Attention-Dependent Representation of a Size Illusion in Human V1

Fang Fang,[1] Huseyin Boyaci,[2] Daniel Kersten,[3] and Scott O. Murray[4],*

[1]Department of Psychology
and Key Laboratory of Machine Perception
    (Ministry of Education)
Peking University
Beijing 100871
People's Republic of China
[2]Department of Psychology
Bilkent University
Ankara 06800
Turkey
[3]Department of Psychology
University of Minnesota
Minneapolis, MN 55455
USA
[4]Department of Psychology
University of Washington
Seattle, WA 98105
USA

Summary

One of the most fundamental properties of human primary visual cortex (V1) is its retinotopic organization, which makes it an ideal candidate for encoding spatial properties, such as size, of objects. However, three-dimensional (3D) contextual information can lead to size illusions that are reflected in the spatial pattern of activity in V1 [1]. A critical question is how complex 3D contextual information can influence spatial activity patterns in V1. Here, we assessed whether changes in the spatial distribution of activity in V1 depend on the focus of attention, which would be suggestive of feedback of 3D contextual information from higher visual areas. We presented two 3D rings at close and far apparent depths in a 3D scene. When subjects fixated its center, the far ring appeared to be larger and occupy a more eccentric portion of the visual field, relative to the close ring. Using functional magnetic resonance imaging, we found that the spatial distribution of V1 activity induced by the far ring was also shifted toward a more eccentric representation of the visual field, whereas that induced by the close ring was shifted toward the foveal representation, consistent with their perceptual appearances. This effect was significantly reduced when the focus of spatial attention was narrowed with a demanding central fixation task. We reason that focusing attention on the fixation task resulted in reduced activity in—and therefore reduced feedback from—higher visual areas that process the 3D depth cues.

Results

Psychophysical Experiment

To assess the magnitude of the size illusion, we used a rendered three-dimensional (3D) scene of a hallway and walls to present two physically identical 3D rings: one was at a close ("front") apparent depth and the other was at a far ("back") apparent depth (Figure 1). The inner and outer radii of the ring were 2.37° and 3.00°, respectively. Subjects adjusted a 2D gray ring located outside the scene until it matched either the front 3D ring or the back 3D ring in angular (image) size. The inner and outer radii of the 2D ring could be adjusted independently. The results showed that, on average, the back ring appeared to be approximately 15% larger than the front ring. Expressed as degrees of visual angles, the inner and outer radii of the back ring were perceived to be 2.56° (standard error [SE] = 0.03°) and 3.21° (SE = 0.02°). In contrast, the inner and outer radii of the front ring were perceived to be 2.27° (SE = 0.03°) and 2.80° (SE = 0.04°).

fMRI Experiments
Stimulus
The stimuli used in the functional magnetic resonance imaging (fMRI) experiments were the same as the psychophysical experiment—rings at near and far apparent depths. The use of rings—as opposed to other shapes such as spheres or disks—offers several advantages for the fMRI measurements. First, the use of rings addresses a critical issue that was not addressed in our previous experiment [1]. Although our earlier study measured a greater distribution of activity in V1 to a perceptually larger sphere, this could result from either a positional shift in the cortical representation of the edges of the spheres or from nonlinearities present in the fMRI signal. Figure 2A illustrates these two possibilities. The top row of Figure 2A demonstrates how an expansion in the edge response to the perceptually larger back sphere, combined with a simple model of the fMRI point-spread function, could result in a spatial distribution of the fMRI signal similar to what we observed in our previous study. This characterizes how we interpreted our original findings—namely, that perceived size differences manifest in a change in the activity distribution.

