activity increased in all three areas in motion conditions relative to the stationary condition. The critical comparison in this experiment is between the velocity-scrambled and SFM conditions. LOC activity increased to the SFM stimulus compared to the velocity-scrambled stimuli whereas V1 and MT+ showed significant activity reductions. All pair-wise comparisons are significant,  $p < .001$ . Error bars are SEM.

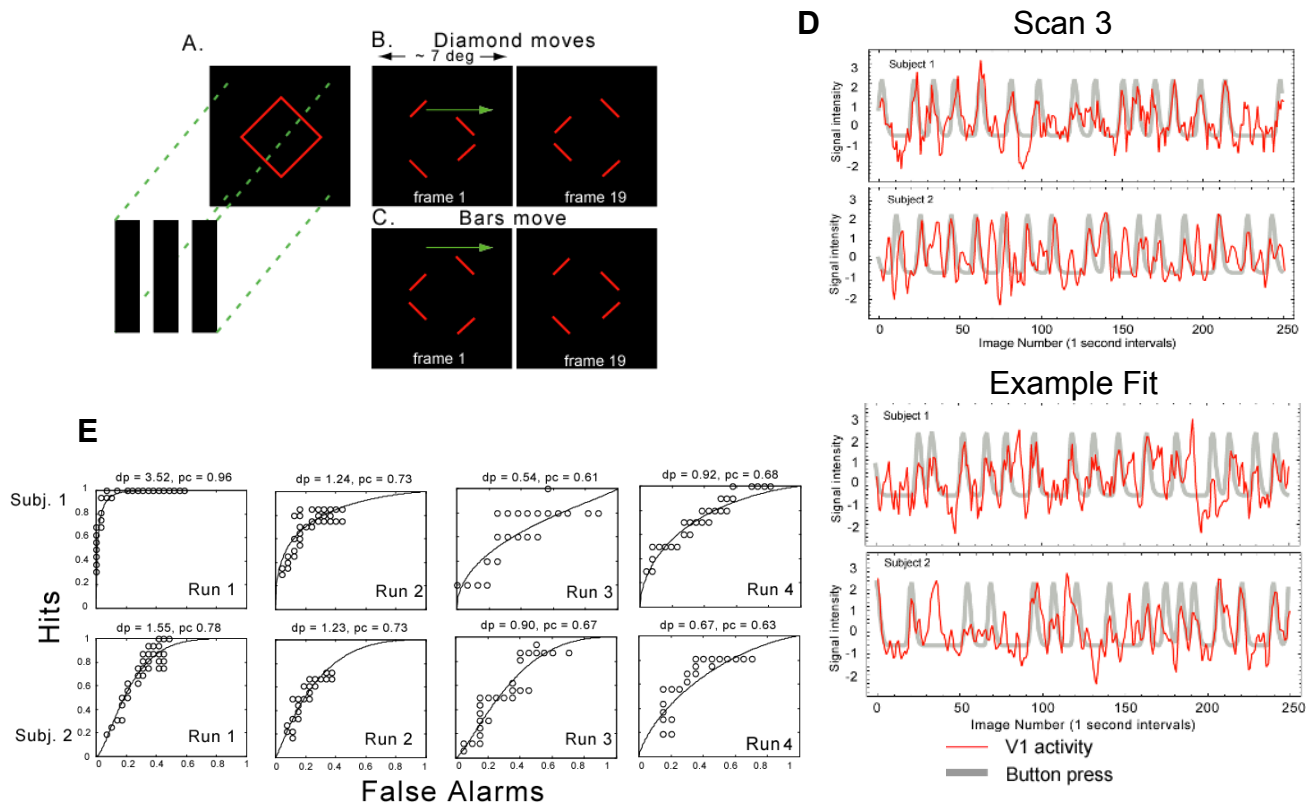


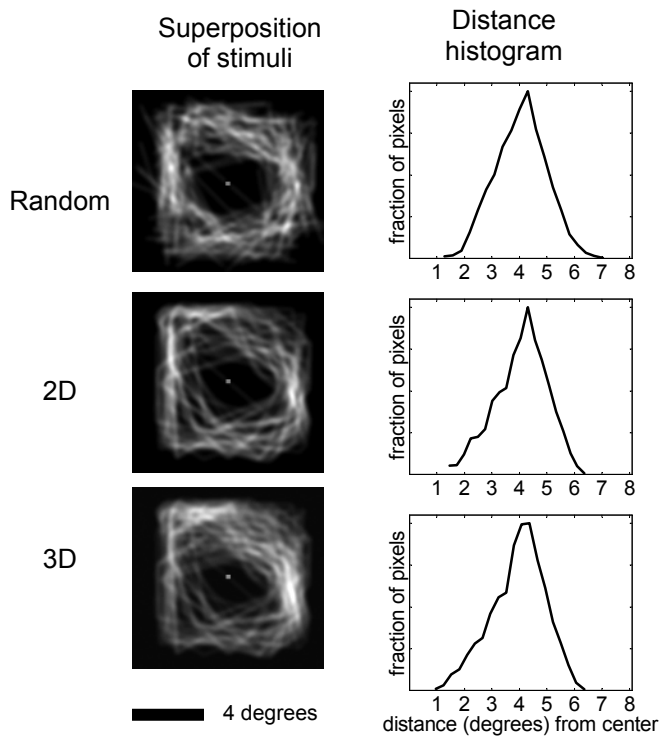
FIGURE 3

Fig. 3. Experiment 3. (A) A red diamond was covered by three black bars that hid the four vertices. There were two stimulus conditions in which either the red diamond moved, or the three occluding black bars moved horizontally back and forth, as shown in *B* (“diamond moves”) and *C* (“bars move”), respectively. The left and right columns in *B* and *C* show the first and last movie frame. The four remaining line segments could either be perceived as a rigid diamond moving horizontally or as individual line segments moving vertically. See supporting videos 1 and 2 for examples of the “diamond moves” stimulus. (D) Activity time course of voxels in V1 (thin red lines) and the superimposed simulated response (thick gray lines) for two subjects. Scan 3 