An Area within Human Ventral Cortex Sensitive to "Building" Stimuli: Evidence and Implications

Geoffrey K. Aguirre,* E. Zarahn, and M. D'Esposito
Department of Neurology
University of Pennsylvania
School of Medicine
Philadelphia, Pennsylvania 19104

Summary

Isolated, ventral brain lesions in humans occasionally produce specific impairments in the ability to use landmarks, particularly buildings, for way-finding. Using functional MRI, we tested the hypothesis that there exists a cortical region specialized for the perception of buildings. Across subjects, a region straddling the right lingual sulcus was identified that possessed the functional correlates predicted for a specialized building area. A series of experiments discounted several alternative explanations for the behavior of this site. These results are discussed in terms of their impact upon our understanding of the functional structure of visual processing, disorders of topographical disorientation, and the influence of environmental conditions upon neural organization.

Introduction

There is now considerable evidence that visual object processing is, to some extent, subserved by specialized neuroanatomical subsystems organized by stimulus class. Most notably, recent neuroimaging (Kanwisher et al., 1997; McCarthy et al., 1997) and neuropsychological (Moscovitch et al., 1997) studies have provided decisive evidence for the existence of an area of extrastriate cortex in humans that is specialized for the perception of faces. Particularly compelling is the evidence (1) that patients with localized lesions within the fusiform gyrus can be specifically impaired in the recognition of faces (i.e., prosopagnosic; Damasio et al., 1982; Farah, 1990) and (2) that neurons within this area respond more vigorously to face stimuli than to any other class of stimulus tested (Allison et al., 1994a; Puce et al., 1995, 1997; Kanwisher et al., 1997). Similar evidence, albeit not as complete, has been amassed for the existence of a word recognition area (Warrington and Shallice, 1980; Shallice and Saffran, 1986; Allison et al., 1994b; Polk and Farah, 1998).

Although infrequently recognized as such, there is neuropsychological evidence for the existence of yet another specialized area within extrastriate cortex. This evidence is in the form of a subset of patients who suffer from "topographical disorientation," a heterogeneous set of neuropsychological deficits that follow different, isolated brain lesions (Aguirre et al., 1998a). One variety of this disorder has been termed "landmark agnosia" (Whiteley and Warrington, 1978; Levine et al., 1985), as these patients, most typically following dextral lesions

*To whom correspondence should be addressed.

of the medial occipital lobe, seem to have particular difficulty using salient environmental features for way-finding. The limited neuropsychological testing that has been performed upon these patients suggests that they are primarily impaired in the perception and recognition of street scenes, landscapes, monuments, and, most notably, buildings (Hécaen et al., 1980; Takahashi et al., 1989; McCarthy et al., 1996; Rocchetta et al., 1996). Given that other higher order perceptual disorders, such as general object agnosia and prosopagnosia, can occur without landmark agnosia (Tohgi et al., 1994), and given that patients with landmark agnosia do not uniformly suffer from these more general impairments as well (Hécaen et al., 1980), it is possible to argue that damage to a cortical substrate separate from that damaged in prosopagnosia and general object agnosia is responsible for this variety of topographical disorientation. Additional evidence for the separability of landmark agnosia and general object agnosia is the compensatory strategy that landmark agnosics employ to way-find following their brain damage. Commonly, these patients report relying upon less salient environmental features (i.e., distinctive doorknobs, mail boxes, park benches) to learn and follow a path.

While other explanations certainly might be offered, a facile account of landmark agnosia is that these patients have sustained damage to a cortical region that is specialized for the perception of visual stimuli with orienting value. Given that impaired recognition of buildings is frequently described by landmark agnosics, and given that buildings as a class of objects have very high landmark value (Lynch, 1960), any account of such a "landmark" region would require that buildings be among the visual objects for which the region is specialized. An area that is specialized for the perception of buildings might be expected to respond with the greatest intensity of neuronal firing to building stimuli, in a manner analogous to that observed for face and word areas. Such selectivity of response should be observable using neuroimaging methods, thus offering the ability to test the hypothesis of a specialized landmark area.

In a series of functional magnetic resonance imaging (fMRI) studies described here, we attempted to identify in normal subjects an area of ventral occipitotemporal cortex that could be specialized for the perception of buildings. We wished, in particular, to rule out several alternative explanations that might be offered for apparent building selectivity. The general approach was to identify candidate regions during an initial study and then repeatedly probe the activity of the area in further studies. Experiment 1 identified candidate building areas in which fMRI signal was greater during the perception and recognition of buildings than during the perception and recognition of faces or general inanimate objects. The regions identified in Experiment 1 were then further investigated during Experiments 2 and 3. During Experiment 2, subjects passively viewed buildings, cars, and several processed versions of the building stimuli. The purpose of Experiment 2 was to test alternate hypotheses based upon low level visual feature
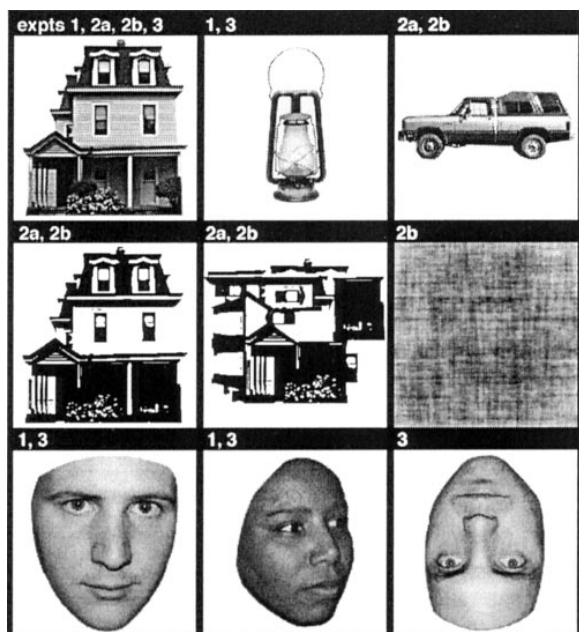
Figure 1. Examples of Stimuli Used in the Three Experiments

Table 1. Mean d′ and RT Scores across Subjects for Each Stimulus Category from Experiment 1

Stimulus	d′	RT (ms)
Faces	3.36	1057
Buildings	3.63	930
Objects	3.68	891

Post hoc tests revealed that d′ values for faces were significantly lower as compared with both buildings and general objects, and that reaction times to faces were significantly slower as compared with both buildings and general objects. The d′ values shown here are lower than the average d′ value for the entire experiment reported in the text. This is because the magnitude of the d′ measure is affected by the number of observations obtained. The d′ measures broken down by category here have one-third the number of observations of the d′ measure obtained for the experiment as a whole.

confounds, hypotheses based upon subordinate categorical membership (Gauthier et al., 1997), and hypotheses based upon passive versus active viewing of stimuli. Experiment 3, as opposed to Experiments 1 and 2, was designed as an event-related fMRI experiment, the purpose of which was to determine if interactions of stimulus type and stimulus order (i.e., the blocked presentation of buildings used in Experiments 1 and 2) could account for the activity within the candidate region (Zarahn et al., 1997b).

Across subjects, a region in the anterior aspect of the right lingual gyrus was identified that possessed the functional correlates predicted for a specialized building area. These results are discussed in terms of their impact upon our understanding of the functional structure of visual processing, disorders of topographical disorientation, and the influence of environmental conditions upon neural organization.