The bottom row of Figure 2A offers a potential alternative explanation of the results and demonstrates how the exact same change in the measured response could result from an increase in the neural activity to the back sphere (with no change in the spatial distribution of the edge) in combination with a saturating nonlinearity in the fMRI response. Figure 2B shows how this issue is addressed by the use of rings rather than spheres. By using rings, we could compare the precise retinotopic positions of the activity peaks in V1. Specifically, we are testing whether the spatial distribution of V1 activation induced by the back ring is shifted toward a more eccentric representation of the visual field, whereas that induced by the front ring is shifted toward the foveal representation (top row, Figure 2B). Importantly, as the bottom row of Figure 2B illustrates, a simple saturating nonlinearity in the fMRI signal will not produce an activity shift in the peak responses.
Experimental Design
We used a block design to present the stimuli in the fMRI experiment. Specifically, subjects fixated a small green point at the center of the front 3D ring (i.e., the intersection of the spokes) for 20 s, and then the front 3D ring was counterphase flickered for 10 s. The green fixation point moved to the center

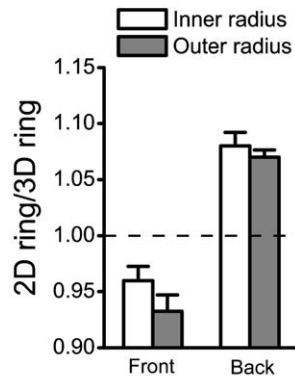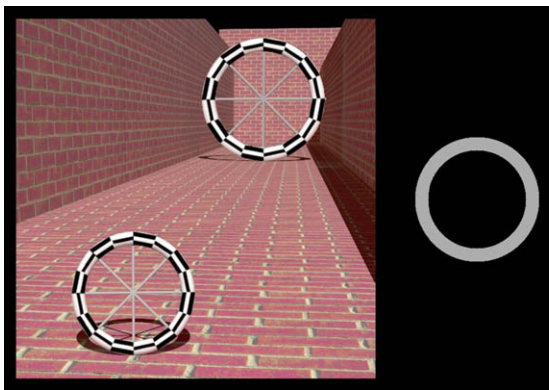*Correspondence: somurray@u.washington.edu

Figure 1. Psychophysical Experiment

(Left) For direct comparison of 2D and 3D size judgments, a 2D ring located outside the 3D scene was adjusted to match the front and back rings. The inner and outer radii of the 2D ring could be adjusted independently.
(Right) The behavioral effect was quantified by dividing the adjusted size of the 2D ring by the 3D ring size. The front and back rings in the 3D scene were judged to be smaller and larger, respectively, than an equivalent 2D ring.
Error bars represent SEM.

of the back 3D ring and the subjects were instructed to follow it and maintain fixation for 20 s, and then the back 3D ring was flickered for 10 s. Inclusion of the spokes allowed the fixation point to always remain directed on the object (3D ring plus spokes). Otherwise, one fixation would have directed attention to the floor and the other to the back wall. The cycle was repeated five times for each scan.

Attention Manipulation

We hypothesized that the previously measured [1] changes in the spatial distribution of activity in V1—which stem from a size illusion—are the result of feedback of 3D contextual information from higher visual areas. The primary cues to depth that gave rise to the size illusion included linear perspective, occlusion, and relative texture size. All of these cues presumably depend on the ability to integrate and compare information across a relatively large area of visual space and thus are likely to rely on the larger receptive fields of higher stages of the visual system. Further, there appear to be cortical regions in the human inferior temporal lobe that are specialized for processing 3D spatial configurations [2, 3].

