12

Making a scene in the brain

RUSSELL A. EPSTEIN AND SEAN P. MACEVOY

12.1 Introduction

Humans observers have a remarkable ability to identify thousands of different things in the world, including people, animals, artifacts, structures, and places. Many of the things we typically encounter are objects – compact entities that have a distinct shape and a contour that allows them to be easily separated from their visual surroundings. Examples include faces, blenders, automobiles, and shoes. Studies of visual recognition have traditionally focused on object recognition; for example, investigations of the neural basis of object and face coding in the ventral visual stream are plentiful (Tanaka, 1993; Tsao and Livingstone, 2008; Yamane *et al.*, 2008).

Some recognition tasks, however, involve analysis of the entire scene rather than just individual objects. Consider, for example, the situation where one walks into a room and needs to determine whether it is a kitchen or a study. Although one might perform this task by first identifying the objects in the scene and then deducing the identity of the surroundings from this list, this would be a relatively laborious process, which does not fit with our intuition (and behavioral data) that we can identify the scene quite rapidly. Consider as well the challenge of identifying one's location during a walk around a city or a college campus, or through a natural wooded environment. Although we can perform this task by identifying distinct object-like landmarks (buildings, statues, trees, etc.), we also seem to have some ability to identify places based on their overall visual appearance.

Vision in 3D Environments, ed. L. R. Harris and M. R. M. Jenkin. Published by Cambridge University Press.

These observations suggest that our visual system might have specialized mechanisms for perceiving and recognizing scenes that are distinct from the mechanisms for perceiving and recognizing objects. There are many salient differences between scenes and objects that could lead to such specialization. Some of these differences are structural. As noted above, objects are spatially and visually compact, with a well-defined contour. In contrast, scenes are spatially and visually distributed, containing both foreground objects and fixed background elements (walls, ground planes, and distant topographical features). Consequently, we tend to speak of the shape of an object but the layout of a scene. Related to these structural differences are differences in ecological relevance. As a consequence of their compactness, objects are typically things that one can move – or, at least, imagine moving – to different locations in the world. In contrast, scenes are locations in the world; thus, it makes little sense to think of "moving" a scene. Put another way, objects are things one acts upon; scenes are things one acts within. Because of this, scenes are relevant to navigation in a way that objects typically are not.

In this chapter, we will review the evidence for specialized scene recognition mechanisms in the human visual system. We will begin by reviewing behavioral literature that supports the contention that scene recognition involves analysis of whole-scene features such as scene layout and hence might involve processes that are distinct from those involved in object recognition. We will then discuss evidence from neuroimaging and neuropsychology that implicates a specific region in the human occipitotemporal lobe – the parahippocampal place area (PPA) – in processing of whole-scene features. We will describe various data analysis techniques that facilitate the use of functional magnetic resonance imaging (fMRI) as a tool to explore the representations that underlie scene processing in the PPA. Finally, we will bring objects back into the picture by describing recent experiments that explore how objects are integrated into the larger visual array.

12.2 Efficacy of human scene recognition

The first evidence for specialized mechanisms for scene perception came from a classic series of studies by Potter and colleagues (Potter, 1975, 1976; Potter and Levy, 1969) showing that human observers can interpret complex visual scenes with remarkable speed and efficiency. Subjects in these experiments were asked to detect the presence of a target scene within a sequential stream of distractors. Each scene in the stream was presented for a brief period of time (e.g., 125 ms) and then immediately replaced by the next item in the sequence, a technique known as *rapid serial visual presentation*. Despite the challenging nature of this task, the subjects were remarkably efficient, detecting the target on 75% of the trials. This high level of performance was observed even when the target scene was defined only by a high-level verbal description (e.g., “a picnic” or “two people talking”) so that subjects could not know exactly what it would look like. Potter concluded that the subjects could not have been doing the task based on simple visual feature mapping: they must have processed the scene up to the conceptual level. The human visual system appears to be able to extract the gist (i.e., overall meaning) of a complex visual scene within 100 ms.

Related results were obtained by Biederman (1972), who observed that recognition of a single object within a briefly flashed (300–700 ms) scene was more accurate when the scene was coherent than when it was scrambled (i.e., jumbled up into pieces). Biederman concluded that the human visual system can extract the meaning of a complex visual scene within a few hundred milliseconds and use it to facilitate object recognition. A notable aspect of this experiment is the fact that the visual elements of the scene were present in both the intact and the scrambled conditions. Thus, it is not the visual elements alone that affect object recognition, but the organization into a meaningful scene. In other words, the layout of the scene matters, lending support to the notion of a scene recognition system that is distinct from and might influence object recognition. Subsequent behavioral work has upheld the proposition that even very complex visual scenes can be interpreted very rapidly (Antes *et al.*, 1981; Biederman *et al.*, 1974; Fei-Fei *et al.*, 2007; Thorpe *et al.*, 1996).

Although one might argue that scene recognition in these earlier studies reduces simply to recognition of one or two critical objects within the scene, subsequent work has provided evidence that this is not the whole story. Scenes can be identified based on their whole-scene characteristics, such as their overall spatial layout, without reducing them to their constituent objects. Schyns and Oliva (1994) demonstrated that subjects could classify briefly flashed (30 ms) scenes into categories (highway, city, living room, and valley) even if the scenes were filtered to remove all high-spatial-frequency information, leaving only an overall layout of coarse blobs, which conveyed little information about individual objects. Computational modeling work has given further credence to this idea by demonstrating that human recognition of briefly presented scenes could be simulated by recognition systems that operated solely on whole-scene characteristics. For example, Renniger and Malik (2004) developed an algorithm that classified scenes from 10 different categories based solely on their visual texture statistics. The performance of the model was comparable to that of humans attempting to recognize briefly presented (37 ms) scenes. Then Greene and Oliva (2009) developed a scene recognition model that operated on seven global properties: openness, expansion, mean depth, temperature, transience, concealment, and navigability. These properties predicted the performance of human observers, insofar as scenes that were more similar in the property space were more often misclassified by the observers.