for both subjects was used to choose significantly correlated voxels (left column). The right column shows the fit for one of the other three scans. (E) ROC curves quantifying the predictability of the perceptual transition from diamond to non-diamond from fluctuations in V1 activity. In all scans for both subjects, V1 activity was an accurate predictor of subjects’ perceptual state.

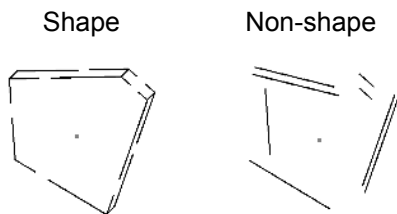
Supporting Video 1. The vertices of a horizontally translating diamond are covered with occluders. The diamond percept remains clear when the occluders are visible. The stimulus in experiment 3 used invisible occluders, an example of which is shown in Supporting Video 2.

Supporting Video 2. The occluders are now invisible and the diamond percept is more difficult to see. With familiarity, the stimulus is bistable switching from ungrouped line segments to a rigid diamond percept.

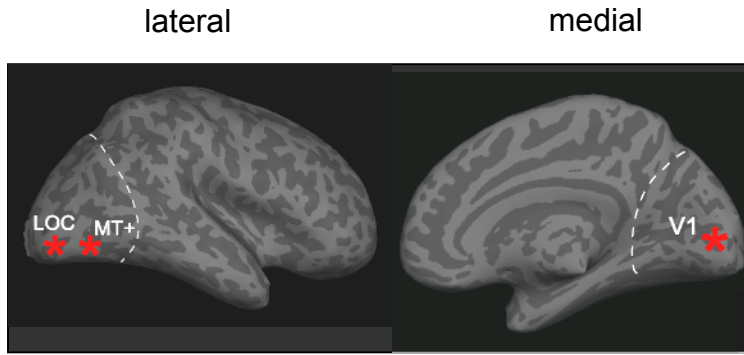




Supporting Figure 1. The lines for all figures in all conditions had similar eccentricities (mean  $4^\circ$ ), shown in the left column by superimposing stimuli in each condition. The three conditions were equated for the mean and variance of distance from fixation, shown in the histograms in the right column.