Because areas in higher visual cortex are strongly modulated by attention [4, 5], one way to test the feedback model is to "silence" activity in areas that process 3D scene cues with an attention manipulation. For measurement of the effect of attention on the spatial distribution of activity in V1, each subject performed two tasks. In the attend-to-ring condition, attention was directed to the ring by asking subjects to detect a 250 ms flickering pause of the 3D ring, which occurred randomly with a mean intertrial interval of 5 s drawn from a uniform distribution with a range of 3.5–6.5 s. This task was designed to be relatively easy so that perceptual resources could be allocated to the entire scene. In the attend-to-fixation condition, the focus of spatial attention was narrowed in order to minimize the perception of the 3D scene. Subjects were asked to perform a very demanding fixation task in which they needed to press one of two buttons to indicate the 200 ms luminance
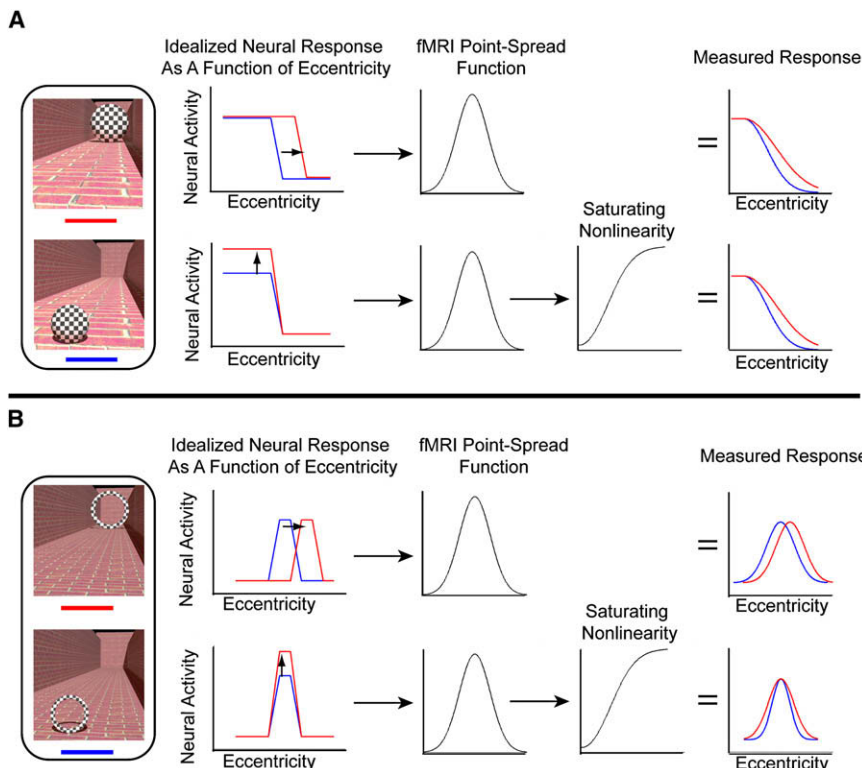

Figure 2. Sphere versus Ring

(A) A greater distribution of activity in V1 to the perceptually larger back sphere could result from either a positional shift in the cortical representation of the edges of the sphere (top) or an increase in the neural activity to the back sphere in combination with a saturating nonlinearity in the fMRI response (bottom). Note that the idealized activity responses are assuming that fixation is maintained at the center of the sphere.
(B) These two alternative explanations can be distinguished by using a ring presented at close and far apparent depths while subjects fixate the center of the ring. Specifically, in the current experiment using rings, we will look for a shift in the location of the maximum response rather than a shift in the distribution.

change (increase or decrease) of the fixation point as soon as possible. The luminance changes occurred randomly with a mean intertrial interval of 1.1 s drawn from a uniform distribution with a range of 1–1.2 s. Subjects reported having little awareness of the 3D rings and the rest of the scene while performing the task.

Behavioral data showed that the subjects had strictly followed the instructions. In the attend-to-ring experiment, they detected almost all the flickering pauses within 1 s of the onset of pause when attending to either the front (hit rate; 0.97 ± 0.02, mean ± standard error of the mean [SEM]) or the back (hit rate; 0.95 ± 0.02, mean ± SEM) ring. In the attend-to-fixation experiment, the hit rates for discriminating the luminance change were 0.81 ± 0.04 (mean ± SEM) when attending to the fixation of the front ring and 0.82 ± 0.04 (mean ± SEM) when attending to the fixation of the back ring. There was no significant difference between the front- and the back-ring conditions in both experiments.

fMRI Results

To first characterize the spatial distribution of activity to the front and back rings, we simply overlaid activation maps from the two stimuli on inflated representations of the cortical surface. In Figure 3A, the green and red regions on the inflated brain were significantly activated by the front and back flickering rings, respectively (both p values < 0.01, Bonferroni corrected, data from Subject S1, attend-to-ring condition). The overlap between these two regions is shown in dark yellow. Compared with the red region, the cortical location of the green region is shifted toward the foveal representation of V1, consistent with the different perceptions of the two rings. The other subjects showed a similar effect.