Results such as these provide an "existence proof" that it is possible to identify scenes on the basis of global properties rather than by first identifying the component objects. Indeed, the similarity of the human and model error patterns in Greene and Oliva's study strongly suggests that we actually use these global properties for recognition. In other words, we have the ability to identify the scenic "forest" without having to first represent the individual "trees." Note that this does not mean that scenes are never recognized based on their component objects (we discuss this scenario later). However, it does argue for the existence of mechanisms that can process these scene-level properties.

A final line of behavioral evidence for specialized scene-processing mechanisms comes from the phenomenon of boundary extension (BE). When subjects are shown a photograph of a scene and are later asked to recall it, they tend to remember it as being more wide-angle than it actually was – as if its boundaries had been extended beyond the edges of the photograph (Intraub *et al.*, 1992). Although it was initially thought to be a constructive error, more recent results indicate that BE occurs even when the scene has only been absent for an interval as short as 42 ms, suggesting that it is not an artifact of memory but indicative of the scene representations formed during online perception (Intraub and Dickinson, 2008). Specifically, a representation of the layout of the local environment may be formed during scene viewing that is more expansive than the portion of the scene shown in the photograph. When the photograph is removed, this layout representation remains, influencing one's memory of the width of the scene. BE occurs only for scenes: it is not found after viewing decontextualized objects. Thus, BE provides another line of evidence for distinct scene- and object-processing systems because it indicates the existence of a special representation for scenic layouts.

12.3 Scene-processing regions of the brain

What are the neural systems that support our ability to recognize scenes? The first inkling that there might be specialized cortical territory for scene recognition came from a 1998 fMRI paper (Epstein and Kanwisher, 1998). Subjects were scanned with fMRI while they viewed indoor and outdoor scenes along with pictures of common objects (blenders, animals, and tools) and faces. A region in the collateral sulcus near the parahippocampal–lingual boundary responded much more strongly to the scenes than to the other stimuli

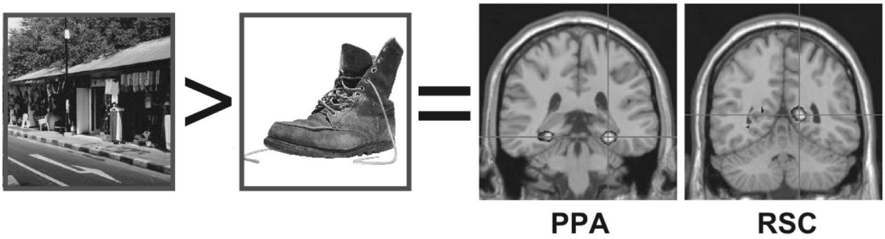

Figure 12.1 Scene-responsive cortical regions. Subjects were scanned with fMRI while viewing scenes and nonscene objects. The parahippocampal place area (PPA) and retrosplenial complex (RSC) respond more strongly to the scenes than to the objects (highlighted voxels). The PPA straddles the collateral sulcus near the parahippocampal–lingual boundary. The RSC is in the medial parietal region and extends into the parietal–occipital sulcus. Scene-responsive territory is also observed in the transverse occipital sulcus region (not shown). A color version of this figure can be found on the publisher's website (www.cambridge.org/9781107001756).

(Figure 12.1). This region was labeled the *parahippocampal place area*, or PPA, because it responded preferentially to images of places (i.e., scenes). These results were similar to contemporaneous findings that the parahippocampal–lingual region responds more strongly to houses and buildings than to faces (Aguirre *et al.*, 1998; Ishai *et al.*, 1999), which makes sense if one considers the fact that the facade of a building is a kind of partial scene.

Subsequent work demonstrated that the PPA responds to a wide range of scenes, including cityscapes, landscapes, rooms, and tabletop scenes (Epstein *et al.*, 2003; Epstein and Kanwisher, 1998). It even responds strongly to "scenes" made out of Lego blocks (Epstein *et al.*, 1999). This last result is particularly interesting because the comparison condition was objects made out of the same Lego materials but organized as a compact object rather than a distributed scene (Figure 12.2a). Thus, the geometric structure of the stimulus appears to play an important role in determining whether the PPA interprets it as a "scene" or an "object."

We refer to the idea that the PPA responds to the overall structure of a scene as the layout hypothesis. Some of the strongest evidence for this idea comes from a study in which subjects viewed rooms that were either filled with furniture and other potentially movable objects or emptied such that they were just bare walls (Figure 12.2b). In a third condition, the objects from the rooms were displayed in a multiple-item array on a blank background. The PPA responded strongly to the rooms irrespective of whether they contained objects or not. In contrast, the response to the decontextualized objects was much lower. This finding suggests that the PPA responds strongly to layout-defining background elements but may

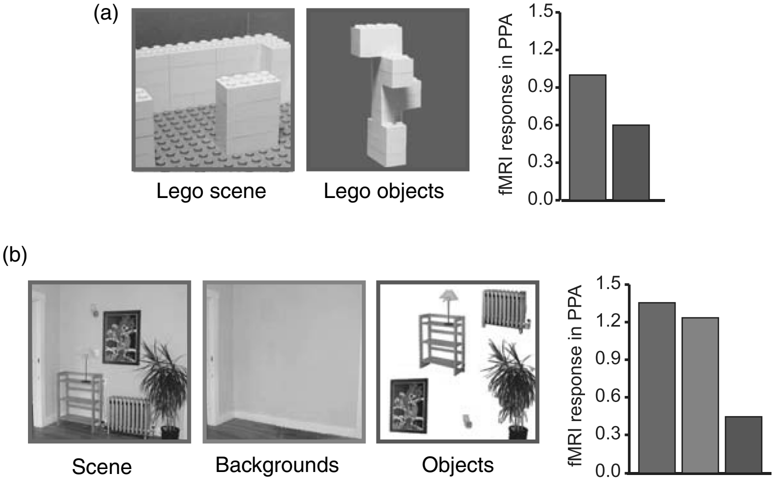

Figure 12.2 The PPA encodes scenic layout. The magnitude of the PPA's response to each of a series of scenes is shown in a bar chart on the right. (a) The PPA responds more strongly to "scenes" made out of Lego blocks than to "objects" made out of the same materials, indicating that the spatial geometry of the stimulus affects the PPA response. (b) The PPA responds almost as strongly to background elements alone as it does to scenes containing both foreground objects and background elements; in contrast, the response to the foreground objects alone is much lower. Note that although the locations of the objects were randomized in this experiment, additional studies indicated that the response in the object-alone condition was low even when the spatial arrangement was preserved from the original scene (Epstein and Kanwisher, unpublished data). The PPA was identified in each subject as the set of voxels in the collateral sulcus region that responded more strongly to real-world scenes than to real-world objects in a separate data set. A color version of this figure can be found on the publisher's website (www.cambridge.org/9781107001756).

