Neuron, Vol. 29, 277–286, January, 2001, Copyright ©2001 by Cell Press

Neural Response to Perception of Volume in the Lateral Occipital Complex

Cassandra Moore* and Stephen A. Engel
Department of Psychology
University of California, Los Angeles
Los Angeles, California 90095

Summary

Projection of a 3D scene onto the 2D retina necessarily entails a loss of information, yet perceivers experience a world populated with volumetric objects. Using simultaneous behavioral and neural (fMRI) measures, we identify neural bases of volume perception. Neural activity in the lateral occipital cortex increased with presentation of 3D volumes relative to presentation of 2D shapes. Neural activity also modulated with *perceived* volume, independent of image information. When behavioral responses indicated that observers saw ambiguous images as 3D volumes, neural response increased; when behavioral data revealed a 2D interpretation, neural response waned. Crucially, the physical stimulus was identical under both interpretations; only the percept of volume can account for the increased neural activity.

Introduction

Natural images contain multiple sources of information that contribute to the percept of volume, including shading, texture, occlusion, and contour, among others. Nonetheless, 2D retinal image information alone is inadequate to produce a unique 3D interpretation of the visual input. An infinite number of combinations of luminance, surface reflectance, and surface shape can produce the same image (Freeman, 1994). The visual system supplements ambiguous image information with stored knowledge of environmental regularities to infer volumetric shape (Knill et al., 1996). Thus, the experience of seeing a volumetric object usually involves both a bottom-up (image-based) and a top-down (knowledge-based) component.

The neural bases of volume recovery, both top down and bottom up, have only been partially identified. Most human studies have focused on processing of particular image-based cues to depth. For example, functional magnetic resonance imaging (fMRI) was used to demonstrate that human V5/MT is preferentially activated by motion cues carrying information about 3D structure (Orban et al., 1999). Positron emission tomography (PET) studies comparing 2D and 3D random dot stereograms have shown cerebral blood flow changes in V1, V2, and the right inferotemporal cortex (Ptito et al., 1993). Comparison of strong and weak shading cues indicated that processing of 3D shape from shading may start as early as striate cortex (Humphrey et al., 1997). The existence of a cortical area sensitive to top-down cues to volumetric shape has yet to be demonstrated. Additionally, areas representing unified volume by combining multiple cues to depth have yet to be identified.

In the macaque, area MT has been implicated in perception of 3D structure defined by motion (Bradley et al., 1998) as well as stereo (DeAngelis et al., 1998), though initial processing of these cues begins in striate cortex (Hubel and Wiesel, 1970). The lower bank of the superior temporal sulcus (STS) has also been implicated in 3D structure from stereo (Janssen et al., 2000). STS is located in a subregion of the inferior temporal cortex (IT), which responds selectively to specific object features (Tanaka, 1996). Macaque IT, a region critical for object recognition (Mishkin et al., 1983; Logothetis and Sheinberg, 1996), may be analogous to the lateral and ventral regions of the human occipital lobe known as the lateral occipital complex, or LOC (Malach et al., 1995; Tootell et al., 1996).

Recent neuroimaging studies have implicated LOC, located posterior to human MT and anterior to retinotopic cortex (Sereno et al., 1995; DeYoe et al., 1996; Tootell et al., 1996), in the visual processing of whole or partial objects. A defining feature of LOC is that it responds more strongly to images of objects than to texture patches (Malach et al., 1995) or scrambled versions of object images (Kanwisher et al., 1996; Grill-Spector et al., 1998b). LOC is also driven by image fragments containing some object structure (Grill-Spector et al., 1998b) and by shapes defined by kinetic (Orban et al., 1999), stereo (Mendola et al., 1999), or illusory contours (Mendola et al., 1999). These studies suggest LOC is a midlevel processing area, more responsive to structured shape, either 2D or 3D, than to random patterns.

We hypothesized that, as a midlevel area sensitive to structure in images, LOC might be the region that supports the earliest representations of volume. We report two studies designed to test the hypothesis that LOC is responsible for integrating volume information from both bottom-up and top-down sources to construct representations of volumetric object shape. Using simultaneous neural (fMRI) and behavioral measures, we recorded neural activity while observers were perceiving volume and while they were not. Crucially, this paradigm enabled the differentiation of neural responses related to bottom-up volume information and neural responses related to top-down volume information.

Sensitivity to bottom-up information was measured by comparing neural responses to rendered volumetric objects with responses to flat shapes (Figure 1, 3D primes versus 2D primes). During scanning, observers indicated whether each 3D object had a plane of symmetry and whether each 2D shape had a horizontal axis of symmetry. The task was designed to focus attention on either the volumetric or flat structure of the depicted object.

Sensitivity to top-down volume information was measured using novel stimuli (Figure 1, 3D two-tones, 2D two-tones) that could be primed to appear either volumetric or nonvolumetric. Similar two-tone images of fa-

* To whom correspondence should be addressed (e-mail: seymour@ucla.edu).