Results

Experiment 1

The purpose of Experiment 1 was to identify regions to be studied in further experiments. During fMRI scanning, subjects performed a visual object recognition task in which 30 s blocks of stimuli from a given category (i.e., faces, buildings, and general objects) were presented together. Candidate regions were required to (1) respond more to buildings than to inanimate objects in general, biasing against areas that respond equally to all objects, and (2) respond more to buildings than to faces, biasing against areas that respond only to subordinate level classification (Gauthier et al., 1997). Example stimuli are shown in Figure 1.

All nine subjects performed the detection task well above chance (mean d′ ± SD = 4.24 ± 0.32). There was a significant effect of stimulus category upon detection

scores [ANOVA $F_{(2, 16)} = 3.9$, $p = 0.042$] and reaction time (RT) [ANOVA $F_{(2, 16)} = 9.9$, $p = 0.002$]. Post hoc tests indicated that the effect was the result of face recognition being slightly, but significantly, more difficult (as indexed by lower d′ scores and slower RTs; see Table 1).

Three initial subjects were studied during Experiment 1 (these subjects did not participate in subsequent experiments). In these subjects, bilateral, ventral extrastriate cortex was searched for voxels with significantly greater fMRI signal during the presentation of buildings compared with either general objects or faces. Two of the three subjects (S1 and S2) had voxels that passed both tests (see Figure 2). Both S1 and S2 had building-sensitive voxels straddling the anterior end of the right lingual sulcus, just posterior to the parahippocampus. S2 also had voxels in a homologous position on the left. The anatomical position of these voxels was determined by reference to the location of the collateral and lingual sulci. S1 and S2 also possessed voxels with significantly greater responses to faces than to buildings or objects. In both subjects, these voxels were located on the right, in the fusiform gyrus.

The right, lingual gyrus site was found in both of the initial subjects who evinced any significant signal changes. As this region accorded with our a priori notions regarding the possible location of building-sensitive regions (based upon case reports of topographical disorientation), we chose to define selective regions of interest in this general area for examination in the six subsequent subjects. These regions averaged ∼200 voxels in volume and included the parahippocampus, the superior and medial portions of the fusiform gyrus, and the anterior portions of the inferior and superior lingual gyri on the right. The locations of these regions are shown in Figure 2.

Five of the subsequent six subjects studied possessed voxels within the focused region of interest that responded with greater fMRI signal to buildings as compared with faces and general objects. These voxels were very similar in location to those observed in the preliminary subjects: generally close to or upon the anterior lingual sulcus and 5–15 mm posterior to the parahippocampus. The voxels identified for subject S8, however, were located in a slightly more anterior position within the collateral sulcus. The coordinates (in Talairach space) of the unweighted centroids of the region in the second
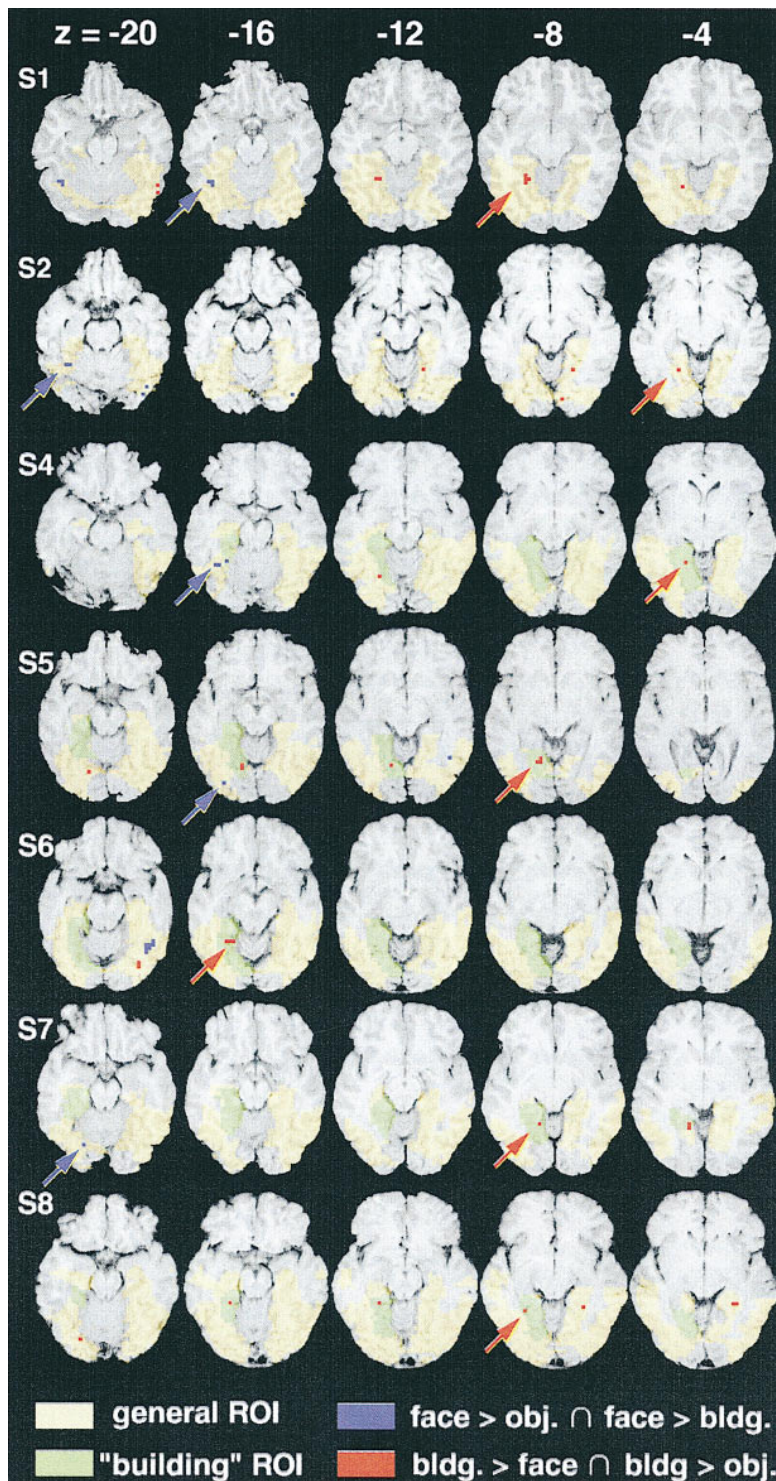
Figure 2. Results from Experiment 1

Shown in gray are axial brain slices in standard (Talairach) space, arranged inferior to superior for seven subjects. Images are displayed using the radiologic convention (left is on the right). The yellow overlay indicates the intersection of the region of interest, defined a priori to include ventral extrastriate cortex, and the locations where adequate fMRI signal was present to test hypotheses. The green overlay is a restricted region of interest guided by the initial subjects who performed Experiment 1 (S1 and S2) and our a priori hypotheses derived from the topographical disorientation literature. As the regions were defined upon the anatomical images collected for each subject in their native spatial frame, the conversion to standard space results in a slightly irregular appearance of the mask. Shown in red are those voxels that evidenced significantly greater fMRI signal during the presentation of buildings as compared with faces and general objects. Shown in blue are voxels that passed analogous tests for faces.

set of subjects were obtained. The mean ($\pm$ SD) location of this candidate building region was (x, y, z): 20.6 $\pm$ 5.0, -53.8 ± 6.7, -9.2 ± 3.9. Figure 3 provides a magnified view of the anatomical location of this area for three subjects. The average percentage signal change for the different stimulus classes within these voxels was calculated across subjects (buildings, 3.29; general objects, 0.29; and faces, -0.16).