play little role in processing the discrete objects within the scene. Originally, we (Epstein and Kanwisher, 1998) hypothesized that the PPA might respond solely to spatial layout – the geometric structure of the scene as defined by these large surfaces. Behavioral data suggest that humans and animals preferentially use such geometric information during reorientation (Cheng, 1986; Cheng and Newcombe, 2005; Hermer and Spelke, 1994); thus, we speculated that the PPA might be the neural locus of a geometry-based reorientation module (Gallistel, 1990). However, the idea that the PPA responds solely to scene geometry while ignoring nongeometric visual features such as color or texture has not been proven, and indeed there is some evidence against this claim (Cant and Goodale, 2007).

In general, though, the neuroimaging data suggest that the PPA plays an important role in processing global aspects of scenes rather than the objects that make them up. (However, see Bar (2004) for a different view.) This conclusion is consistent with neuropsychological data, which support the proposed role of the posterior parahippocampal/anterior lingual region in scene recognition. As one might expect from the neuroimaging literature, the ability to recognize scenes is often seriously degraded following stroke damage to this region. This syndrome is sometimes referred to as topographical agnosia because the patient has difficulty identifying a wide variety of large topographical entities, such as buildings, monuments, squares, vistas, and intersections (Mendez and Cherrier, 2003; Pallis, 1955). Interestingly, this deficit is usually reported as an inability to identify specific places. To our knowledge, a deficit in being able to recognize scenes at the categorical level (e.g., beach vs. forest) is not spontaneously reported, perhaps because there are alternative processing pathways for category analysis that do not require use of the PPA.

Topographical agnosia patients often complain that some global organizing aspect of the scene is missing (Epstein *et al.*, 2001; Habib and Sirigu, 1987); for example, a patient described by Hécaen (1980) reported that "it's the whole you see, in a very large area I can identify some minor detail or places but I don't recognize the whole." This inability to perceive scenes as unified entities is also evident in a compensatory strategy that is often observed: rather than identifying places in a normal manner, topographical agnosia patients will often focus on minor visual details such as a distinctive mailbox or door knocker (Aguirre and D'Esposito, 1999). An important contrast to these patients is provided by DF, a patient whose PPA is preserved but whose object-form-processing pathway (including the lateral occipital complex (LOC)) is almost completely obliterated. Despite an almost total inability to recognize objects, DF is able to classify scenes in terms of general categories such as city, beach, or forest; furthermore, her PPA activates during scene viewing (Steeves *et al.*, 2004). These results demonstrate a double dissociation between scene and object processing: the PPA appears to be part of a distinct processing stream for scenes that bypasses the more commonly studied object-processing pathway. This contention is also supported by anatomical connectivity studies (Kim *et al.*, 2006).

In sum, the PPA appears to play a critical role in scene recognition. Note, however, that this does not mean that it is the only region involved. At least two other regions respond more strongly to scenes than to other stimuli: a retrosplenial/medial parietal region that has been labeled the retrosplenial complex (RSC; Figure 12.1) and an area near the transverse occipital sulcus (TOS). We have published several studies exploring the idea that the PPA and RSC may play distinct roles in scene processing, with the PPA more concerned with scene perception/recognition and the RSC more involved in linking the local scene to long-term topographical memory stores (Epstein and Higgins, 2007; Epstein *et al.*, 2007b). In this chapter, we focus primarily on the PPA because of its critical role in visual recognition. A perhaps even more important point, however, is to remember that a region does not have to respond preferentially to scenes to play an important role in scene recognition. Although it might be possible to recognize scenes based on whole-scene characteristics such as layout, they might also be identified in part through analysis of their constituent objects, in which case object-processing regions such as the Lateral Occipital Complex (LOC; Malach *et al.*, 1995) and fusiform gyrus might be involved. We will explore this idea later in this chapter.

12.4 Probing scene representations with fMRI

What is the nature of the scene representations encoded by the PPA? Broadly speaking, we can imagine at least three very different ways in which the PPA might encode scenes. First, the PPA might encode the "shape" or "geometry" of the scene as defined primarily by the large bounding surfaces. In this view, the representations supported by the PPA would be inherently three-dimensional, insofar as they code information about the locations of surfaces, boundaries, affordances, and openings in the scene. Second, the PPA might encode a "visual snapshot" of what scenes look like from particular vantage points. In this case, the PPA representation would be inherently two-dimensional, insofar as the coded material would be the distribution of visual features across the retina rather than the locations of surfaces in three-dimensional space. Finally, the PPA might encode a spatial coordinate frame that is anchored to the scene but does not include details about scene geometry (Shelton and Pippitt, 2007). In this case, the geometry of the scene would be processed, but only up to the point at which it is possible to determine the principal axis of the environment and to determine the observer's orientation relative to this axis. Such a spatial code might be less useful for identifying the scene as a particular place or type of place, but might be ideal for distinguishing between different navigational situations.

To help differentiate among these scenarios, an important question is whether the pattern of activation evoked in the PPA by a given scene changes with the observer's viewpoint. That is, are two different views of the scene coded the same or differently? In the first scenario, the scene geometry could be could be defined relative to either a viewer-centered or a scene-centered axis (Marr, 1982). In the second scenario, we would expect PPA scene representations to be entirely viewpoint-dependent because the distribution of visual features on the retina necessarily changes with viewpoint. In the third scenario, one would also expect some degree of viewpoint-dependent coding, but the PPA might be more sensitive to some viewpoint changes than to others. For example, it might be more sensitive to viewpoint changes caused by changes in orientation but less sensitive to viewpoint changes caused by changes in position in which the orientation relative to the scene remains constant (Park and Chun, 2009).