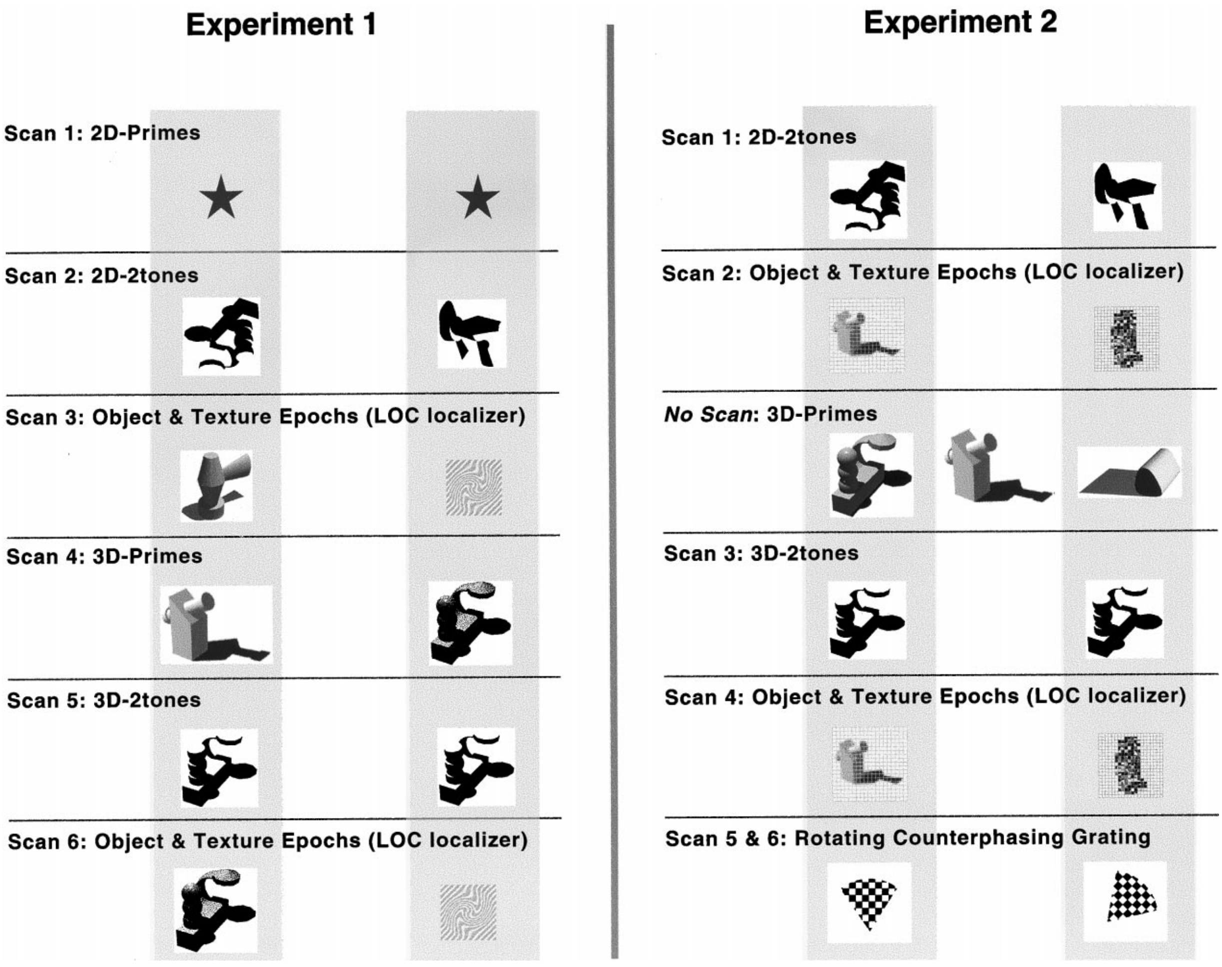

Figure 1. Stimuli

(Experiment 1) Scan 1 primed observers with 2D shapes (for brevity, 2D primes) to encourage a nonvolumetric interpretation. Scan 2 tested observers' behavioral and neural responses to the two-tone images (2D two-tones). Scan 3 was an LOC localizer containing epochs of objects and epochs of texture patches. Scan 4 primed observers with grayscale images of 3D volumes (3D primes), and scan 5 tested observers' response to the volume-primed two-tone images (3D two-tones). Scan 6 was a replication of the LOC localizer.
(Experiment 2) Scan 1 tested the two-tones without any primes (2D two-tones). Scan 2 was an LOC localizer, containing epochs of scrambled and intact objects. During a 5 min period between scans, observers studied grayscale objects (3D primes). Scan 3 tested the volume-primed two-tones (3D two-tones); scan 4 replicated the LOC localizer. Scan 5 and scan 6 mapped retinotopic cortex as observers watched a rotating wedge containing a contrast-reversing checkerboard pattern.

miliar objects have proven advantageous in studies of face perception (Kanwisher et al., 1998) and perceptual learning (Dolan et al., 1997), because the objects they depict are not immediately identifiable. Embedding a familiar object in noise (Dolan et al., 1997) or inverting a face (Kanwisher et al., 1998) can disguise the identity of the depicted object. Then, if attention is drawn to a distinctive feature of the object or the face is righted, the same two-tone image suddenly contains a nameable object.

Two-tone images of *novel* objects need not be disguised. Unprimed objects in two tone usually appear to be black shapes on a white background (Moore and Cavanagh, 1998). However, exposure to grayscale images of a similarly shaped 3D objects can give rise to a very different percept. After priming with grayscale images of 3D objects (Figure 1, 3D primes), novel objects in two-tone images appear volumetric (C. Moore and S. A. Engel, 2001). Black and white regions are integrated to form a unified volume, and cast shadows are parsed from the object. We used this phenomenon to create percepts of flatness, then volume, without altering the stimulus.

Perceptual experience of volume was measured by a behavioral task (Figure 2) performed during scanning. The observer decided whether a bright red spot was "On" or "Off" the object depicted in each two-tone image (Figure 2). If the object in two-tone appeared 2D, observers would say all spots on black regions were On the object and spots on white regions were On the background. If the two-tone objects appeared volumetric, some spots in the black region would appear on the cast shadow (i.e., Off the object), and some spots in the white image region would appear on the object surface. Thus, responses differentiating object surfaces, cast shadows, and background would indicate that the observer perceived volume.

The cortical locations queried for differential response

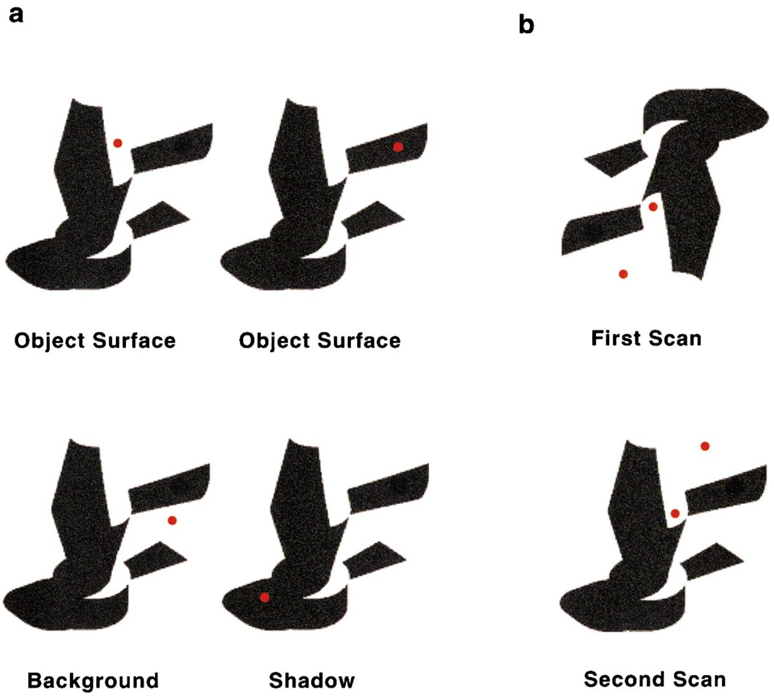

Figure 2. Tasks: Experiments 1 and 2

(A) In experiment 1, perceived volume was measured by requiring observers to decide whether a red spot was on the surface of each two-tone object and register their response by pressing a button. The red spot appeared in one of four locations in each image. First row: white object surface, black object surface. Second row: (white) background, (black) cast shadow. Previous studies (Moore and Cavanagh, 1998; C. Moore and S. A. Engel, 2001) have shown that when two-tone images are not seen as volumetric a figure/ground segregation is performed. Black image regions, including cast shadows, are seen as unified shapes, and white regions are seen as background. When volume is perceived, object surfaces, background, and cast shadows are differentiated; black and white surfaces are integrated to form a volumetric object.