Extrastriate cortical areas outside of the focused ROI were also examined for significant signal changes, albeit with reduced statistical power (due to a more stringent statistical threshold required for the larger search region). This was done (1) to identify areas that responded with greater signal change to building stimuli outside of our region of interest and (2) to search for areas with greater responses to faces or general objects. Several subjects (S5, S6, and S8) possessed additional voxels that responded significantly more to buildings than the
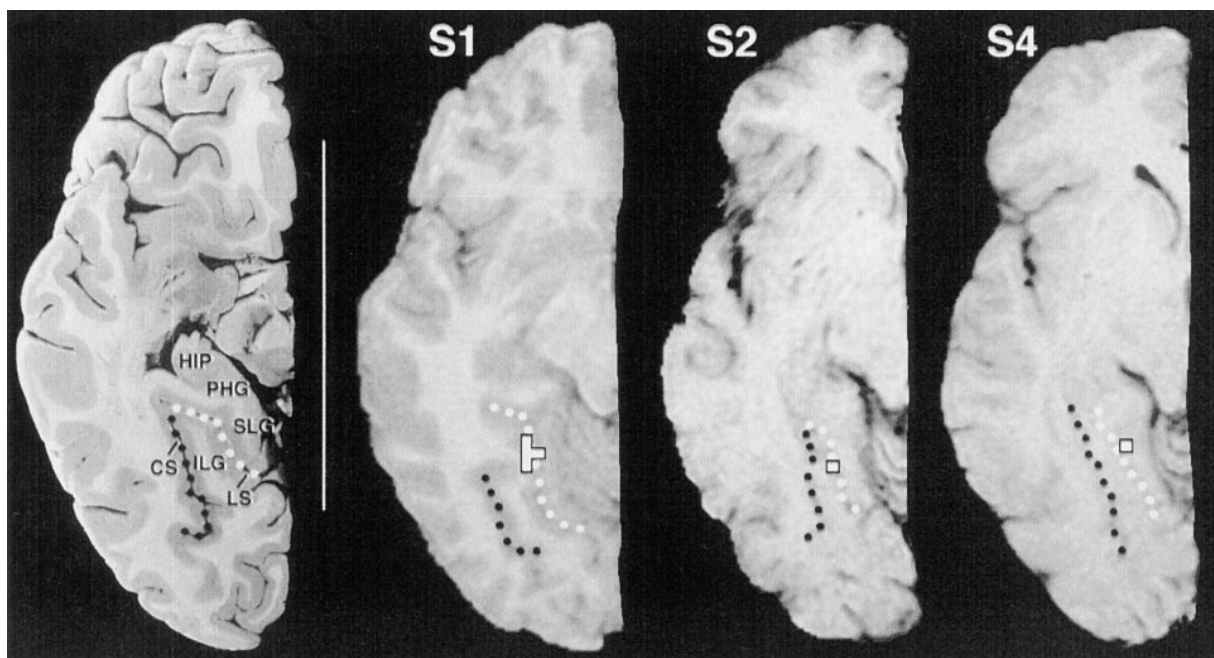
Figure 3. Anatomic Location of Building-Sensitive Voxels

Shown on the left is an axial tissue section of the ventral human brain (adapted from Duvernoy, 1991). Shown on the right are data from three different subjects from Experiment 1. These subjects were selected because of the relative clarity of the sulcal structures in their T1 MRI images. Abbreviations: HIP, hippocampus; PHG, parahippocampal gyrus; SLG, superior lingual gyrus; ILG, inferior lingual gyrus; CC, collateral sulcus; and LC, lingual sulcus.

other stimulus categories, but the locations of these voxels were not consistent across subjects. Several subjects (S4, S5, S6, and S7) also possessed voxels that responded with significantly greater signal to faces than to either buildings or general objects. For three of these subjects, the suprathreshold voxels were on the right in the fusiform cortex, inferior and lateral to the building-sensitive voxels. A fourth subject (S6) had face-sensitive voxels only on the left side. The mean ($\pm$ SD) location of the three face-sensitive regions on the right was (x, y, z): 29.0 $\pm$ 5.6, -70.0 ± 19.0, -17.3 ± 2.3. Interestingly, no voxels were discerned in any subject, either within the focused or within the expanded regions of interest, in which the response to general objects was significantly greater than the response to both buildings and faces.

Experiment 2

During Experiment 1, regions were identified in which fMRI signal was greater during the presentation of building stimuli compared with the presentation of general object or face stimuli. In Experiment 2, we wished to test a number of alternative explanations for the observed responses to building stimuli. During scanning, subjects passively viewed stimuli, in 30 s blocks, from the following categories: (1) gray-scale buildings, (2) thresholded (black-and-white) buildings, (3) scrambled thresholded buildings, and (4) gray-scale cars. Two subjects also viewed stimuli derived from phase-randomized pictures of buildings, in addition to those just listed. The fMRI signal obtained during Experiment 2 was averaged within the building-sensitive regions identified in Experiment 1 to create a single time series for each subject.

Several alternative hypotheses were tested by examining the magnitude of this signal under different stimulus conditions. As Experiment 2 was a passive viewing task, no behavioral measures were obtained.

The building-sensitive areas identified during Experiment 1 were found to have significantly greater signal during the perception of gray-scale buildings relative to cars for three of the five subjects (see Table 2). The two subjects who participated in Experiment 2b (in which random phase buildings were presented) also possessed significantly greater fMRI signal during the presentation of gray-scale buildings compared with the phase-randomized stimuli within this region. Finally, for three of the five subjects, the interrogated region responded with a greater signal change to thresholded (black and white) pictures of buildings compared with scrambled versions of these stimuli.

The parameters estimated from the time series data collected from each subject were entered into a mixed-effect group model to determine if, across subjects, gray-scale buildings evoked significantly greater signal than cars and if thresholded buildings produced a greater signal change than these same stimuli scrambled. This test was not conducted for the comparison versus phase-randomized stimuli as, with only two subjects, insufficient degrees of freedom (df) were present to conduct a reasonably sensitive test. Across subjects, the presentation of gray-scale buildings was found to evoke greater signal compared with gray-scale cars [$t(4 \text{ df}) = 5.17$, $p = 0.003$]. Thresholded buildings also evoked significantly greater signal compared with scrambled stimuli [$t(4 \text{ df}) = 2.56$, $p = 0.031$]. The average

Table 2. Results of Within-Subject (Fixed-Effect) and Across-Subject (Mixed-Effect) Statistical Tests for Experiments 2 and 3

Contrast	S4	S5	S6	S7	S8	Population
Experiment 2						
Building versus Car	1.40	**2.02**	**2.26**	**2.95**	1.39	**5.17**
Threshold versus Scrambled	**3.36**	**5.49**	**4.80**	−0.87	1.11	**2.56**
Building versus Random	—	—	—	**3.70**	**4.09**	—
Experiment 3						
Building versus Object	**2.09**	**3.73**	**2.67**	−0.03	**1.85**	**2.93**
Building versus Face	0.31	**4.78**	**2.07**	0.50	**1.79**	**2.92**

Presented are t-values, with significant results in bold. Experiment 2, within subjects, had 121 effective degrees of freedom (eff df). Experiment 3, within subjects, had 1087 eff df. Tests for the group (mixed-effect model) for both experiments had 4 df.

percentage change values obtained for the covariates that modeled different stimulus categories relative to fixation across subjects were obtained as well (buildings, 4.06; thresholded buildings, 3.50; cars, 2.27; scrambled buildings, 1.45; and phase-randomized: −0.22).