How can we address such a question with neuroimaging? Although the precise way that cognitive "representations" relate to neural activity is unclear (Gallistel and King, 2009), it is usually assumed that two items are representationally distinct if they cause different sets of neurons to fire. For example, neurons in area MT are tuned for motion direction and speed. As not all neurons have the same tuning, stimuli moving in different directions activate different sets of neurons in this region. Ideally, we would like to know whether two views of the same scene activate the same or different neuronal sets in the PPA. Although the spatial resolution of fMRI is far too coarse to address this question directly, two data analysis methods have been used to get at the question indirectly.

The first method is multivoxel pattern analysis (MVPA). The basic unit of fMRI data is the voxel (volume element), which can be though of as the three-dimensional equivalent of a pixel. Each fMRI image typically consists of thousands of such voxels, each corresponding to a cube of brain tissue that typically measures 2–6 mm on a side. For example, a region such as the PPA might include from several dozen to several hundred 3 × 3 × 3 mm voxels. In standard fMRI analysis, one averages the responses of these voxels together. This gives one an estimate of how much the region as a whole activates in response to each condition. Although such univariate analyses are very useful for determining the kind of stimulus the region prefers (e.g., scenes but not faces in the case of the PPA), they are less useful for determining representational distinctions within the preferred stimulus class (e.g., whether the PPA distinguishes between forests and beaches).

MVPA gets around this problem by doing away with the averaging. The voxel-by-voxel response pattern is treated as a multidimensional vector and various tests are done to determine whether the response vectors elicited by one condition are reliably different from the response vectors elicited by another. For example, in the paper that first popularized this technique, Haxby *et al.* (2001) used MVPA to demonstrate that the ventral occipitotemporal cortex distinguished between eight object categories (faces, houses, cats, scissors, bottles, shoes, chairs, and scrambled nonsense patterns). A more recent study by Walther *et al.* (2009) used MVPA to demonstrate that the PPA, the RSC, and the lateral occipital object-sensitive region can reliably distinguish between scene categories such as beaches, buildings, forests, highways, mountains, and industrial scenes.

In theory, it should be possible to use MVPA to investigate the issue of viewpoint sensitivity. In particular, if different views of the same scene were to activate distinguishable patterns in the PPA, this would establish viewpoint specificity. However, a negative result would not necessarily demonstrate viewpoint-invariant coding (at least, not at the single-neuron level). The extent to which MVPA can be used to interrogate representations that are organized at spatial scales much smaller than a voxel is a matter of considerable debate (Drucker and Aguirre, 2009; Kamitani and Tong, 2005; Sasaki *et al.*, 2006). One could imagine a scenario in which individual PPA neurons are selective for different views of a scene, but in which these neurons are tightly interdigitated within each voxel such that it is impossible to distinguish between different views of a scene using MVPA. In this scenario, MVPA would not reveal the underlying neural code. It would, however, reveal aspects of the representation that are implemented at a coarser (supraneuronal) spatial scale; for example, whether neurons responding to different views of the same scene (or scene category) are clustered together.

The second method for addressing such representational questions is fMRI adaptation (sometimes also referred to as fMRI attenuation or fMRI repetition suppression). Here one looks at the reduction in fMRI response when a stimulus is repeated (Grill-Spector *et al.*, 2006; Grill-Spector and Malach, 2001). The critical question is whether response reduction occurs when the repeated stimulus is a modified version of the original. If it does, one infers that the original stimulus and the modified stimulus are representationally similar. For example, one might examine whether a previously encountered scene elicits a response reduction if shown from a previously unseen viewpoint. If so, we conclude that the representation of the scene is (at least partially) viewpoint-invariant. This technique is motivated by neurophysiological findings indicating that neurons in many regions of the brain respond more strongly to the first presentation of a stimulus than to later presentations (Miller *et al.*, 1993), as well as by behavioral studies indicating that experience with a stimulus can lead to an adapted detection threshold (Blakemore and Nachmias, 1971).

In a typical implementation of an fMRI adaptation paradigm, two stimuli are presented within an experimental trial, separated by interval of 100–700 ms (Kourtzi and Kanwisher, 2001). For example, the two images could be different scenes, different views of the same scene, or the same view of the same scene (i.e., identical images). In this example, the “different-scene” condition is taken to be the baseline; that is, the condition for which there is no adaptation. The question of interest is then whether the response is reduced compared with this baseline when a scene is repeated from a different viewpoint (implying some degree of viewpoint invariance) or reduced only when a scene is repeated from the same viewpoint (implying viewpoint specificity). Across several experiments, we have consistently found that adaptation using this paradigm is viewpoint-specific (Epstein *et al.*, 2003, 2005, 2007a, 2008). No (or very little) response reduction is observed in the same-scene/different-view condition, suggesting that, at least for large enough viewpoint changes, two views of the same scene are as representationally distinct as two different scenes.

This would seem to resolve the issue. However, somewhat puzzlingly, one can get a somewhat different answer by implementing the fMRI adaptation paradigm in a slightly different way. Rather than repeating stimuli almost immediately within an experimental trial, one can repeat stimuli over longer intervals (several seconds or minutes) during which many other stimuli are observed (Henson, 2003; Vuilleumier *et al.*, 2002). In this case, one treats the response to a previously unviewed scene ("new scene") as the baseline and examines whether the response is lower for scenes that were previously viewed from a different viewpoint ("new view") or reduced only for scenes that were previously viewed from the same viewpoint ("old view"). This gives a somewhat different result. Significant response reductions are observed for new views compared with new scenes, suggesting some generalization of processing across views (Epstein *et al.*, 2007a, 2008). The adaptation effect is not entirely viewpoint-invariant, as we observe a further reduction of the response for old views compared with new views; nevertheless, the pattern is quite different from that observed with the short-interval repetition paradigm.