(B) In the first two-tone scan of experiment 2, observers decided whether the spots on the image could be connected by a single straight line that did not pass through the black image region. The images were rotated either 90° or 180° in the picture plane to enhance their 2D appearance. Spot placement caused connecting lines to pass very close to the edge of the black shape. In the second two-tone scan, after priming with volumetric grayscale objects, observers reported whether either spot was on the object.

to volume (both top down and bottom up) were defined by two criteria. Anatomically, the region of interest, LOC, was restricted to the inferior and middle occipital gyri, the lateral and inferior occipital sulci, and the sulcus lunatus (Figure 3). Functionally, LOC was restricted to voxels within the anatomical region that produced a significantly greater response to images of 3D grayscale objects than to texture patterns (Figure 1) in the LOC localizer scans.

In experiment 1 (Figure 1), scan 1 primed observers with 2D shapes (for brevity, 2D primes) to encourage a nonvolumetric interpretation. Scan 2 tested observers' behavioral and neural responses to the two-tone images (2D two-tones). Scan 3 was an LOC localizer containing epochs of objects and epochs of texture patches. Scan 4 primed observers with grayscale images of 3D volumes (3D primes), and scan 5 tested observers' response to the volume-primed two-tone images (3D two-tones). Scan 6 was a replication of the LOC localizer.

In experiment 2 (Figure 1), scan 1 tested the two-tones without any primes (2D two-tones). Scan 2 was an LOC localizer, containing epochs of scrambled and intact objects. During a 5 min period between scans, observers studied grayscale objects (3D primes). Scan 3 tested the volume-primed two-tones (3D two-tones); scan 4 replicated the LOC localizer. Scan 5 and scan 6 mapped retinotopic cortex as observers watched a rotating wedge containing a contrast-reversing checkerboard pattern.

We reasoned that if the neural response in LOC modulated with the observer's perceptual experience of volume, then activity should increase when the two-tones appeared volumetric. This result would provide evidence of LOC sensitivity to top-down volume information. However, if LOC was *not* sensitive to top-down information, activity should remain constant across two-tone image conditions, regardless of whether the observer perceived the depicted object as a volume. Anal-

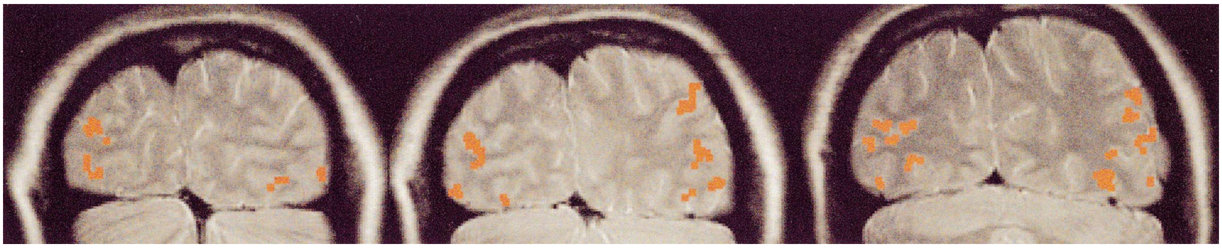

Figure 3. Anatomical and Functional Definitions of LOC

Anatomically, LOC included the inferior and middle occipital gyri, the lateral and inferior occipital sulci, and the sulcus lunatus, when present. Functionally, LOC was limited to voxels within the anatomical area producing a greater response to objects than textures; threshold pictured is $p < 0.05$.

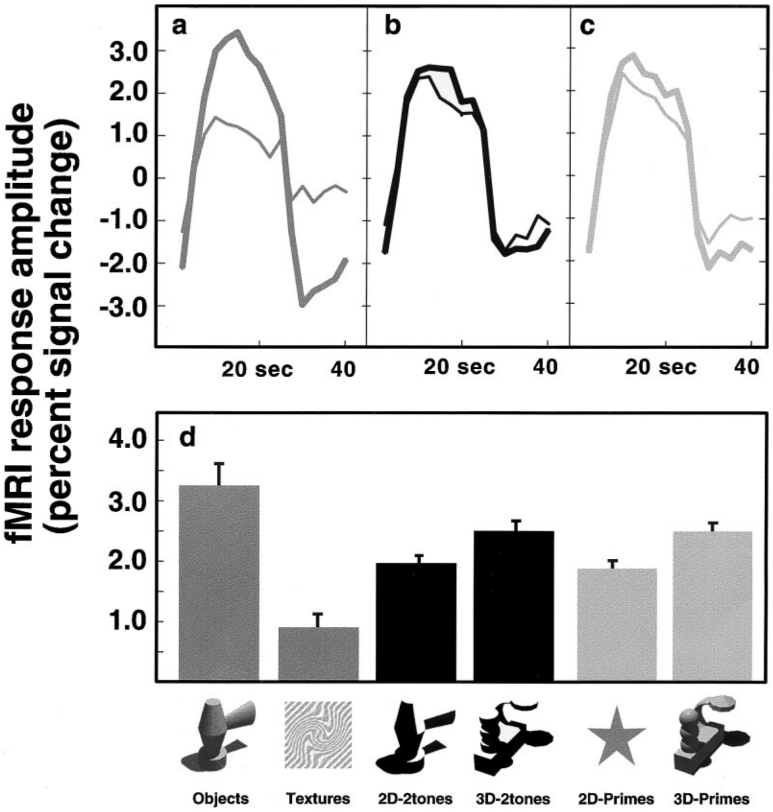

Figure 4. Results: Experiment 1

(A–C) MR time courses. (A) Objects and textures, (B) 2D two-tones and 3D two-tones, and (C) 2D primes and 3D primes, averaged across six epochs per condition for the seven observers. The bolder lines depict the condition with the greatest response within a comparison pair.

(D) Response amplitudes. Grayscale objects (3D primes) produced significantly greater ($p < 0.003$, corrected $p < 0.017$) neural activity than 2D primes (right light gray bars). Two-tone images primed by 3D objects (right black bar) produced a significantly greater ($p < 0.003$, corrected $p < 0.017$) response than identical images primed by 2D shape (left black bar). The dark gray bars (left) depict object > texture differences in these voxels. Note also that 2D primes produced a greater response than textures ($p < 0.0001$, corrected $p < 0.017$). Bars depict mean response amplitudes with one standard error (SE) of the difference of the two tones and of the primes and one SE of the individual means for objects and for textures.

ogously, if LOC was sensitive to bottom-up information about volume (e.g., shading, occlusion), neural responses collected during the 3D prime scan should be greater than those collected during the 2D prime scan. Such an outcome would provide evidence for processing of bottom-up volume information. Finally, if the same regions were responsive to both types of information, LOC might be a region in which top-down and bottom-up information is integrated to form the unified representation of volumetric objects human perceivers experience.