For a more precise comparison of the spatial distributions of V1 activation induced by the front and back rings, six regions of interest (ROIs) in V1 corresponding to six different eccentricities along the radius of the ring were defined with counter-phase-flickering 2D rings. The size of the 3D ring used in the hallway partially overlapped with the third- and fourth-largest rings used to define the ROIs. The psychophysical data showed that the front and back rings were perceived to be smaller and larger than their matched 2D sizes, respectively. Thus, we expected that compared to other ROIs, the third and fourth ROIs would be most strongly activated by the front and back rings, respectively. Event-related signal averages for each subject were calculated with the fixation periods serving as baseline and the average response 6–10 s after the start of flickering served as a peak measurement for each ROI.

Figure 3B shows the peak measurements in each of the six ROIs for the front and back rings. The ROIs are arranged in increasing order of eccentricity as determined by the size of the flickering annuli used to define the ROIs. In the attend-to-ring experiment, all four subjects' data showed that the spatial distribution of V1 activation induced by the back ring was shifted toward a more eccentric position in V1, whereas that induced by the front ring was shifted toward the more foveal representation in V1. The distributions clearly peak at different cortical positions (the 3rd or 4th ROI). A two-way repeated-measures ANOVA (eccentricity × ring-position) was performed. In addition to the obvious main effect of eccentricity, there was a significant eccentricity × ring-position interaction, $F(3,5) = 18.0$, $p < 0.0001$. The shifts between the front and back ring were further confirmed by planned comparisons at each eccentricity. There were statistically significant differences between the signals induced by the front and back rings at the 2nd, 3rd, 4th, and 5th ROIs (all $p < 0.05$). This result was consistent with

the perceptual appearances of the front and back rings and our prediction. In the attend-to-fixation experiment, a repeated-measures ANOVA also revealed a significant interaction, $F(3,5) = 5.9$, $p = 0.003$. However, although a peak difference was found in three out of the four subjects, the shifts of the spatial distributions of V1 activation were smaller than those in the attend-to-ring experiment. The only significant difference between the front and back ring was at the 3rd ROI ($p < 0.05$).

To further characterize the shifts in the distributions, we fit each subject's data with a three-parameter Gaussian function (width, height, and position along the eccentricity dimension). Next, for each attention condition, we calculated the shift in the fMRI response by subtracting the position parameters for the fits for the front and back responses. This resulted in a single metric that characterized the shift in the spatial distributions of V1 activation for the front and back spheres. The average shift in the attend-to-ring condition was 0.55 degrees, and the average shift in the attend-to-fixation condition was 0.17 degrees. A paired two-tailed t test comparing the shift for the attend-to-ring and attend-to-fixation conditions was significant ($p < 0.05$). The width and height parameters were not significantly different.

There were significant between-subject variations in the strength of the size illusion measured in the psychophysics experiment. To assess whether these differences could be explained by differences in the fMRI signal, we compared each subject's fMRI difference between the front and back rings (assessed by the Gaussian parameter estimates) and the magnitude of their size illusion measured in the psychophysics experiment. This relationship is shown in Figure 4A. With the limited number of subjects as a noted caveat, there is a relationship between size of the fMRI difference and the magnitude of the size illusion ($r = 0.94$) in the attend-to-ring condition (Figure 4A, left). In other words, the spatial shift in peaks in the fMRI response on the cortical surface for each subject appears to predict the magnitude of each subject's size illusion. This relationship is not present ($r = 0.0$) in the attend-to-fixation condition (Figure 4A, right), which further demonstrates the role of attention in mediating the effects of 3D context in V1. However, more data are needed to assess the statistical reliability of these between-subject effects.

Although the differences in the spatial distribution of activity are clearly smaller when attention is directed away from the 3D scene in the attend-to-fixation condition, a significant difference nonetheless remains. This may simply be the result of residual processing of the 3D scene as a result of incomplete removal of attention. Alternatively, the change in the spatial distribution of activity in V1 may partially be a stimulus-driven process and independent of perception of the 3D scene. To attempt to differentiate these possibilities, we examined the relationship between task performance in the fixation task and the fMRI difference measured in the attend-to-fixation condition. We reasoned that higher task performance would be associated with increased removal of attention from the scene, thereby reducing the separation of the fMRI peaks to the front and back ring. Consistent with this prediction, a negative correlation ($r = −0.77$) was observed (Figure 4B)—greater task performance at fixation resulted in a smaller spatial separation of the fMRI peaks.