What are we to make of this discrepancy? The short-interval and long-interval fMRI adaptation paradigms appear to be indexing different aspects of scene processing. To verify that these effects were indeed distinct, we implemented an experiment in which both kinds of fMRI adaptation could be measured simultaneously (Epstein *et al.*, 2008). Subjects were scanned while viewing scene images that were paired together to make three kinds of short-interval repetition trials (different scene, same-scene/different-view, and same-scene/same-view). Critically, the subjects had been familiarized with some but not all of the scene images before the scan session (see Figure 12.3a). This allowed us to cross the three short-interval (i.e., within-trial) repetition conditions with three long-interval repetition conditions (new scene, new view, and old view) to give nine trial types in which the short-interval and long-interval repetition states were independently defined. Thus, for example, different-scene trials could be constructed from new scenes, from new views, and from old views. When we measured the fMRI response to these nine conditions, we observed a complete lack of interaction between the short-interval and long-interval

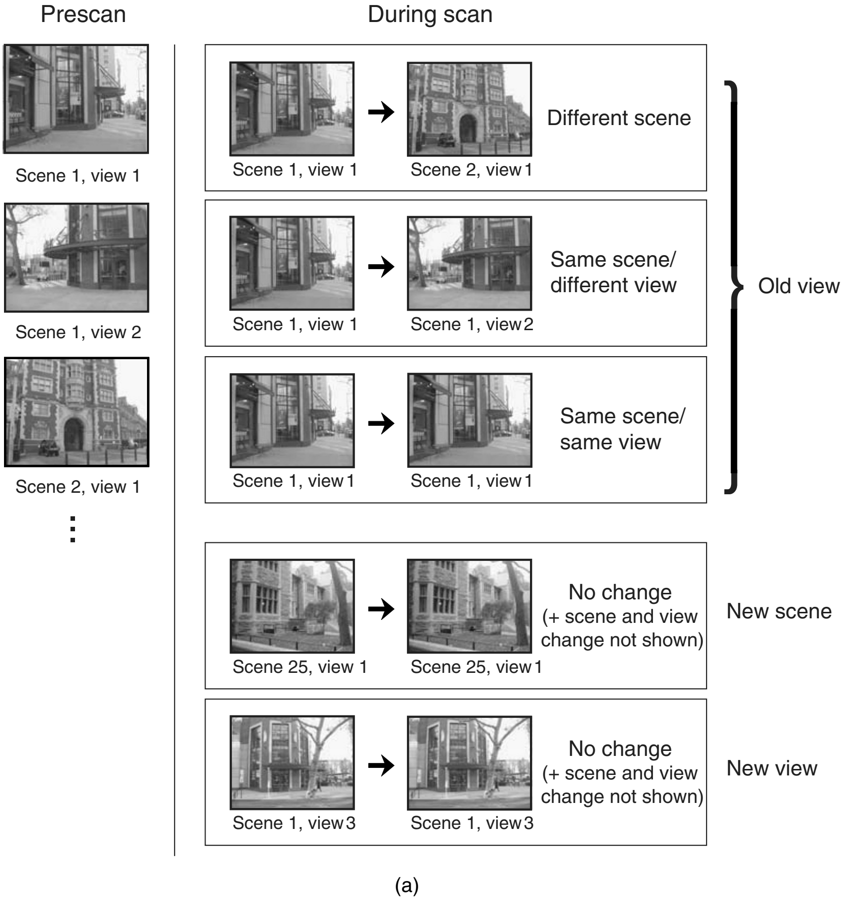

Figure 12.3 Long-interval and short-interval fMRI adaptation in the PPA. (a) Design of experiment. Subjects studied 48 scene images (two views each of 24 campus locations) immediately prior to the scan. The stimuli shown during the scan were either the study images ("old views"), previously unseen views of the studied locations ("new views"), or locations not presented in the study session ("new place"). Long-interval adaptation was examined by comparing fMRI response across these three conditions. Short-interval adaptation, on the other hand, was examined by measuring the effect of repeating items within a single experimental trial (different place, same place/different view, and same place/same view). The three short-interval repetition conditions were fully crossed with the three long-interval repetition conditions to give nine conditions in total, five of which are illustrated in the figure. (b) Results. Short-interval adaptation is almost entirely viewpoint-specific: response reduction is observed only when scenes are repeated from the same viewpoint. In contrast, long-interval adaptation is much more viewpoint-invariant: repetition reduction is observed even when scenes are presented from a different viewpoint. These data argue for distinct mechanisms underlying the two adaptation effects. A color version of this figure can be found on the publisher's website (www.cambridge.org/9781107001756).

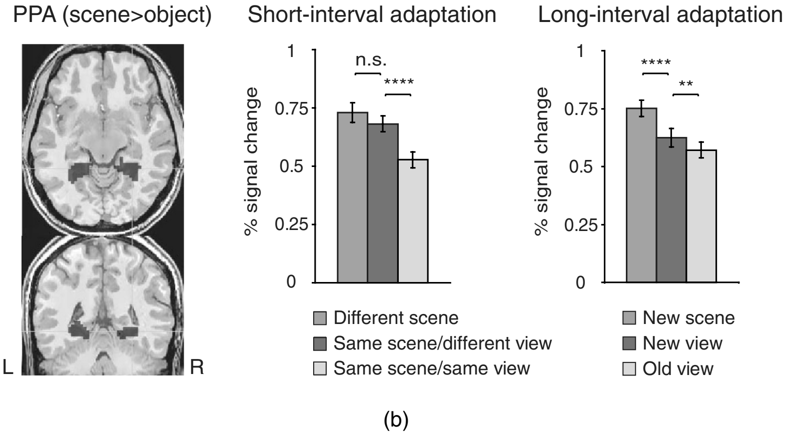

Figure 12.3 Continued

repetition effects, suggesting that they are caused by different underlying mechanisms. Furthermore, in a manner consistent with our earlier results, the short-interval adaptation effects were completely viewpoint-specific, while the long-interval adaptation effects showed a high degree of viewpoint invariance (Figure 12.3b).