Experiment 3

In Experiment 3, we sought to replicate the finding of greater responses to buildings as compared with faces and general objects within the context of a randomized, event-related fMRI design. As in Experiment 1, Experiment 3 was conducted as a detection task.

All five subjects performed well above chance (mean $d' \pm SD = 4.28 \pm 0.30$). This performance level was not significantly different from that observed during Experiment 1 [$t(4\ df) = 0.75$, $p = NS$]. The regions identified in Experiment 1 were interrogated for the building versus object and building versus face contrasts.

The signal obtained for subjects S5, S6, and S8 displayed significantly greater evoked responses to buildings compared with the other stimulus types (Figure 4 and Table 2). As can be seen, the response to the presentation of building stimuli appears similar in shape to the fMRI hemodynamic response function that has been observed in response to brief neural events (Aguirre et al., 1998b). This shape of response is thus consistent with the proposal that building stimuli evoke transient increases in neural firing in this area. A mixed-effect model was used to test the hypothesis of greater responses to building stimuli in this region across subjects. Buildings were found to evoke significantly greater fMRI signal when compared both with general objects [$t(4\ df) = 2.93$, $p = 0.021$] and with faces [$t(4\ df) = 2.92$, $p = 0.022$].

A final analysis was conducted within subjects to account for the possibility of misregistration of the functional data across scanning sessions (see Experimental Procedures for details and rationale). This post hoc analysis was conducted in subjects S4 and S7, who did not evidence significantly greater responses to buildings within the target region. In subject S7, a region was found one voxel anterior to the original search region that displayed the responses shown in Figure 4. The response to buildings was significantly greater than the response to the other stimuli after accounting for the number of voxels examined in the post hoc test. Subject S4 did not have any significant evoked responses within the searched region.

Discussion

The series of experiments presented here were motivated by our consideration of the case literature of topographical disorientation. Over the last century, several dozen reports have presented patients who seem to have selectively lost the ability to find their way within their locomotor environment. These patients are, however, rather heterogeneous, and closer inspection of the cases reveals that different groups of patients possess substantially different underlying impairments. This might be expected, given that way-finding is a complex behavior and that any one of many different underlying cognitive impairments might lead to an overt inability to travel from place to place. In particular, lesions to the ventral occipitotemporal cortex (e.g., Landis et al., 1986; McCarthy et al., 1996) have been suggested to result in a variant of disorientation termed landmark agnosia, in which the patient is unable to use salient environmental features for the purposes of orientation but evidences spared representations of the spatial arrangement of places. In particular, these patients are greatly impaired at recognition tests involving premorbidly familiar buildings.

How can an isolated lesion produce an inability to recognize and make use of landmarks in the absence of other object recognition deficits? The notion that was tested in this set of experiments is that there is an area of ventral cortex that is specialized for the perception of stimuli with orienting value. Sufficiently isolated damage to this area would impair the ability of the patient to make use of a class of high salience environmental features that are valuable for way-finding. This would produce the observed deficits of landmark agnosia while sparing general object recognition and face perception. We proposed that if such a region exists, it should display relatively selective responses to building stimuli, given the prominent role that buildings play in navigation in urban environments (Lynch, 1960). Thus, the landmark area hypothesis would be supported by the identification of a cortical area that responds maximally to buildings.

Tests of the Hypothesis

The experiments conducted here had two primary goals: first, to identify candidate building regions and assess the spatial concordance of such regions both across subjects and with respect to the lesion literature; second, to test a number of alternative hypotheses that
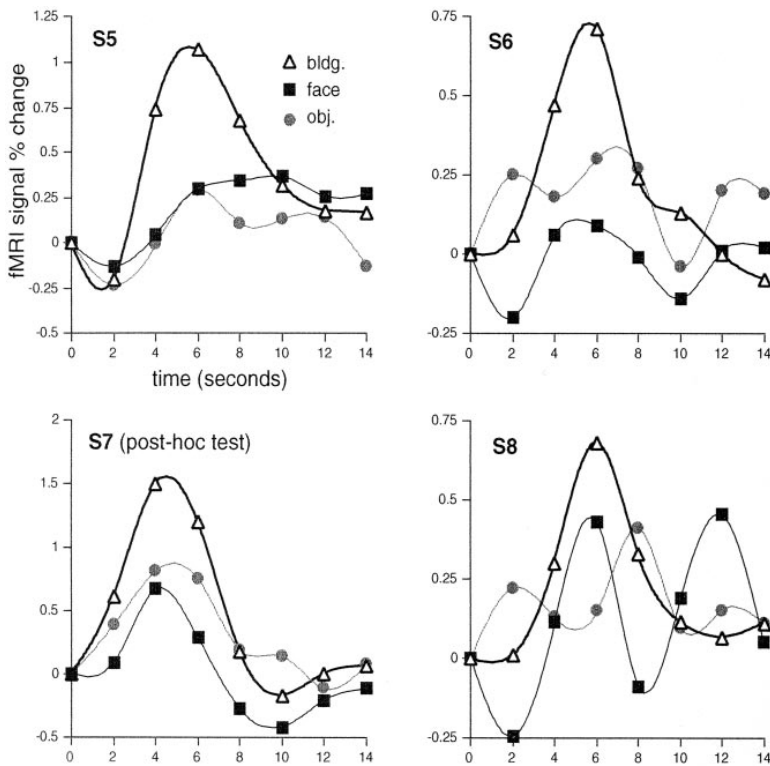
Figure 4. Results from Experiment 3

Stimuli were presented in a random order every 16 s. The plots show the trial averaged signal (with nuisance effects removed; see Experimental Procedures) evoked for different stimulus categories for four different subjects. The smooth evoked response to buildings (thick black line) approximates the shape of the hemodynamic response of the BOLD fMRI system (Aguirre et al., 1998b) and is thus likely the result of a brief period of neural activity following the presentation of the building stimuli. Note that the plots for S7 are not from the original region defined in Experiment 1 but from a region one voxel anterior to this original area, which was identified by a post hoc test (see Results and Experimental Procedures).

might explain activity in the area aside from the possibility that it is simply most responsive to buildings. Experiment 1, which searched for putative building regions both within bilateral, ventral extrastriate cortex and a more focused dextral region of interest, identified candidate voxels in seven of nine subjects. The location of these activations, near the anterior portion of the right lingual sulcus, is in good agreement with the lesion sites that have been reported to produce landmark agnosia. In addition, the location of this site was rather consistent across subjects, both in terms of gross neuroanatomical landmarks (i.e., the collateral and lingual sulcus) and standard Talairach coordinates.