Control Experiments

Our results thus far have shown that removing attention from the 3D scene reduces the cortical separation of the responses to the front and back rings. Our contention is that this reduces
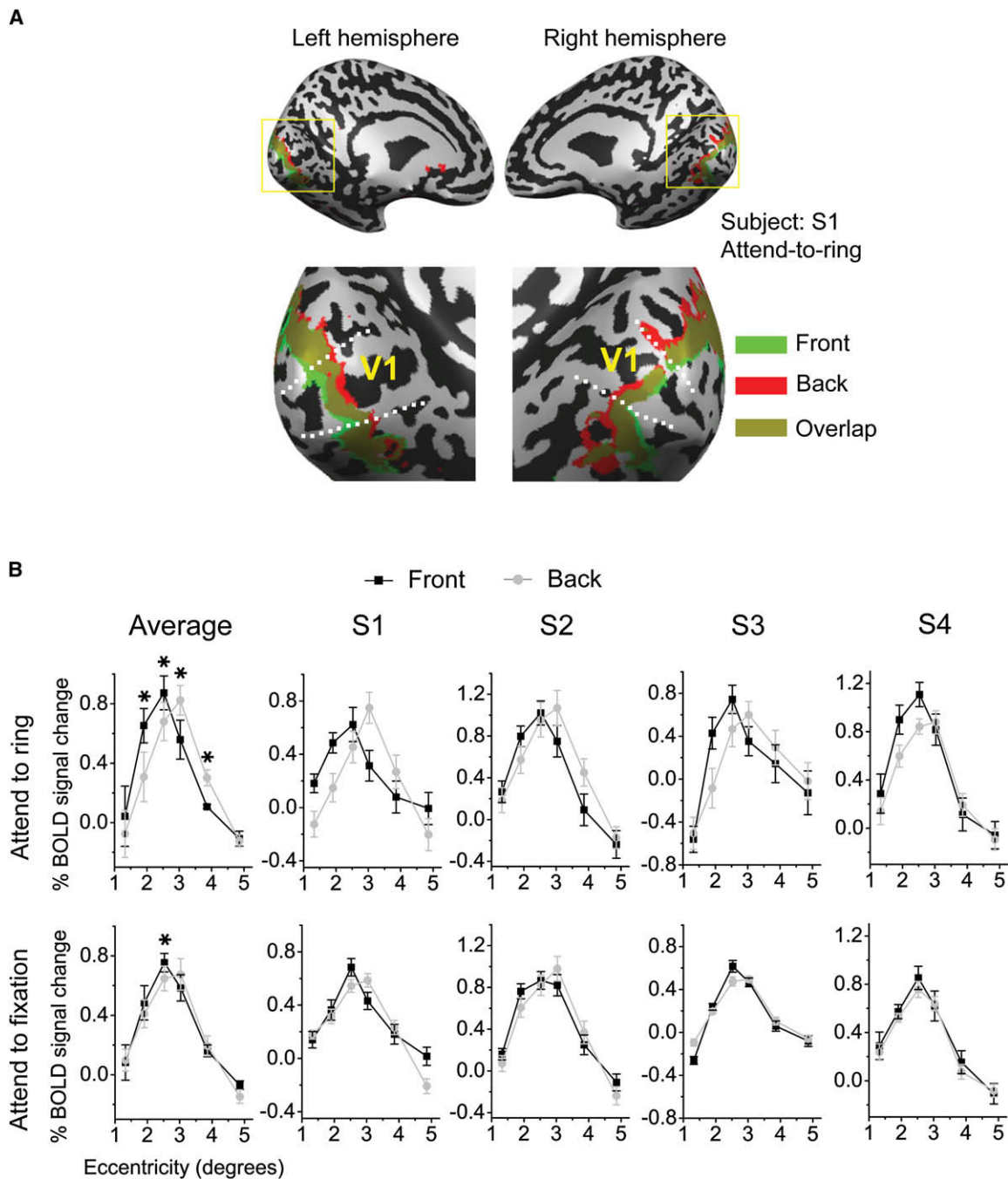
A

B

Figure 3. fMRI Results

(A) Cortical activation maps induced by the front and back flickering rings. The back view of the inflated left and right hemispheres from S1 is shown in the upper part. The regions in the yellow boxes are amplified and shown in the lower part. V1 is defined by retinotopic mapping, and its boundaries are indicated by the white dashed lines. The green and red regions were activated by the front and back flickering rings, respectively (for both, p < 0.01, Bonferroni corrected). The overlap between these two regions is shown in dark yellow. Compared with the red region, the green region shifts toward the foveal representation of V1.