The fact that long-interval adaptation effects are more viewpoint-invariant than short-interval adaptation effects suggests that the long-interval effects might be initiated somewhere later in the processing stream. One possibility is that long-interval adaptation effects in the PPA might be caused by top-down modulation from other cortical regions. Indeed, long-interval effects are typically more prominent in areas involved in "mnemonic" processing such as the lateral frontal lobes (Buckner *et al.*, 1998; Maccotta and Buckner, 2004) and the hippocampus (Epstein *et al.*, 2008; Gonsalves *et al.*, 2005), while short-interval effects are typically more prominent in early visual regions (Epstein, 2008; Ganel *et al.*, 2006). Furthermore, transcranial magnetic stimulation (TMS) studies have linked the behavioral priming effects normally associated with long-interval adaptation to processing in the frontal lobes (Wig *et al.*, 2005). Thus, the idea that long-interval adaptation reflects viewpoint-invariant representations coded outside the PPA, whereas short-interval adaptation effects reflect viewpoint-specific representations inherent to the PPA, has some appeal. However, it is worth noting that whole-brain analyses revealed no clear high-level source for the long-interval repetition effect. Thus, at present, the hypothesis that short- and long-interval adaptation originate in distinct cortical regions must be considered speculative. Functional connectivity analyses might be useful for investigating this hypothesis further.

A second possibility is that short- and long-interval adaptation effects reflect two separate mechanisms that are both anatomically localizable to the PPA. For example, short-interval fMRI adaptation might reflect short-term changes in the synaptic inputs to the PPA, perhaps caused by synaptic depression (Abbott *et al.*, 1997), which is believed to operate on a relatively short timescale of less than 2 s (Muller *et al.*, 1999). Long-interval adaptation, on the other hand, might reflect more permanent changes in interregional connectivity (Wiggs and Martin, 1998). Under this scenario, the PPA takes visual inputs in which different views of the same scene are representationally distinct and calculates a new representation in which different views of the same scene are representationally similar. This new representation could then be used as the basis for scene recognition.

Although the interpretation of the fMRI adaptation data in terms of two different adaptation mechanisms is somewhat speculative, it is consistent with recent neurophysiological data. In an intriguing single-unit recording study, Sawamura *et al.* (2006) recorded from object-sensitive neurons in the monkey inferior temporal (IT) cortex and examined the adaptation effect when identical object images were repeated after a short interval (300 ms). In a second condition, they measured the cross-adaptation effect caused by presenting two different objects in sequence. Critically, these two different objects were chosen such that the neurons responded equally strongly to both in the absence of adaptation. Despite the fact that the neuron considered these objects to be “the same” in terms of firing rate, the objects were distinguishable in terms of adaptation. Specifically, there was more adaptation when the same object was presented twice than when two “representationally identical” objects were presented in sequence. Sawamura *et al.* hypothesized that these repetition reductions might reflect adaptation at the synaptic inputs to the neuron, which would be nonoverlapping for the two objects. In this scenario, short-interval fMRI adaptation may reflect the selectivity of the neurons that provide input to a region rather than reflecting the selectivity of the neurons within that region itself.[1]

[1] This scenario might also explain a recent failure to find orientation-specific adaptation in V1 using the short-interval repetition paradigm (Boynton and Finney, 2003). The fact that V1 neurons are tuned for the orientation of lines and gradients has been well established using direct recording methods (Hubel and Wiesel, 1962). Thus, the absence of orientation-specific adaptation when two gratings were presented within a trial was surprising. A subsequent experiment replicated this finding (Fang *et al.*, 2005) and also demonstrated that orientation-specific adaptation could be observed if a different adaptation paradigm was used in which adapting gratings were presented for several seconds in order to induce neural fatigue (Fang *et al.*, 2005). One possible explanation for these discrepant results is that short-interval adaptation effects might be dominated by suppression in the thalamic inputs to V1 which do not contain information about grating orientation. A later experiment by the same group observed similar results for faces within the fusiform face area (Fang *et al.*, 2007). The “fatigue” paradigm

Note that, even under this second scenario, individual neurons in the PPA might not exhibit viewpoint-invariant responses. Although long-term adaptation effects show a much greater tolerance for viewpoint changes than do short-interval adaptation effects, they retain some degree of viewpoint specificity. Furthermore, it is possible that long-interval effects might reflect changes in network connectivity that do not relate directly to the tuning of individual neurons. For example, one could imagine that neurons corresponding to distinct views of a particular building might be intermixed within a cortical column. If long-interval adaptation operates on columns rather than neurons, one would observe viewpoint-invariant adaptation even though the individual neurons within the column were viewpoint-tuned. Indeed, a likely scenario is that PPA neurons are analogous to the object-sensitive neurons in the monkey IT, which respond in a viewpoint-sensitive manner. Recent studies have demonstrated that object identity can be “read out” from ensembles of such neurons even though there are almost no individual neurons that respond to object identity under all possible transformations of lighting, position, and viewpoint (DiCarlo and Cox, 2007; Hung *et al.*, 2005; Tsao *et al.*, 2006). Furthermore, the same set of neurons can provide information about both object category and object identity. Similarly, ensembles of PPA neurons might be readable by other cortical regions in terms of both the category and the identity of the scene being viewed, even if individual neurons are tuned to specific views of specific scenes.