Several other aspects of Experiment 1 are noteworthy. First, separate cortical regions with greater responses to faces than to either buildings or general objects were observed in six of the nine subjects. The existence of separate but anatomically proximal regions responsive to face and building perception was hypothesized based upon the topographical disorientation literature. These face-sensitive regions were located within the fusiform gyrus, inferior and lateral to the building-sensitive voxels, in rough agreement with previous studies (Kanwisher et al., 1997; McCarthy et al., 1997) (although variability between subjects in precise location is evident here as elsewhere; see Kanwisher et al., 1997). Notably, the simultaneous observation of building-sensitive and face-sensitive regions (a neuroimaging double dissociation) makes several alternative explanations for the findings rather unlikely. For example, it would be difficult to argue that the building area is only responding to within-category discriminations (Gauthier et al., 1997), as such a proposed cognitive process would seem to apply equally to face perception. It is also important to note that subject accuracy was lower and reaction times were slower

for recognition judgments of faces compared with buildings. This finding argues against an alternative account in which activation in the building area is due to the increased difficulty of discriminating building stimuli. Finally, the use of a varied stimulus set of inanimate objects allows us to discredit the proposal that the responses of the building-sensitive region result from greater variety or novelty within the building stimulus set as compared with other stimulus sets.

Experiment 2 was designed to reject a number of additional alternatives. One might claim, for example, that the responses of the building-sensitive region are in fact driven by low level visual features that are confounded with buildings as a stimulus class. By comparing the presentation of two-tone buildings to scrambled versions of these stimuli, we were able to test the hypothesis that the region is simply responsive to particular textures that are present in building stimuli. Also, comparison of gray-scale buildings to these same stimuli phase-randomized tested the notion that the region is responsive only to spatial frequencies that are present within building stimuli. Both of these alternative accounts were rejected. More sophisticated alternative accounts were tested by comparing gray-scale buildings to a set of car stimuli. Cars provide a useful stimulus set to compare against buildings for several reasons. First, cars are large, inanimate objects found in the same visual environment as are buildings and, second, they constitute a class with equal face validity as do buildings. Importantly, however, cars are poor landmarks as they rarely maintain a constant position with respect to way-points. Thus, equal responses to cars and buildings would admit the possibility of several different alternative accounts: the area might be involved in within-category discriminations of inanimate objects, or simply

respond to large objects, or represent all external environment features. The evoked fMRI signal in response to buildings was nearly twice as large as that observed in response to cars, consistent with the assertion that the area is primarily organized for the representation of landmarks and/or buildings. Notably, however, cars did evoke a response relative to fixation that was larger than any other stimulus class besides buildings. We comment on this finding below.

Experiment 3 was designed to eliminate an entire class of confounds—those that are associated with the blocked order of trial stimulus presentation used in Experiments 1 and 2 (Zarahn et al., 1997b). In particular, the approach removes the possibility of long duration behaviors (e.g., general arousal, anticipation, boredom) associated with particular stimulus sets. Additionally, the experiment was an independent replication of the finding of greater activity in response to buildings in this area compared with faces and general objects. The finding that buildings evoke greater fMRI signal as compared with faces and general objects under this setting further strengthens the assertion that the responsiveness of the region is driven by the building stimuli themselves.

The Implications of a Building Region

Neuroimaging experiments can never prove the involvement of a cortical region in a given cognitive process (Aguirre et al., 1998a). Thus, it cannot be claimed absolutely, based on these or any other imaging tests, that the region under study is necessary for the perception of buildings (or landmarks in general, for that matter). Furthermore, it will never be possible to demonstrate completely that the region responds most strongly to buildings, given practical limitations on the number of stimuli that may be tested. Nonetheless, we examined and rejected a number of reasonable alternatives to the proposal that the identified area is maximally sensitive to the perception of buildings. Further evidence for the building selectivity of this area can be found in the preliminary report of Ishai and colleagues (1997, Neuroimage, abstract), who observed that an area of cortex close to the collateral sulcus demonstrated greater fMRI signal change in response to buildings as compared with faces and chairs. We therefore provisionally accept that this region responds disproportionately to buildings themselves, as opposed to some simple confound of the stimulus class. We suggest further that, even if we were to ultimately find another stimulus class to which the area responds with equal or greater firing, the peculiar selectivity of response demonstrated here deserves attention.

How might a consistently localized cortical area, with selective responses to a particular class of stimuli, in this case buildings, come to be? For some functions (e.g., motion perception), it is possible to propose that the location and behavior of a specialized region is dictated by genetics. This is because such functions are evolutionarily old and are shared with other species (Polk and Farah, 1998). However, some functional specializations, because of their evolutionary recency and cultural variability, cannot be explained as innate and

genetically preprogrammed (e.g., letter areas; Warrington and Shallice, 1980; Shallice and Saffran, 1986; Allison et al., 1994b; Polk and Farah, 1998). Instead, such functional areas might be the result of the organizing effect of Hebbian learning upon a plastic area of cortex (Polk and Farah, 1995). Such a mechanism could explain the existence of a building-sensitive region, given that buildings tend to cooccur in the environment and are clustered together repeatedly and extensively in urban areas (Lynch, 1960). This also accounts for the observation of nonzero responses to pictures of cars in the studied region, as cars tend to frequently cooccur with buildings in urban environments.

Unexplained, however, is the consistent anatomical location of this region across subjects. If we assume an initially equipotent, callow area of cortex, then the final location of a specialized area created by Hebbian learning within that expanse should be a function of random fluctuations in initial connection strengths: a stochastic process (Polk and Farah, 1995). In other words, if we find an area responsive to buildings in a roughly consistent location across subjects, then there must be a reason why that particular area tends to develop building representations as opposed to, for example, letter representations. What might predispose this area of the right lingual gyrus to develop selective responses to buildings? The explanation that we advance here is that the region is predisposed, by dint of its connections to other cortical areas, to develop representations of stimuli commonly used for the purposes of orientation. This category might include large natural terrain features, hallways and rooms, and buildings. These classes of stimuli could be preferentially represented because, for example, they behave in distinctive ways within optic flow fields, or because they are a class of objects typically seen from a restricted number of viewing angles. While the current study certainly does not prove this supposition, as only building stimuli were examined, it was capable of refuting the proposal. Further imaging experiments will be able to test other predictions of this hypothesis, including the prospect that subjects raised in rural environments will have a different set of functional responses within this region compared with subjects raised in urban environments.

One might object to the possibility of a cortical region specialized for the perception of landmarks with the observation that any sufficiently distinctive object can serve as a landmark. While this is true, it does not therefore follow that an area specialized for stimuli frequently used for the purposes of orientation could not exist. Consider that, while almost any symbol could be used as a character in a written language, only a subset of all possible symbols are used for this purpose. A cortical area specialized for perception of letters of the alphabet would develop because a particular set of symbols are encountered frequently and in a particular context. In a similar fashion, while almost any object might be used as a landmark, if there exists a subset of stimuli that are more commonly regarded for this purpose, and that tend to be encountered in similar contexts, then the proper conditions exist for cortical specialization guided by Hebbian learning.