Results

In the first scan, 2D primes (flat shapes like stars and squares) were presented in order to suggest that the two-tones presented in the second scan would be 2D. Perceptual sensitivity to volume information was measured by behavioral responses to two-tone stimuli. Signal detection techniques (see Experimental Procedures) provided a measure of volume detectability (d′) for each epoch in each two-tone scan.

After seeing the 2D primes in scan 1, observers detected little or no volume during scan 2, the 2D two-tone scan, ($d' = -0.044 \pm 0.27$, 95% confidence interval [abbreviated CI hereafter]). There was a pronounced bias to call all black regions On and all white regions Off the object surface, suggesting observers saw the black image region, shadow included, as a flat black object on a white background. Across sequential epochs of the 2D two-tone scan, a simple regression showed no linear trend in performance (slope $= -0.02$, $t[30] = 0.28$, $p > 0.78$), and an ANOVA of the residuals provided no evidence of nonlinear trends ($F[5,24] = 1.36$, $p > 0.28$). These results indicate the two-tones consistently appeared to be 2D black shapes on a white background.

Presentation of the 3D primes (grayscale images of volumetric objects) in scan 4 created an alternative perceptual expectation that included volumetric objects, shadows, and highlighted object surfaces. During scan 5, the 3D two-tone scan, the two-tones appeared consistently volumetric; the object surface included both black (nonilluminated) and white (highlighted) regions. Volume detectability scores were high ($d' = 1.22 \pm 0.1$, 95% CI), and both variability and bias were low. Neither a linear trend in performance (slope $= -0.09$, $t[30] = 1.09$, $p > 0.28$) nor a nonlinear trend in residuals ($F[5,24] = 0.04$, $p > 0.99$) was apparent, suggesting that perception of volume was relatively consistent across epochs in the scan.

The change in perceived volume from the first to the second two-tone scan was highly significant ($t[30] = 8.72$, $p < 0.0001$). Importantly, the stimuli themselves did not become more volumetric; the change in percept was the result of changing perceiver expectations about object dimensionality.

Corresponding neural activity was measured in lateral occipital cortex. LOC was defined as described earlier (see Experimental Procedures for details). Within LOC, the amplitude of the best-fit sinusoid for the average fMRI signal time course for each epoch of each experimental condition was calculated. This "response amplitude" served as the measure of neural activity in all

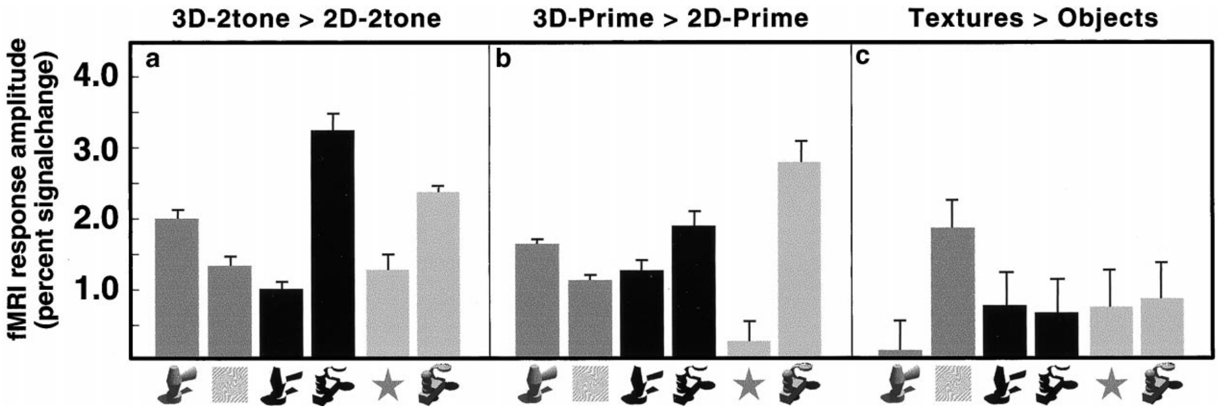

Figure 5. Additional Analyses: Experiment 1

(A) Includes only voxels with a reliable ($p < 0.05$) 3D two-tone > 2D two-tone (top-down volume) preference. These voxels also produced a reliable ($p < 0.004$, corrected $p < 0.008$) 3D prime > 2D prime (bottom-up volume) response.

(B) Includes only voxels with a reliable ($p < 0.05$) 3D prime > 2D prime (bottom-up volume) preference. These voxels also produced a reliable ($p < 0.001$, corrected $p < 0.008$) 3D two-tone > 2D two-tone (top-down volume) response.

(C) Illustrates that sensitivity to volume is not a ubiquitous property of occipital cortex. Voxels that produced a texture > object ($p < 1$) response were quite variable and showed no preference for volume in either the bottom-up ($p > 0.42$) or top-down ($p > 0.57$) conditions.

subsequent analyses. Mean response amplitudes were compared in paired t tests of the relevant conditions.

Neural activity in LOC modulated with the perceptual experience of the observer, not the unchanging stimulus. Response amplitudes, like d′ scores above, were significantly higher (Figure 4) in scan 5, the 3D two-tone scan, than scan 2, the 2D two-tone scan ($t[6] = 4.035$, $p < 0.003$). The Bonferroni correction for multiple tests yielded a $p < 0.017$, hereafter refered to as a "corrected p." When the two-tone objects were perceived as volumetric, neural activity was greater than when the same two-tones did not appear volumetric. Thus, LOC appears to be sensitive to top-down, knowledge-based volume information.

Comparison of the neural responses in scan 4, the 3D prime scan, and scan 1, the 2D prime scan, revealed that LOC was also sensitive to bottom-up, image-based volume information. The 3D primes produced a significantly greater response than the 2D primes ($t[6] = 4.36$, $p < 0.003$, corrected $p < 0.017$), suggesting an LOC preference for volumetric objects over 2D shapes (Figure 4). Interestingly, the 2D prime response was greater than the response to texture patches in the localizer scan ($t[6] = 6.81$, $p < 0.0001$, corrected $p < 0.017$). This result supports earlier studies showing a preference for organized 2D shapes (Mendola et al., 1999) and suggests LOC may have a graded response to different levels of image structure.

The previous analyses established that regions of the lateral occipital cortex were sensitive to both bottom-up and top-down volume information. The next comparisons (Figure 5) confirmed that (1) the same cortical regions were sensitive to top-down and bottom-up information, (2) sensitivity to volume information was not a ubiquitous property of occipital cortex, and (3) increased activity in the 3D two-tone scan was not simply due to stimulus repetition (Henson et al., 2000).