(B) Peak fMRI signals from the six ROIs in V1 were plotted as a function of eccentricity for four individual subjects and their average. The spatial distribution of V1 activation induced by the back ring (perceptually larger) was shifted toward a more eccentric representation of the visual field in V1, whereas that induced by the front ring (perceptually smaller) was shifted toward the foveal representation. This resulted in peaks at cortical locations representing different eccentricities. The shift was significantly larger in the attend-to-ring experiment than the attend-to-fixation experiment. * p < 0.05. Error bars represent SEM.

feedback from higher-level visual areas that process the 3D scene. However, the evidence that the attention manipulation reduces processing of 3D scene cues in higher visual areas is only a conjecture based on the results of previously reported

attention experiments. Unfortunately, assessing the response in high-level visual areas was precluded by our high-resolution scanning sequence, which did not reliably cover the visual areas mostly likely involved in processing the 3D scene—the
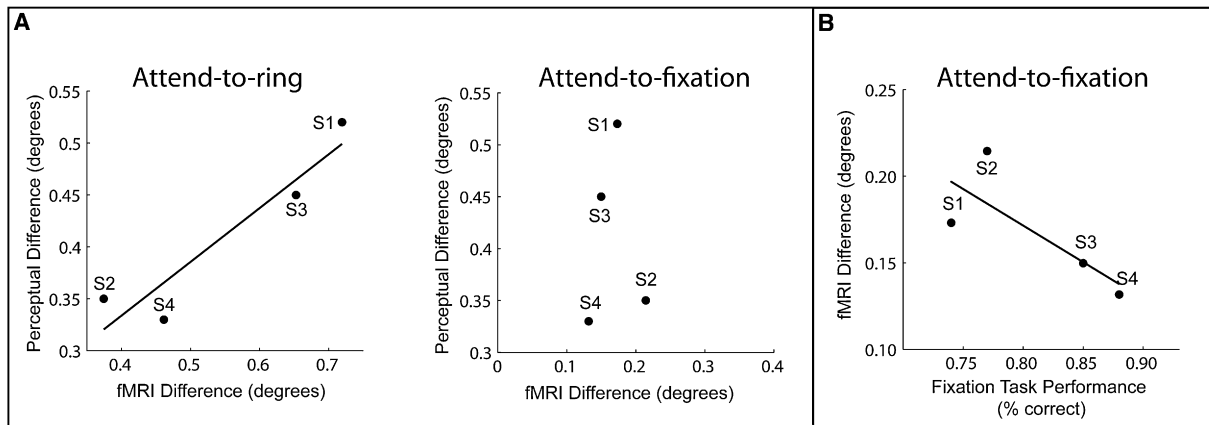
Figure 4. Between-Subject Analysis

(A) The perceptual difference between the front and back rings for each subject plotted against the difference in the cortical position of the peak of the fMRI response in the attend-to-ring condition (left) and the attend-to-fixation condition (right).

(B) The fMRI difference in the attend-fixation condition plotted against fixation-task performance.

lateral occipital complex (LOC) and the parahippocampal place area (PPA). To address this issue directly, we used standard imaging resolution and a simple 14 s block design to collect fMRI data from two subjects while they alternated between the two attention tasks. There was a large effect observed in the two ROIs. Specifically, activities in the LOC and PPA were significantly reduced during the attend-to-fixation task relative to the attend-to-ring task (see Supplemental Data 2), suggesting that the attention manipulation is effective at reducing activity in—and feedback from—these higher-level visual areas.