It is noteworthy that the area identified here is adjacent

to the parahippocampus. Several neuroimaging studies have been taken as evidence that the posterior portion of the parahippocampal gyrus is involved in the representation of large-scale place (Aguirre et al., 1996; Aguirre and D'Esposito, 1997; Maguire et al., 1996, 1998; Epstein and Kanwisher, 1998; reviewed by Aguirre et al., 1998a). These studies have used either virtual reality navigation tasks or the presentation of static or moving images of places to evoke activity in this area. It should be noted however, that spatial smoothing was used in all of these studies, and in some cases the results were obtained through intrasubject averaging. These methodological considerations make precise localization difficult, especially as the anatomic boundary between the superior lingual gyrus and the parahippocampus is ill defined (Duvernoy, 1991). As a result, we cannot determine if the region we have studied here overlaps entirely, partially, or not at all with the regions reported in these previous studies. Further experiments, in which the behavioral paradigms of the current and previous studies are examined in the same subject, will be necessary to resolve this ambiguity.

Within this context, the recent study of Epstein and Kanwisher (1998) is particularly relevant. The authors observed greater responses to indoor and outdoor scenes compared with ingenious and appropriate control stimuli within a bilateral region identified as the parahippocampus. What is the relationship of the region identified by Epstein and Kanwisher to that studied here? Two observations are noteworthy. First, examination of the anatomical images provided in that report suggest that the position of the activation is fairly posterior, judging from the width of the cerebellum and the presence of the atrium of the lateral ventricle in some of the coronal slices shown. At this posterior position, there is no clear border between the parahippocampal and superior lingual gyri (Duvernoy, 1991). It is thus possible that the region identified by Epstein and Kanwisher is in the same anatomical position as that studied here. Second, the region identified by Epstein and Kanwisher was found to respond to isolated pictures of buildings with a magnitude of fMRI signal change comparable to that observed in response to scenes. Thus, the functional behavior of the region studied by Epstein and Kanwisher is similar, on at least one count, to that studied here. It is therefore reasonable to propose that the two regions are one and the same.

Epstein and Kanwisher interpreted the functional behavior they observed as indicating that the region responds to the perception of the layout of local space and dubbed the region "the parahippocampal place area". The responsiveness of the region to isolated building stimuli, however, would seem to be in conflict with this account, as a single building does not by itself define an extended space any more than does a single coffee cup (or any other object). Furthermore, as the buildings were unfamiliar, one cannot argue that the subjects were prompted by the stimuli to recall familiar places. Epstein and Kanwisher explained the building responses by noting that buildings "play an important role in defining the geometry of local space." A reasonable modification of this account, as proposed above, is that the region responds to stimuli that have orienting

value (i.e., isolated landmarks as well as "scenes" of the environment), as opposed to space per se. Further neuroimaging and patient studies may prove the appellation "lingual landmark area" to be a more apt label for the region under study.

Experimental Procedures

MRI Acquisition
Imaging was carried out on a 1.5T SIGNA scanner (GE Medical Systems) equipped with a fast gradient system for echoplanar imaging. A standard radiofrequency head coil was used with foam padding to comfortably restrict head motion. High resolution sagittal and axial T1-weighted images were obtained in every subject. A gradient echo, echoplanar sequence was used to acquire data sensitive to the blood oxygenation level–dependent (BOLD) signal at a TR of 2000 ms and a TE of 50 ms. Resolution was 3.75 mm × 3.75 mm in plane and 4 mm through plane, with no skip in between planes. Ten axial slices were acquired in all subjects. These images were positioned to cover the ventral extrastriate cortex completely, ranging from below the most inferior extent of the temporal lobes to the superior aspect of the striate cortex. Twenty seconds of dummy gradient and RF pulses preceded the actual data acquisition to approach steady state magnetization.

Subjects viewed a backlit projection screen from within the magnet bore through a mirror mounted on the head coil. For those tasks in which the subject was required to indicate a detection, the subject made a bilateral button press with both thumbs on a fiber optic game pad.

General Data Processing
Offline data processing was performed on SUN Sparc workstations using programs written in Interactive Data Language (Research Systems, Boulder, CO). After image reconstruction, the data were sinc interpolated in time to correct for the fMRI acquisition sequence. The data were then motion corrected. First, a six parameter, rigid body, least squares realignment routine was used (part of SPM96b package; Friston et al., 1995b). Next, a slice-wise motion compensation method was utilized, which removed spatially coherent signal changes via the application of a partial correlation method to each slice in time (Zarahn et al., 1997a). Data from subjects who moved over 5 mm during the course of any scan were discarded. The data were not smoothed in space, as maximum anatomical resolution was desired. The univariate statistical analyses employed are described below for each experiment.

A six-parameter coregistration algorithm (Friston et al., 1995b), guided by the anatomical T1-weighted images, was used to transfer regions of interest defined during Experiment 1 to the data obtained during other scanning sessions for a given subject. For presentation purposes, the regions investigated for each subject were converted to a standard (Talairach and Tournoux, 1988) spatial frame using a least squares, 12 parameter transformation with nonlinear deformations (part of SPM96b) guided by the anatomical T1s.

Subjects
A total of 12 subjects, naive to the hypotheses of the study, participated in at least one portion of the protocol. Two subjects were rejected for motion during scanning. One subject participated in Experiment 3 but refused to participate in the additional scans required for Experiments 1 and 2. The remaining nine subjects (five female, all right-handed; mean age, 27; range, 18–38) are reported here. Three of these subjects participated in only part 1 of the study. The remaining six performed the three experiments over the course of two different scanning sessions separated by several days to weeks. Experiment 3 was conducted prior to, and on a different day from, Experiments 1 and 2 in all subjects (S4–S8) except S9. The order of Experiments 1 and 2 during the scanning session was varied across subjects. All subjects provided informed consent.

Stimuli
All experiments made use of sets of gray-scale (8 bit) pictures, 256 × 256 pixels in resolution. Sets of 50 photos of faces, houses, cars,

and general small, manipulatable inanimate objects were obtained. Face pictures were obtained in both full and three-quarter view (45° angle) orientations. The models for the face photographs were faculty and staff from the University of Pennsylvania Department of Neurology and volunteers from the hospital cafeteria. The racial composition of the face stimulus set reflects that of the general University community. A portable digital camera was used to take pictures of cars and buildings from the urban Philadelphia area and its residential and commercial environs. These photos were edited to remove background details. The face stimuli were edited to remove details exterior to the face proper, including the ears and details above the hairline and below the chin. A set of 25 inverted face stimuli (from an entirely different set of models) were also obtained. (Inverted face stimuli were included to test hypotheses not addressed in this paper.) Stimulus examples are provided in Figure 1. Additional, derivative sets of stimuli were generated for use in Experiment 2 and are described in detail below.

Behavioral Tasks and Statistical Analysis
Experiment 1
Two fMRI scans, each 5 min in duration (i.e., 300 fMRI observations per voxel), were obtained during this experiment. Prior to each scan, subjects studied one target picture from each of three categories: faces, buildings, and general inanimate objects. The pictures to be learned were shown in order twice for 15 s per picture per presentation. Subjects were instructed to study and remember the pictures for a subsequent memory test. During scanning, subjects viewed blocks of 10 pictures from each category, with each picture presented for 3000 ms. If the subject detected a picture that matched one studied, she was to make a bilateral button press. If the picture was new, the subject was instructed not to make any overt response. Targets were infrequent (11% of stimuli). Neither targets nor distractors were repeated within an experiment. Full and three-quarter view faces were randomly intermixed. The order of the three stimulus blocks and a fixation condition was fixed within subjects but varied across subjects.