We identified voxels within the anatomical confines of LOC that were preferentially responsive to 3D two-tones (compared to 2D two-tones) and tested these for sensitivity to 2D primes versus 3D primes and objects versus textures (Figure 5A). The results indicated that voxels sensitive to top-down volume information were also sensitive to bottom-up volume information ($t[6] = 3.94$, $p < 0.004$, corrected $p < 0.008$) and more responsive to objects than to textures ($t[6] = 3.92$, $p < 0.003$, corrected $p < 0.008$). Analogously, voxels responsive to bottom-up volume information, those with a 3D prime response greater than 2D prime response, were tested for sensitivity to top-down volume information (3D two-tones versus 2D two-tones) and objects versus textures (Figure 5B). Voxels sensitive to bottom-up volume information were also sensitive to top-down volume information ($t[6] = 4.78$, $p < 0.001$, corrected $p < 0.008$) and more responsive to objects than to textures ($t[6] = 9.14$, $p < 0.00002$, corrected $p < 0.008$).

We also located voxels in the anatomical region defined as LOC that had a greater response to textures than objects. On average, these voxels had a texture response that was somewhat greater than previously selected voxels, but, across conditions, response amplitudes were low, and variance was high (Figure 5C). The voxels were tested for differential response to 2D versus 3D two-tones, and 2D versus 3D primes in order to determine whether *all* voxels in LOC had a preferential response to volume. Neither a preference for 3D two-tones ($t[6] = -0.173$, $p > 0.57$, corrected $p < 0.008$) nor a preference for 3D primes ($t[6] = 0.211$, $p > 0.42$, corrected $p < 0.008$) was found. Sensitivity to volume was not ubiquitous in the lateral occipital cortex.

Finally, we compared response amplitudes across scans and across epochs (see Experimental Procedures, Comparing Epochs) to exclude the possibility that the increased activity recorded in the 3D two-tone scan was due to stimulus repetition rather than perception of volume. We reasoned that if stimulus repetition was responsible for increased neural activity, differential activation should be evident in (1) a comparison of earlier and later epochs of a scan in which stimuli were repeated and/or (2) a comparison of earlier and later scans in which stimuli were repeated.

Stimuli were repeated in the six sequential epochs of four scans: 2D two-tones, 3D two-tones, 2D primes, and 3D primes. Each stimulus was repeated approximately three times in each scan. Only the 3D prime scan revealed a significant change in response amplitude over sequential epochs (slope $= -0.002$, $t = 2.38$, $p < 0.02$, corrected $p < 0.013$). The negative slope indicates that neural response *diminished* with repeated viewing of the stimuli. The other three scans also had negative slopes but did not reach statistical significance (all $p > 0.19$). A test of residuals revealed no further trends in any of the scans (all $p > 0.09$).

Stimuli were also repeated in the two localizer scans (3 and 6). The same objects doubled as 3D primes in scan 4 and were represented as two-tones in scans 2 and 5. Thus, by scan 6, each object had been presented three times in the 3D prime scan, twice in the localizer scan, and six times in the two-tone image scans. Despite heavy repetition, there was no discernable change in response amplitude across sequential epochs (slope $= -0.001$, $t = 1.12$, $p > 0.27$) nor any trends in residuals ($p > 0.74$).

No significant increase in response amplitude across epochs or across scans was revealed in five separate tests. Instead, neural response appeared to *decrease* with stimulus repetition. Thus, it seems the enhanced response recorded in the 3D two-tone scan was due to perception of volume, not some vagary of stimulus repetition.

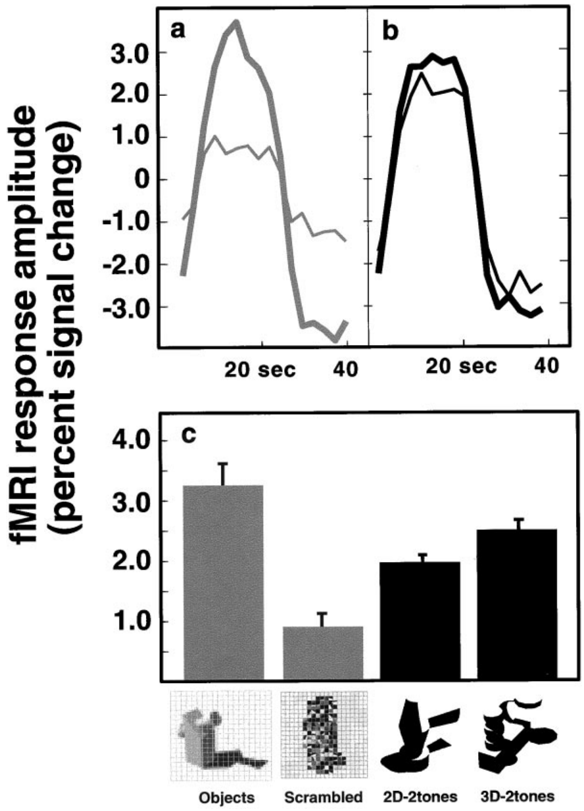

Figure 6. Results: Experiment 2

(A and B) MR time courses. (A) Objects and textures, (B) 2D two-tones, and 3D two-tones, averaged across six epochs per condition for the seven observers. The bolder lines depict the condition with the greater response within a comparison pair.

(C) Response amplitudes. Two-tone images primed by 3D objects (right black bar) produced a significantly greater ($p < 0.009$) response than identical nonprimed images (left black bar). Mean response amplitudes are depicted with one SE of the difference for the two-tone conditions and one SE for object and for textures.

Experiment 2 replicated and extended the results of experiment 1. Once again, neural signal in LOC increased when viewed objects were perceived as volumetric (Figure 6). In contrast, earlier visual areas did not respond preferentially to perceived volume (Figure 7). Mapping of retinotopic organization in cortex (Engel et al., 1994; Sereno et al., 1995) enabled identification of a ventral area corresponding to V4, VP, and V8; a dorsal region corresponding approximately to V3, V3a, and V7; and a early area encompassing V1 and V2. No retinotopically organized regions were included in LOC.

From each area, voxels showing a significantly greater response to grayscale objects than fixation epochs were chosen (see Experimental Procedures, Defining Retinotopic ROIs). The response to 3D two-tones was not significantly different from the response to 2D two-tones (Figure 7) in the early ($t[6] = 1.15$, $p > 0.143$) and ventral ($t[6] = 1.14$, $p > 0.146$) areas. LOC again responded more strongly to 3D two-tones than 2D two-tones ($t[6] = 4.39$, $p < 0.002$, corrected $p < 0.013$). Dorsal areas (Figure 7) produced a numeric but unreliable difference ($t[6] = 1.22$, $p > 0.132$, corrected $p < 0.013$) due to within-subject variation. The suggestion of sensitivity to volume in this area may reflect a dorsal stream concern with motor interaction with objects or perhaps to spatial layout of 3D objects casting shadows on a ground plane (Mishkin et al., 1982; Nagahama et al., 1996; Andersen et al., 1997; Sakata et al., 1997).