So, what do these MVPA and fMRI adaptation results tell us about the function of the PPA? The MVPA data of Walther and colleagues suggest that the PPA encodes information that allows scene categories to be distinguished. However, as noted above, these results do not necessarily indicate that PPA neurons are tuned for category (in the sense that there are “kitchen neurons”). Indeed, to our knowledge, there have been no reports of category-specific adaptation (kitchen A primes kitchen B) in the PPA, and some recent unpublished data from our laboratory have failed to find such an effect (MacEvoy and Epstein, 2009a; Morgan *et al.*, 2009). Rather, fMRI adaptation data indicate that the PPA adapts to repetition of a specific scene (i.e., two images of my kitchen

revealed orientation-tuned adaptation that decreased gradually as the angular difference between the adapting and the adapted face increased, in a manner consistent with the orientation-selective tuning curves observed in neurophysiological data. The short-interval adaptation paradigm, on the other hand, revealed adaptation effects that were much more orientation-specific, insofar as adaptation was observed only when the adapting and adapted faces were shown from identical viewpoints. These data are similar to our own insofar as they indicate that measurements of short-interval fMRI adaptation effects may overestimate the degree of stimulus specificity within a region.

rather than two images depicting two different kitchens), with incomplete tolerance for viewpoint changes (and a large degree of tolerance for retinal-position changes (MacEvoy and Epstein, 2007)). Thus, our best guess is that PPA neurons encode visual or spatial quantities that vary somewhat with viewpoint and differ strongly between individual scenes of the same category. Furthermore, the neurons responding to these quantities are clustered unequally within the region such that different scene categories elicit different voxelwise activation patterns. However, the precise nature of these quantities – whether they are geometric shape parameters, 2D visual features, or spatial relationships – is unknown. Indeed, the question is almost entirely unexplored.

12.5 Integrating objects into the scene

The literature reviewed thus far clearly implicates the PPA in scene recognition. We have argued that this recognition probably involves analysis of whole-scene quantities such as spatial layout. We are skeptical of the idea that the PPA encodes the individual objects within the scene, although it is worthwhile to note that recent MVPA studies indicate that PPA activation patterns can provide information about individual objects when they are presented in isolation (Diana *et al.*, 2008). However, there are clearly circumstances in which information about object identity can be an important cue for recognizing a scene. Indeed, part of our ability to quickly understand the "gist" of a scene must involve integration of information about object identities; for example, understanding that a computer, a desk, a whiteboard, a lamp, and some chairs make an office. How might this object information be integrated together into a scene?

We addressed this question in a recent study (MacEvoy and Epstein, 2009b) in which we examined the fMRI response to multiple-item object arrays. Although these stimuli were not "scenes"– they contained no background elements and had no three-dimensional layout – they allowed us to explore the rules by which the visual system combines object representations when more than one object appears on the screen. Subjects were scanned while viewing blocks containing either single objects (chairs, shoes, brushes, or cars) or two-object arrays (with each object from a different category). To ensure that the subjects attended equally to both of the objects in the pairs, we asked them to detect stimulus repetitions that could occur randomly at either object location. We then used MVPA to decode the response to both single-object categories and object-category pairs. Our analyses focused on the LOC, which is the area of the brain that appears to be critically involved in processing object identity. Recent studies indicate that LOC response patterns can be used to decode scene categories (Walther *et al.*, 2009) and also objects within scenes (Peelen *et al.*, 2009).

We found that the multivoxel response patterns in the LOC contained information not only about single objects (as demonstrated previously by Haxby and colleagues) but also about object pairs. In other words, one can use the distributed pattern of the fMRI response to tell not only whether the subject is looking at a shoe or a brush but also whether he or she is looking at a shoe and a brush together. Moreover, we were able to reliably distinguish between patterns evoked by object pairs that shared an object, such as shoe + brush, shoe + car, and shoe + chair. Thus the pattern evoked by each object array was uniquely prescribed by its particular combination of objects.

This "uniqueness" comes from the particular way in which patterns evoked by pairs are constructed in the LOC. The patterns evoked by pairs were not random; rather, they obeyed an ordered relationship to the patterns evoked by each of their component objects when those objects were presented alone. Specifically, pair patterns were very well predicted by the arithmetic mean of the patterns evoked by each of the component objects (Figure 12.4). This relationship was strong enough that we were able to decode with 75% accuracy (chance = 50%, $p < 10^{-5}$) the patterns evoked by pairs using synthetic pair patterns created by averaging pairs of single-object patterns. In this way, pair patterns not only are unique with respect to each other, but also ensure that the identity of each of their component objects is preserved.

We hypothesized that this averaging rule may be a general solution to the problem of preserving the representations of multiple simultaneous objects in a population of broadly tuned neurons. Because averaging is a linear operation (i.e., summation followed by uniform scaling), simple deconvolution can recover the patterns evoked by each object in an array. Although simple summation without scaling would theoretically achieve the same ends, the scaling step that accompanies averaging preserves linearity under the practical constraint imposed by the limited response ranges of individual neurons. It is no surprise, then, that the same rule is encountered repeatedly throughout the visual hierarchy (Desimone and Duncan, 1995; MacEvoy *et al.*, 2009; Zoccolan *et al.*, 2005). But rather than as a result of interstimulus "competition" for representational resources, as some have previously suggested (Kastner *et al.*, 1998), we see this rule as the outcome of a "cooperative" scheme aimed at preserving as much information as possible about each stimulus in an array.

This result gives us a novel framework in which to understand several perceptual phenomena that affect multiobject arrays such as scenes. For instance, consider change blindness (Rensink *et al.*, 1997; Simons and Rensink, 2005). This striking phenomenon is observed when two versions of a scene are alternated