Three subjects participated in only Experiment 1. For the analysis of the data from these preliminary subjects, regions of interest were defined that included all of inferior extrastriate cortex bilaterally. These regions contained ~1000 voxels. To identify candidate building regions, two contrasts were evaluated: (buildings − faces) and (buildings − general objects). For a voxel to be retained, it had to surpass a t-value corresponding to an $\alpha = 0.05$ for both comparisons, Bonferroni corrected for the number of voxels within the search area. These t-values were on the order of 4.0. The results of a preliminary analysis of the data from a subset of the subjects who participated in Experiment 1 have been described previously (Aguirre et al., 1998a).

Using the results from the preliminary subjects as a guide, more restrictive regions of interest were defined to constrain the search area and thus improve sensitivity. These regions averaged ~200 voxels in volume and included the parahippocampus, the superior and medial portions of the fusiform gyrus, and the anterior portions of the inferior and superior lingual gyri on the right. As above, retained voxels had to demonstrate significantly greater responses to buildings for both contrasts, evaluated at a t-value corresponding to a region-wise $\alpha = 0.05$ for each comparison, Bonferroni corrected for the number of voxels within the search area. These t-values were ~3.6.

Exploratory regions of interest were also created to cover the remaining extrastriate cortex. These regions were searched for building-sensitive, face-sensitive, and object-sensitive voxels. The regions contained ~1000 voxels and the critical t-value for each comparison was ~4.0.

Data were analyzed using a general linear model for serially correlated error terms (Worsley and Friston, 1995), modified to accommodate the null hypothesis distribution of power observed in our laboratory (1/f model; Aguirre et al., 1997; Zarahn et al., 1997a). An empirically derived (Zarahn et al., 1997a) hemodynamic transfer function was used to smooth both the temporal data and a boxcar model of idealized neural activity. Global signal covariates were also included as the global signal across subjects was found not to be

significantly influenced by the task design [as evaluated by a mixed-effect model test, as described below; all contrasts, t(8 df, two-tailed) < 1.56, p = NS] (Aguirre et al., 1997). Sine and cosine regressors were used to remove frequencies below that of the task. This analysis has been empirically demonstrated to hold the map-wise false positive rate at or below tabular values (Zarahn et al., 1997a).

Experiment 2
Two versions of Experiment 2 were conducted. All six subjects performed Experiment 2a, while two subjects of the six also performed Experiment 2b. Additional sets of stimuli were derived from the building stimulus set for Experiment 2a. First, two-tone (i.e., thresholded black-and-white) versions of the building stimuli were produced. Next, scrambled two-tone buildings were produced by rearranging the components of the two-tone images to produce stimuli that were no longer recognizable as buildings (see Figure 1). During the two fMRI scans (375 fMRI observations per voxel), subjects passively viewed these stimuli in 30 s blocks (as in Experiment 1) in addition to sets of gray-scale building and car stimuli, as well as a fixation condition. The order of these blocks was fixed within subjects and varied across subjects. An additional condition was added to this design for two subjects to create Experiment 2b. A new stimulus set was created that consisted of the original gray-scale building images following randomization of each photograph's phase image (see Figure 1). The two subjects who participated in Experiment 2b passively viewed these images after phase randomization, as well as the other blocks of images described above for Experiment 2a, during fMRI scanning (450 fMRI observations per voxel).

The statistical models used to analyze the fMRI data from each subject were as described above for Experiment 1 with two changes. First, global signal covariates were not included, as the global signal across subjects was found to be significantly correlated with the thresholded buildings versus scrambled buildings contrast [t(4 df, two-tailed) = 4.08, p = 0.015] (Aguirre et al., 1997). Second, the dependent data used was the average signal obtained from the building-sensitive areas identified in Experiment 1. The significance of the contrasts outlined above was assessed at $\alpha = 0.05$, with no correction necessary for multiple independent comparisons (as only a single time series was being examined).

A group analysis was also performed upon the data collected during Experiment 2. The estimated β (i.e., $\hat{\beta}$) values for each contrast from each subject were obtained. These $\hat{\beta}$ values were then scaled by the intercept term from each model to normalize the values to percentage change units. Paired t tests were then performed upon the $\hat{\beta}$ values obtained across subjects for the buildings versus cars and thresholded buildings versus scrambled buildings contrasts. This mixed-effect model tests the null hypothesis that a difference does not exist between the mean level of evoked signal for the compared conditions across subjects. It should be noted that a statistical model such as this that explicitly accounts for subject variability is necessary to make inferences regarding the population from which the subjects were drawn (Woods, 1996).

Experiment 3
This experiment utilized a detection task similar to that of Experiment 1 but was implemented as an event-related design (Zarahn et al., 1997b). Prior to each of five scans (288 fMRI observations per voxel per scan, 1440 total), subjects studied one picture from each of the four categories (face, inverted face, building, and general object). During scanning, pictures were presented in a pseudo-random order for 2000 ms. If the subject judged that the picture matched one studied, she was to make a bilateral button press. If the picture was new, the subject was instructed to simply look at the picture and not make any overt response. Pictures were presented every 16 s, with a fixation cross occupying the center of the screen during the intertrial interval. On average, targets were presented on 11% of trials. Full and three-quarter view faces were randomly intermixed. Comparisons among the various orientations of face stimuli were not conducted here. For the contrasts evaluated here, full and three-quarter view faces were combined to create a single "upright face" covariate.

Because of the temporal spacing of the trials, this design allowed the analysis of the data within a trial-based framework (Zarahn et al., 1997b). Independent variables were created for each stimulus

type (i.e., buildings, faces, general objects, and inverted faces). In addition, all trials in which a target stimulus was presented or in which the subject made a response were modeled by a separate covariate. Thus, the primary covariates for each stimulus category were insensitive to signal changes which might have resulted from overt recognition or response. The independent variables themselves were formed using the first three principle components (eigenvectors) derived from a set of hemodynamic response functions from an independent group of subjects (Aguirre et al., 1998b). The first eigenvector resembles an across-subject, representative hemodynamic response function and it was the relationship between this covariate and the fMRI signal that was evaluated. The other two covariates were included to model nuisance variance components.

No temporal smoothing of the data with a hemodynamic response function was undertaken, in order to retain maximal temporal resolution. The K matrix (Worsley and Friston, 1995) contained a representation of the empirically observed null hypothesis power distribution in fMRI data (the 1/f model) and a filter designed to remove low frequency confounds (below 0.025 Hz) and high frequency noise at and around the Nyquist frequency (above 0.244 Hz). Covariates modeling the trial means were also included, ensuring that comparisons between different trial types were not confounded by poorly modeled low frequency fluctuations in signal. Application of this analysis method to null hypothesis data (similar to that performed by Aguirre et al., 1997; Zarahn et al., 1997a) demonstrated an empirical false-positive rate not significantly different from tabular values.

The average time series from within the regions defined in Experiment 1 served as the dependent data for the analysis, and each contrast was evaluated at an $\alpha = 0.05$ level. A mixed-effect group test (similar to that described above for Experiment 2) was also conducted using the $\hat{\beta}$, scaled to percentage signal change units. For display purposes, the time series data within the candidate regions were adjusted for nuisance covariates (Friston et al., 1995a) and trial averaged. Sinc interpolation was used to generate the smooth curves that are displayed.