Experiment 2 also confirmed that the results of experiment 1 were independent of observer effort or attentional effects. In experiment 1, the integration involved in seeing the two-tone objects as volumes (3D two-tones) may have required more effort than the figure/ground segregation performed when the two-tones did not appear volumetric (2D two-tones). If so, observer effort may have influenced neural response. In experiment 2, the 2D two-tone task was made more difficult than the 3D two-tone task. We reasoned that if effort, not volume, was responsible for the neural activity differences in experiment 1, then pairing an easy task with 3D two-tones and a difficult task with 2D two-tones should eliminate or reverse that difference.

Two red spots appeared on the white region of each two-tone image (Figure 2). In the 2D two-tone scan, observers decided whether the spots could be connected by a single straight line that did not pass through the black image region. Task difficulty was increased by placing spots such that a connecting line would pass very close to the edge of the black shape. The images were rotated either 90° or 180° in the picture plane to enhance their 2D appearance. Behavioral measures confirmed that the task used in the 2D two-tone scan ($d' = 0.774 \pm 0.26$, 95% CI; mean RT = 401) was more difficult than the task used in the 3D two-tone scan ($d' = 3.08 \pm 0.64$, 95% CI; mean RT = 368). A paired t test of the d' scores from each condition for the seven observers produced a significant difference ($t[6] = 3.83$, $p < 0.006$), whereas mean reaction times did not ($t[6] =$

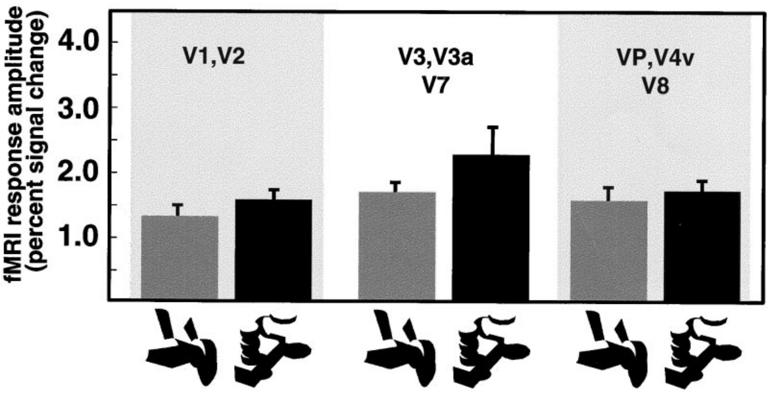

Figure 7. Retinotopic ROI: Experiment 2

None of the three retinotopic regions preferred two-tone images that appeared volumetric over those that did not (all $p > 0.13$). Although the difference in the dorsal region (V3, V3a, and V7) is not reliable, it suggests some sensitivity to the spatial relationships of surfaces in the images. Mean response amplitudes and one SE for seven observers are shown for each condition in each visual area.

1.59, $p > 0.10$). There was no evidence of a linear trend across epochs in the 2D two-tone scan (slope = 0.01, $t = 0.17$, $p > 0.87$) nor any trends in residuals ($p > 0.69$).

In the 3D two-tone scan, which had been primed by grayscale objects, observers reported whether either red spot was on the two-tone object. Volume detectability was high ($d' = 3.08 \pm 0.64$, 95% CI; mean RT = 368), and there was no evidence of a linear trend across epochs (slope = 0.07, $t = 0.37$, $p > 0.71$) nor any trends in residuals ($p > 0.79$).

Neural activity in LOC increased when the two tones were seen as volumes. Response amplitudes (Figure 6) from the 3D two-tone scan were reliably higher than those from the 2D two-tone scan ($t = 3.2$, $p < 0.009$). Once again, comparisons across epochs in the two-tone scans and comparisons across the two localizer scans were conducted to insure that stimulus repetition effects were not responsible for the rise in neural activity associated with the percept of volume. Neural activity decreased significantly across epochs in both the 2D two-tone scan (slope = -0.002, $t = 3.07$, $p < 0.004$, corrected $p < 0.013$) and the 3D two-tone scan (slope = -0.002, $t = 2.71$, $p < 0.01$, corrected $p < 0.013$). Neural activity was unchanged across localizer scans (slope = -0.001, $t = 1.15$, $p > 0.26$), even though objects in the last localizer scan had been seen up to 20 times in different formats. No tests of residuals were significant (all $p > 0.46$).

Using a new and more difficult behavioral task, the results of the first experiment were replicated, indicating that perception of volume, not observer effort or stimulus repetition, was the crucial factor.

Discussion

Perception of volume substantially increased neural activity in the LOC. A crucial feature of these experiments was the demonstration of a strong link between perceptual experience and neural activity. When primed with 2D shapes, objects in the two-tone images were not perceived as volumes; when primed with 3D objects, they were. Simultaneously recorded neural activity increased when volume was perceived. Thus, we conclude that LOC is a neural substrate of volume perception.

We have tested and rejected several alternative explanations of our results. However, one alternative merits further consideration. In both experiments, 2D two-tone responses were collected before 3D two-tone responses. During the intervening periods, observers studied grayscale versions of the objects that were subsequently presented in two tone. How can we know that perceived volume, rather than some aspect of stimulus repetition, caused the increase in neural activity?

Existing literature is equivocal on the topic. Repetition of stimuli typically causes a *decrease* in neural response (e.g., Ungerleider, 1995; Tootell et al., 1996; Buckner et al., 1998; Grill-Spector et al., 1999; James et al., 1999). Repetition suppression in LOC has been demonstrated when the format of the depicted object was changed from a line drawing to a grayscale photograph. This result suggests that the presentation of grayscale primes might have reduced the neural response to the two-tone objects. However, Henson et al. (2000) reported an increase in neural response to repetition of unfamiliar symbols and faces in the inferotemporal fusiform region.

Since perceiving volume in the two-tone scans appears to *require* exposure to grayscale images, repetition and volume cannot be directly dissociated. However, if stimulus repetition was responsible for increased neural activity, differential activation should be evident in (1) a comparison of earlier and later epochs of the same scan in which stimuli are repeated and/or (2) a comparison of earlier and later scans in which stimuli are repeated. Of the ten comparisons we tested, none showed an increased response with stimulus repetition. Neural response in both two-tone scans in experiment 2 decreased significantly over time. Response in the other eight comparisons decreased, but slopes were not statistically significant.

Thus, the priming procedure we used may actually underestimate the magnitude of the LOC preference for volume due to repetition suppression. Despite stimulus repetition, response amplitudes were higher in the later (3D two-tone) scan, suggesting the preference for perceived volume in area LOC was substantial.

We suggest that LOC is the first region in the ventral processing stream (Mishkin et al., 1983) to support a unified representation of volumetric object shape. Little or no evidence of a preferential response to volume was seen in earlier visual areas. Individual image-based cues to depth (e.g., stereo) may be processed separately in early visual cortex (Hubel and Wiesel, 1970), then integrated in LOC. Knowledge-based information about structural regularities in the world might supplement image-based information, resolving residual ambiguities. The interaction of information from multiple sources would enable the perceiver to make accurate hypotheses about the shape of objects in the world.