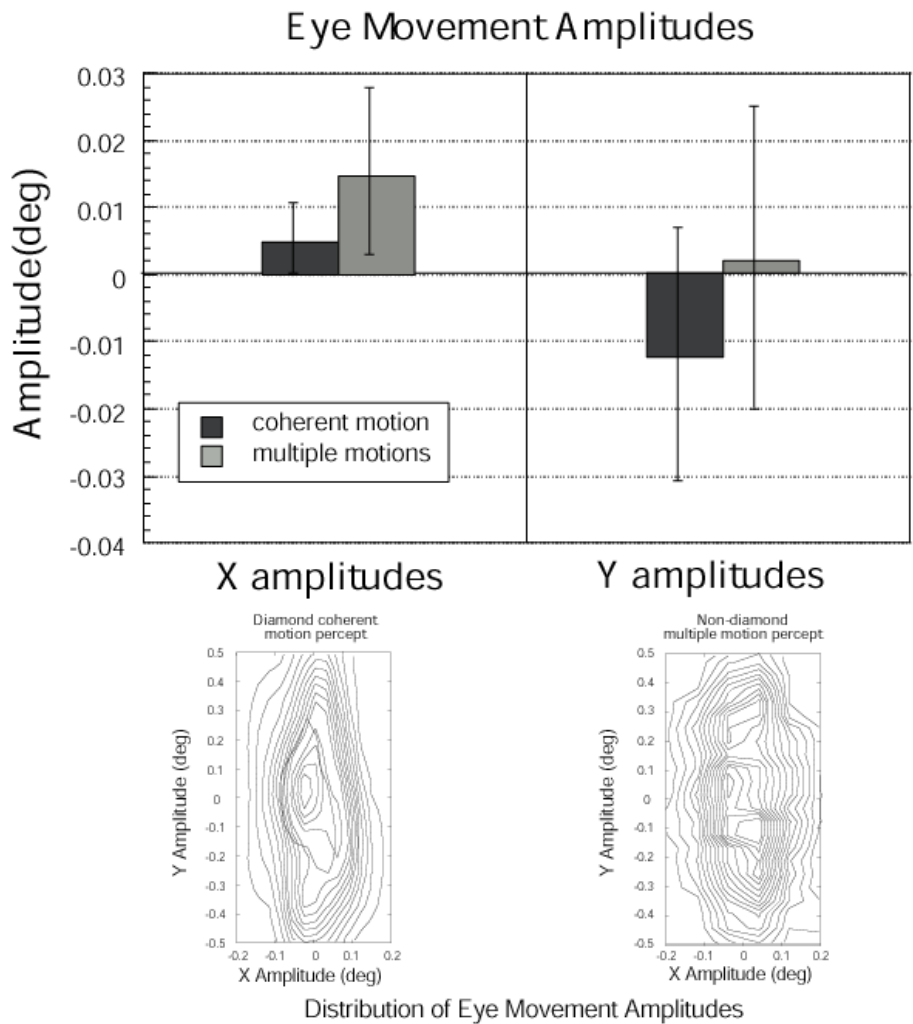


Supporting Figure 2. Examples of the non-shape and shape stimuli used in the control experiment. The stimuli were equated for number of line terminations and parallel lines.



Supporting Figure 3. Relative locations of V1, LOC, and MT+ shown on an inflated cortex.

Dashed line shows location of 'cuts' for the flattened representations.



Supporting Figure 4. Eye movement data were collected using a video-based eye movement tracker, while observers viewed the bistable stimulus and recorded their perceptual state via a key-press. Tracking data was low-pass filtered and saccadic eye movements were detected by filtering the data for step changes exceeding a peak velocity of 50deg/sec. The bar chart shows the mean of the saccade amplitudes in the horizontal (x) and vertical (y) direction that occurred during the two perceptual states for 1000 saccades, and error bars represent the standard deviation of the amplitudes. The contour plots represent Gaussian kernel density estimates of the

2-D saccade amplitude distribution. The figure demonstrates that there were no differences in eye movements during the two perceptual states.