Our results in the main experiment demonstrate that removing attention from the 3D scene reduces the cortical separation of the responses to the front and back rings—a result we argue is consistent with the reduction of feedback from higher visual areas that process the 3D depth cues. However, there is a separate question about the possible role of spatial attention. Specifically, could any difference in the allocation of spatial attention be driving the basic effect of a difference between the front and back ring? For example, in the attend-to-ring condition, perhaps subjects attended more to the outer edge of the back ring and the inner edge of the front ring, thereby leading to the differences in the spatial distribution of V1 activity. Although there is no a priori reason to expect such differences in attention strategies to the front and back ring, ruling out this possibility is important. We performed a behavioral experiment that required subjects to simultaneously perform the flickering-pause detection task while also detecting a small target that could appear on either the inner or the outer edge of the ring. The target-detection performance was identical for the front and back rings at the inner and outer edges, suggesting that there are no differences in the allocation of spatial attention (see Supplemental Data 3).

Finally, to demonstrate that effects we observed in the main experiment were due to the 3D context, rather than to some other factors (e.g., different eye positions for the front and back rings), we performed a control experiment with S1 and S2 from the main experiment. The stimuli and task were the same as the attend-to-ring experiment, but the 3D background was replaced with a uniform gray background. No perceptual differences were observed between the two rings, and no

shifts of the spatial distributions of V1 activation were found (see Supplemental Data 4).

Discussion

We have demonstrated that perceived eccentricity differences are reflected by shifts in the distribution of activity across the surface of V1. The activation in response to the perceptually larger ring occurred in a more eccentric position in V1 compared to the perceptually smaller ring. We believe that it is extremely unlikely that the shifts in cortical location can be explained by local interactions between the front and back rings and the 3D scene (e.g., local contrast) because the local interaction would have increased or decreased the overall magnitude of neural activity, rather than induced the shifts.

In addition to demonstrating that perceived size influences the location of activity in V1, we demonstrate that attention modulates this effect—narrowing the focus of spatial attention with a demanding central fixation task reduced the magnitude of the eccentricity shift in V1. All subjects reported being largely unaware of the rings and 3D scene when performing the demanding central fixation task, suggesting that explicit perception of the 3D scene is necessary for the spatial shifts in V1. This finding is important because it rules out other—stimulus-based—potential explanations for the spatial shifts in activity. One interpretation of the attention result is that the manipulation is altering the feedback of information from higher visual areas. Specifically, the depth cues that give rise to the perceived size difference include linear perspective and texture cues, both of which are probably processed beyond V1. We reason that narrowing the focus of spatial attention away from the 3D scene leads to reduced activity in, and therefore feedback from, the visual areas that process these cues. Given that shifts of attention occur within several hundred milliseconds, it suggests that the remapping in V1 can occur on a fast timescale.

What are the neural mechanisms that allow for the change in spatial distribution of activity? One of the most fundamental properties of V1 is the precise mapping of visual angle subtense arising from the specific neural connection pattern from the lateral geniculate nucleus. We do not claim that the direct feed-forward mapping of connections into V1 is changing.

In fact, our results clearly show that the primary determinant of the spatial distribution of activity in V1 is the retinotopic position of the stimulus. But our results also show that this spatial pattern can be significantly modified by 3D context and is consistent with the perceptual appearance of the stimulus. Given the dynamic nature of the remapping and the fact that the 3D cues are most likely processed in higher visual areas, we suggest that V1 activity is being modified by cortical feedback. The specific neural mechanisms that support these dynamic changes in V1 maps, along with their degree of flexibility, are important questions for future studies.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures and four figures and can be found with this article online at http://www.current-biology.com/supplemental/S0960-9822(08)01251-7.

References

1. Murray, S.O., Boyaci, H., and Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. Nat. Neurosci. *9*, 429–434.
2. Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598–601.
3. Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? Neuron *23*, 115–125.
4. Wojciulik, E., Kanwisher, N., and Driver, J. (1998). Modulation of activity in the fusiform face area by covert attention: An fMRI study. J. Neurophysiol. *79*, 1574–1579.
5. Murray, S.O., and He, S. (2006). Contrast invariance in the human lateral occipital complex depends on attention. Curr. Biol. *16*, 606–611.