Previous research indicates LOC is insensitive to semantic information about objects and thus unlikely to be an object recognition area. Both novel and nameable objects activate LOC more than texture patterns (Malach et al., 1995; Kourtzi and Kanwisher, 2000) and may do so equally well (Kanwisher et al., 1996).

Similar studies indicate that LOC is indifferent to low-level image features used to represent shape and object contour. Neural response in LOC appears to be strong when an object is present, regardless of whether its shape is defined by texture, motion (Grill-Spector et al., 1998a), or luminance contours and shading (Malach et al., 1995).

Similarly, we found that LOC responded to volume regardless of whether volume was defined by bottom-up or top-down information. Image properties such as cast shadows, shading gradients, and contour junctions typically encourage a volumetric interpretation of grayscale objects. Although these and all other known image-based depth cues were absent from the two-tone images (Moore and Cavanagh, 1998), when primed by top-down information they were interpreted as volumes. Importantly, neural response in LOC increased when volume was perceived in the grayscale images and when it was perceived in the two-tone images.

These results conflict with one interpretation of a recent study (Kourtzi and Kanwisher, 2000) in which LOC response to line drawings of 3D objects and to outline silhouettes of the same objects was found to be statistically equivalent. The silhouettes themselves contained no interior contours as bottom-up cues to depth; however, the intact objects were presented immediately prior to the silhouettes, possibly providing top-down depth information. The authors concluded LOC was unresponsive to depth information but acknowledged the possibility that the intact objects may have primed recovery of volume from the silhouettes. Our data support the latter interpretation.

Our finding that LOC is responsive to volume in both two-tone and grayscale images of objects indicates this region is responsive to both the knowledge-based and the image-based components of 3D object structure. Furthermore, it suggests that sensitivity to volume in LOC is independent of individual depth cues such as shadows or shading. LOC appears to be the region in which information from multiple sources is integrated to support the subjective experience of perceiving volume.

Experimental Procedures

MR Parameters

In each experiment, seven adults participated in six scans (EPI at 3T, TR = 2.5, FA = 80, TE = 40, 12 4 mm planes perpendicular to the calcarine, FOV = 20 cm, 64 $\times$ 64 voxel images interpolated to 128 $\times$ 128). The scanning session lasted $\sim$90–120 min, including shimming, anatomical, and functional imaging. Stimulus images were presented in 20 s epochs alternating with 20 s epochs of a blank gray screen with a small fixation cross in the center. Images were presented once per second for 250 ms each; the short presentation minimized the opportunity for eye movements. Subjects fixated the cross during the 750 ms intertrial interval and during fixation epochs.

Data Analyses: fMRI

Defining LOC

Data were analyzed using the general linear model implemented in MATLAB (Mathworks, Natick, MA). First, the raw MR data at each voxel was converted to percent change from the scan mean. Percent change data were then regressed onto a sinusoidal model hemodynamic response function for each stimulus/fixation epoch pair. The GLM package estimated response amplitudes (the scale factor that produced the best fit of the model response to the data) and enabled statistical comparison of the amplitudes corresponding to different conditions (Friston et al., 1994). The result was a t statistic map of the reliability of the contrast at each voxel.

The main map used for voxel selection contrasted object versus texture responses. LOC was defined as voxels within an anatomically restricted region (see Figure 3) that produced an object > texture difference with a threshold of $p < 0.05$. At this threshold, 105 $\pm$ 91 voxels per observer were included in the main analyses of experiments 1 and 2. The particular threshold chosen did not affect further analyses; similar results were obtained with $p < 0.005$ (65 $\pm$ 19 voxels/observer) and $p < 0.001$ (50 $\pm$ 32 voxels/observer).

Supplementary Difference Maps

In experiment 1, three additional difference maps were created using the procedure described above: 3D two-tone > 2D two-tone, 3D prime > 2D prime, and texture > object. Using a threshold of $p < 0.05$, $\sim$304 $\pm$ 148 voxels per observer, 203 $\pm$ 153 voxels per observer, and 96 $\pm$ 81 voxels per observer were included in the three analyses.

Defining Retinotopic ROI

In experiment 2, retinotopic areas were identified using the results of scans in which subjects viewed slowly rotating 10° wedges of contrast-reversing checkerboard (Engel et al., 1994; Sereno et al., 1995; DeYoe et al., 1996). Three groups of visual areas were selected for analysis; areas V1 and V2 were analyzed together as one region; V3, V3a, and V7 constituted a second region; and VP, V4v, and V8 a third. Within each retinotopic region, active voxels were chosen using a difference map that compared fMRI responses during object viewing (from the LOC localizer scans) to fixation epochs. The main analyses included only voxels in an object > fixation map with a threshold of $p < 0.05$; $\sim$30% of the voxels in each retinotopically identified region were analyzed. Again, the particular threshold chosen did not affect further analyses; similar results were obtained with a threshold of $p < 0.01$.

Comparing Conditions

Average time courses were computed for LOC and retinotopic regions for each experimental condition. The response amplitude was calculated as the amplitude of the sinusoid that best fit the average fMRI time courses for each epoch of each experimental condition. For each of the seven observers, mean response amplitudes from each pair of conditions (e.g., 3D primes versus 2D primes) were analyzed with paired t tests, implementing a random effects model. When multiple comparisons were made, Bonferroni corrections were used.

Comparing Epochs

First, we tested for a linear trend across sequential epochs in each scan in both experiment 1 and 2 by regressing response amplitude onto epoch number. A significant slope would indicate a linear trend. Second, we tested for nonlinear trends by conducting a single-factor ANOVA on the residuals of the regression. A significant difference in residuals from different epochs would reveal any effects not detected by the regression. The probability of family-wise error due to multiple tests was reduced with a Bonferroni correction.

Data Analyses: Behavior

Behavioral data from each of the two-tone scans in experiment 1 were analyzed using signal detection methodology; volume was the signal to be detected. Observers decided whether a red spot was on the object (Figure 2). An On response when the spot appeared either on the white or black object surface was a hit, and an Off response was a miss. An Off response when the spot appeared either on the white background or the black cast shadow was a correct rejection, and an On response was a false alarm. Previous studies (Moore and Cavanagh, 1998; C. Moore and S. A. Engel, 2001) have shown that when two-tone images are seen as 2D shapes, volume detectability scores are low, and bias scores are high. When volume is perceived, detectability increases, and bias is negligible.

The tasks used in experiment 2 required a yes/no response, making them amenable to signal detection analysis as well. Reaction times were collected for these responses.