As Experiment 3 was conducted during a different scanning session from Experiments 1 and 2 for most subjects, there was some concern that misregistration of the echoplanar images would bias the study toward negative results. This misregistration might result from distortion of the echoplanar images due to differences in static magnetic field inhomogeneities across scanning sessions, perhaps due to differential orientation of the sinuses. Thus, a post hoc test was conducted in those subjects in whom replication of building sensitivity failed during Experiment 3. The coregistered region of interest was expanded into a sphere with a radius of two voxels, encompassing 33 voxels in volume. This search region was then examined for voxels with significantly greater responses to buildings than to either faces or objects, evaluated at an $\alpha = 0.05$ level, Bonferroni corrected for the volume of the enlarged region. The use of this second, post hoc test inflates α slightly, but this minor increase in Type I error was deemed to be an acceptable exchange for the elimination of a recognized source of bias toward Type II errors.

References

Aguirre, G.K., Detre, J.A., Alsop, D.C., and D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. Cereb. Cortex 6, 823–829.

Aguirre, G.K., and D'Esposito, M. (1997). Environmental knowledge is subserved by separable dorsal/ventral neural areas. J. Neurosci. 17, 2512–2518.

Aguirre, G.K., Zarahn, E., and D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. Neuroimage 2, 199–212.

Aguirre, G.K., Zarahn, E., and D'Esposito, M. (1998a). Neural components of topographical representation. Proc. Natl. Acad. Sci. USA 95, 839–846.

Aguirre, G.K., Zarahn, E., and D'Esposito, M. (1998b). The variability of human BOLD hemodynamic responses. Neuroimage, in press.

Allison, T., Ginter, H., McCarthy, G., Nobre, A.C., Puce, A., Luby, M.,

and Spencer, D.D. (1994a). Face recognition in human extrastriate cortex. J. Neurophysiol. 71, 821–825.

Allison, T., McCarthy, G., Nobre, A., Puce, A., and Belger, A. (1994b). Human extrastriate visual cortex and the perception of faces, words, numbers, and colors. Cereb. Cortex 4, 544–554.

Damasio, A.R., Damasio, H., and Van Hoesen, G.W. (1982). Prosopagnosia: anatomic basis and behavioral mechanisms. Neurology 32, 331–341.

Duvernoy, H. (1991). The Human Brain: Surface, Three-Dimensional Sectional Anatomy, and MRI (New York: Springer-Verlag).

Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature 392, 568–601.

Farah, M.J. (1990). Visual Agnosia: Disorders of Object Recognition and What They Tell Us about Normal Vision (Cambridge, MA: MIT Press).

Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., and Turner, R. (1995a). Analysis of fMRI time series revisited. Neuroimage 2, 45–53.

Friston, K., Ashburner, J., Frith, C., Poline, J.-B., Heather, J., and Frackowiak, R. (1995b). Spatial registration and normalization of images. Hum. Brain Map. 2, 165–189.

Gauthier, I., Anderson, A.W., Tarr, M.J., Skudlarski, P., and Gore, J.C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. Curr. Biol. 7, 645–651.

Hécaen, H., Tzortzis, C., and Rondot, P. (1980). Loss of topographic memory with learning deficits. Cortex 16, 525–542.

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Landis, T., Cummings, J.L., Benson, D.F., and Palmer, E.P. (1986). Loss of topographic familiarity. Arch. Neurol. 43, 132–136.

Levine, D., Warach, J., and Farah, M. (1985). Two visual systems in mental imagery: dissociation of what and where in imagery disorders due to bilateral posterior cerebral lesions. Neurology 35, 1010–1018.

Lynch, K. (1960). The Image of the City (Cambridge, Mass: MIT Press).

Maguire, E.A., Frackowiak, R.S.J., and Frith, C.D. (1996). Learning to find your way: a role for the human hippocampal formation. Proc. R. Soc. Lond. B Biol. Sci. 263, 1745–1750.

Maguire, E.A., Frith, C.D., Burgess, N., Donnett, J.G., and O'Keefe, J. (1998). Knowing where things are: parahippocampal involvement in encoding object locations in virtual large-scale space. J. Cogn. Neurosci. 10, 61–76.

McCarthy, G., Puce, A., Gore, J.C., and Allison, T. (1997). Face-specific processing in the human fusiform gyrus. J. Cogn. Neurosci. 9, 605–610.

McCarthy, R.A., Evans, J.J., and Hodges, J.R. (1996). Topographical amnesia: spatial memory disorder, perceptual dysfunction, or category specific semantic memory impairment? J. Neurol. Neurosurg. Psychiatry 60, 318–325.

Moscovitch, M., Winocur, G., and Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. J Cogn. Neurosci. 9, 555–604.

Polk, T.A., and Farah, M.J. (1995). Brain localization for arbitrary stimulus categories: a simple account based on Hebbian learning. Proc. Natl. Acad. Sci. USA 92, 12370–12373.

Polk, T.A., and Farah, M.J. (1998). The neural development and organization of letter recognition: evidence from functional neuroimaging, computational modeling, and behavioral studies. Proc. Natl. Acad. Sci. USA 95, 847–852.

Puce, A., Allison, T., Gore, J.C., and McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. J. Neurophysiol. 74, 1192–1199.

Puce, A., Allison, T., Spencer, S.S., Spencer, D.D., and McCarthy, G. (1997). Comparison of cortical activation evoked by faces measured by intracranial field potentials and functional MRI: two case studies. Hum. Brain Map. 5, 298–305.

Rocchetta, A.I., Cipolotti, L., and Warrington, E.K. (1996). Topographical disorientation: selective impairment of locomotor space? Cortex *32*, 727–735.

Shallice, T., and Saffran, E.M. (1986). Lexical processing in the absence of explicit word identification: evidence from a letter-by-letter reader. Cogn. Neuropsychol. *3*, 429–458.

Takahashi, N., Kawamura, M., Hirayama, K., and Tagawa, K. (1989). [Non-verbal facial and topographic visual object agnosia—a problem of familiarity in prosopagnosia and topographic disorientation]. [Japanese]. No to Shinkei Brain Nerve *41*, 703–710.

Talairach, J., and Tournoux, P. (1988). Co-planar Stereotaxic Atlas of the Human Brain (New York: Thieme).

Tohgi, H., Watanabe, K., Takahashi, H., Yonezawa, H., Hatano, K., and Sasaki, T. (1994). Prosopagnosia without topographagnosia and object agnosia associated with a lesion confined to the right occipitotemporal region. J. Neurol. *241*, 470–474.

Warrington, E.K., and Shallice, T. (1980). Word-form dyslexia. Brain *103*, 99–112.

Whiteley, A.M., and Warrington, E.K. (1978). Selective impairment of topographical memory: a single case study. J. Neurol. Neurosurg. Psychiatry *41*, 575–578.

Woods, R.P. (1996). Modeling for intergroup comparisons of imaging data. Neuroimage *4* (suppl.), 84–94.

Worsley, K.J., and Friston, K.J. (1995). Analysis of fMRI time series revisited—again. Neuroimage *2*, 173–182.

Zarahn, E., Aguirre, G.K., and D'Esposito, M. (1997a). Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. Neuroimage *5*, 179–197.

Zarahn, E., Aguirre, G.K., and D'Esposito, M. (1997b). A trial-based experimental design for fMRI. Neuroimage *6*, 122–138.