Stimuli

In both experiments, the fixation image consisted of a gray screen with a black cross in the center. In experiment 2, the rotating wedge of contrast-reversing checkerboard used to map retinotopic organization in cortex (see Engel et al., 1997) subtended 60° of arc and swept out a circle with a radius of $\sim$5° visual angle. The wedge completed a full rotation in 45 s and reversed contrast at a rate of 8 Hz. All other stimulus images were $\sim 9° \times 9°$ visual angle in size.

The 3D objects were modeled, illuminated, and rendered in VIDI ModelPro and Presenter Professional. Each novel object consisted of one to five simple volumetric parts with gray matte surfaces. Two-tone images were created by thresholding the grayscale images; all pixels above the mean luminance of the image were set to white, and all pixels below to black.

The texture patches used in experiment 1 contained symmetric and asymmetric patterns, large and small texture elements, and various shades of gray. Texture patches were replaced in experiment 2 with scrambled versions of the grayscale objects. The object image was divided into 400 10×10 pixel squares; then the contents of each square were randomly reassigned to another location in the grid. A 20×20 grid of gray lines was placed over both the intact and scrambled images to equate local image features.

Acknowledgments

We appreciate the assistance of Mark Cohen with fMRI. For their support, we also thank John Mazziotta, the UCLA Brain Mapping Medical Organization, the Ahmanson Foundation, the Pierson-Lovelace Foundation, the Tamkin Foundation, the Jennifer Jones-Simon Foundation, and the Whitehall Foundation.

Received July 17, 2000; revised December 5, 2000.

References

Andersen, R.A., Snyder, L.H., Bradley, D.C., and Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annu. Rev. Neurosci. *20*, 303–330.

Bradley, D.C., Chang, G.C., and Andersen, R.A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. Nature *392*, 714–717.

Buckner, R.L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., and Dale, A.M. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. Neuron *20*, 285–296.

DeAngelis, G.C., Cumming, B.G., and Newsome, W.T. (1998). Cortical area MT and the perception of stereoscopic depth. Nature *394*, 677–680.

DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., and Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad. Sci. USA *93*, 2382–2386.

Dolan, R.J., Fink, G.R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R.S., and Friston, K.J. (1997). How the brain learns to see objects and faces in an impoverished context. Nature *389*, 596–599.

Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., and Shadlen, M.N. (1994). fMRI of human visual cortex. Nature *369*, 525.

Engel, S.A., Glover, G.H., and Wandell, B.A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. Cereb. Cortex *7*, 181–192.

Freeman, W.T. (1994). The generic viewpoint assumption in a framework for visual perception. Nature *368*, 542–545.

Friston, K.J., Jezzard, P., and Turner, R. (1994). Analysis of functional MRI time-series. Hum. Brain Mapp. *1*, 153–171.

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., and Malach, R. (1998a). Cue-invariant activation in object-related areas of the human occipital lobe. Neuron *21*, 191–202.

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzchak, Y., and Malach, R. (1998b). A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum. Brain Mapp. *6*, 316–328.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203.

Henson, R., Shallice, T., and Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. Science *287*, 1269–1272.

Hubel, D., and Wiesel, T. (1970). Stereoscopic vision in macaque monkey. Cells sensitive to binocular depth in area 18 of the macaque monkey cortex. Nature *255*, 41–42.

Humphrey, G.K., Goodale, M.A., Bowen, C.V., Gati, J.S., Vilis, T., Rutt, B.K., and Menon, R.S. (1997). Differences in perceived shape from shading correlate with activity in early visual areas. Curr. Biol. *7*, 144–147.

James, T.W., Humphrey, G.K., Gati, J.S., Menon, R.S., and Goodale, M.A. (1999). Repetition priming and the time course of object recognition: an fMRI study. Neuroreport *10*, 1019–1023.

Janssen, P., Vogels, R., and Orban, G.A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. Science *288*, 2054–2056.

Kanwisher, N., Chun, M.M., McDermott, J., and Ledden, P.J. (1996). Functional imagining of human visual recognition. Brain Res. Cognitive Brain Res. *5*, 55–67.

Kanwisher, N., Tong, F., and Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. Cognition *68*, B1–B11.

Knill, D.C., Kersten, D., and Mamassian, P. (1996). Implications of a Bayesian formulation of visual information for processing for psychophysics. In Perception as Bayesian Inference, D.C. Knill and W. Richards, eds. (New York: Cambridge University Press), pp. 239–286.

Kourtzi, Z., and Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. J. Neurosci. *20*, 3310–3318.

Logothetis, N.K., and Sheinberg, D.L. (1996). Visual object recognition. Annu. Rev. Neurosci. *19*, 577–621.

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., and Tootell, R.B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc. Natl. Acad. Sci. USA *92*, 8135–8139.

Mendola, J.D., Dale, A.M., Fischl, B., Liu, A.K., and Tootell, R.B. (1999). The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. J. Neurosci. *19*, 8560–8572.

Mishkin, M., Lewis, M.E., and Ungerleider, L.G. (1982). Equivalence of parieto-preoccipital subareas for visuospatial ability in monkeys. Behav. Brain Res. *6*, 41–55.

Mishkin, M., Ungerleider, L.G., and Macko, K.A. (1983). Object vision and spatial vision: two cortical pathways. Trends Neurosci. *6*, 414–417.

Moore, C., and Cavanagh, P. (1998). Recovery of 3D volume from 2-tone images of novel objects. Cognition *67*, 45–71.

Moore, C. and Engel, S.A. (2001). Mental models change rapidly with implicitly acquired information about the local environment: a 2-tone image study. J. Exp. Psychol. Hum. Percept. Perform., in press.

Nagahama, Y., Takayama, Y., Fukuyama, H., Yamauchi, H., Matsuzaki, S., Magata, Y., Shibasaki, H., and Kimura, J. (1996). Functional anatomy on perception of position and motion in depth. Neuroreport *7*, 1717–1721.

Orban, G.A., Sunaert, S., Todd, J.T., Van Hecke, P., and Marchal, G. (1999). Human cortical regions involved in extracting depth from motion. Neuron *24*, 929–940.

Ptito, A., Zatorre, R.J., Petrides, M., Frey, S., Alivisatos, B., and Evans, A.C. (1993). Localization and lateralization of stereoscopic processing in the human brain. Neuroreport *4*, 1155–1158.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., and Tanaka, Y. (1997). The TINS lecture. The parietal association cortex in depth perception and visual control of hand action. Trends Neurosci. *20*, 350–357.

Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., and Tootell, R.B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science *268*, 889–893.

Tanaka, K. (1996). Inferotemporal cortex and object vision. Annu. Rev. Neurosci. *19*, 109–139.

Tootell, R.B., Dale, A.M., Sereno, M.I., and Malach, R. (1996). New images from human visual cortex. Trends Neurosci. *19*, 481–489.

Ungerleider, L.G. (1995). Functional brain imaging studies of cortical mechanisms for memory. Science *270*, 769–775.