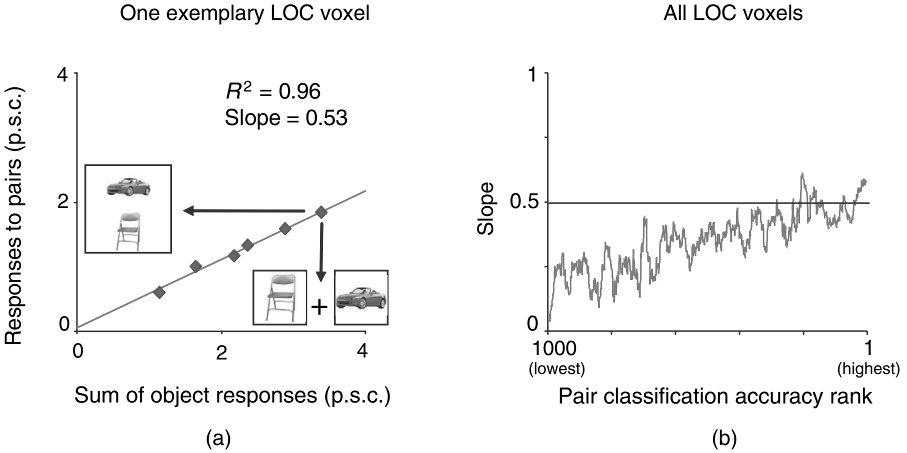

Figure 12.4 Relationship between single-object and paired-object responses in the lateral occipital complex. (a) Data from a single exemplary voxel, illustrating the averaging rule (p.s.c., percentage signal change). The response to each of the six object pairs is plotted against the sum of the responses to each of the constituent objects when these objects are presented alone. The data are well fitted by a straight line with a slope of 0.53, indicating that the pair response is approximately the average of the two single-object responses. (b) Data from across the LOC. We predicted that the linear relationship observed in (a) would be most evident in voxels from subregions of the LOC that were most informative about pair identity. We used a "searchlight" analysis to identify such subregions. For each voxel in the LOC, a 5 mm spherical neighborhood (or "searchlight") was defined, within which two quantities were calculated: (i) the pair-classification performance based on the voxelwise pattern within the sphere, and (ii) the mean slope of the regression line relating the pair and single-object responses, as in (a). The graph shows the slope plotted as a function of pair-classification performance, ranked from the worse-performing to the best-performing voxels. The slope approaches 0.5 for voxels located within the most informative subregions. A color version of this figure can be found on the publisher's website (www.cambridge.org/9781107001756).

on a screen. Even though the two versions of the scene might have quite substantial visual differences, subjects will take a surprisingly long time to notice these differences if low-level visual transients are camouflaged, for example by an intervening blank screen. The subjective impression is that the scene is "the same" every time – at least until the area of difference is specifically noted as a distinct and changing scene element. We hypothesize that this phenomenon occurs for two reasons. First, the scene representation in the PPA takes little note of individual elements; hence, as far as the PPA is concerned, the scene really is “the same” every time. Second, because the representation of the scene in the LOC is the average of the representations of the individual objects, changing any single object will have a negligible effect on the overall activity pattern. In much the same way, perceptual crowding can be seen as the outcome of the visual system’s attempt to preserve information about each element of an object array. At a certain array size, however, the averaging strategy that works well for small numbers of objects produces a pattern of population activity that is indistinguishable from noise, and the identity of individual elements can no longer be discerned. In both change blindness and crowding, veridical perception can be rescued by attention, which in this framework is a mechanism evolved to combat the signal-to-noise penalty caused by using response averaging (albeit with its own price of vastly reduced sensitivity to unattended objects).

Returning to the issue of rapid scene recognition, if we consider the simple object arrays that we used in our experiment to be the very simplest forms of scenes (even if the PPA does not see them that way), then in our results we perhaps have a potential physiological correlate of gist: a pattern that, in preserving the identity of each object in a multiobject scene, forms a unique signature of that scene. Two other notable aspects of our results resonate with gist. First, our results were derived under task conditions designed to ensure that attention was distributed evenly between both objects in each pair, mimicking the attentional state of subjects viewing a briefly presented scene in the original gist experiments. Second, in a way that was consistent with the schematic nature of scene gist, the patterns evoked by pairs did not contain information about the relative positions of the objects (each object pair was presented in two spatial configurations, e.g., shoe above brush, and brush above shoe.) Although MVPA could easily distinguish between patterns evoked by different positions of single objects, demonstrating information about absolute stimulus position, our attempts to decode the spatial configurations of pairs yielded chance performance. This finding is consistent with behavioral data indicating that, in the absence of focal attention, it is possible to extract “gist” information relating to object identity from scenes without determining the locations of the individual objects (Evans and Treisman, 2005).

Somewhat surprisingly, then, our neuroimaging results may also allow us to say something more about what the sensation of gist actually is. In particular, if the pattern evoked by a multiple-object scene is linearly related to the patterns evoked by its constituent objects, then gist perception might correspond to an initial hypothesis about the set of objects contributing to this pattern (but not necessarily the recognition of each or any one of those objects) and a judgment about the category of scene that is most likely to contain these objects. In other words, gist perception does not need to follow object recognition, but could be a parallel assessment of the pattern evoked by multiple simultaneous objects, a pattern which independently feeds object recognition. This assessment of the object-related pattern within the LOC might also proceed in parallel with an assessment of scene layout in the PPA, with both analyses providing information about scene category.

12.6 Conclusions

The basic fact that human observers can rapidly and accurately identify complex visual scenes has been known for over 30 years. Despite this, the study of the cognitive and neural mechanisms underlying visual scene perception is just beginning. Behavioral work strongly supports the idea that scenes are recognized in part through analysis of whole-scene properties such as layout; complementarily to this, neuropsychological and neuroimaging data point to specific brain regions such as the parahippocampal place area in the mediation of these analyses. Gist perception might also involve rapid analysis of the objects within the scene, perhaps through extraction of a summary or mean signal processed by the lateral occipital and fusiform regions. We expect that our understanding of the neural basis of scene recognition will advance rapidly in the next few years through the deployment of advanced fMRI data analysis techniques such as MVPA and fMRI adaptation, which can be used to isolate the representations that underlie these abilities. Ultimately, we believe that it will be possible to develop a theory of scene recognition that ties together multiple explanatory levels, from the underlying single-neuron physiology, through systems neuroscience, to cognitive theory and behavioral phenomena.

Acknowledgments

We thank Emily Ward for comments on the manuscript and assistance with figures. We also thank the many colleagues who contributed to the work described here. This work was supported by the National Institutes of Health (EY-016464 to RE) and the National Science Foundation (SBE-0541957).

References

Abbott, L. F., Varela, J. A., Sen, K., and Nelson, S. B. (1997). Synaptic depression and cortical gain control. *Science*, 275: 220–224.

Aguirre, G. K. and D'Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain*, 122: 1613–1628.

Aguirre, G. K., Zarahn, E., and D’Esposito, M. (1998). An area within human ventral cortex sensitive to “building” stimuli: evidence and implications. *Neuron*, 21: 373–383.

Antes, J. R., Penland, J. G., and Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychol. Res.*, 43: 277–292.

Bar, M. (2004). Visual objects in context. *Nature Rev. Neurosci.*, 5: 617–629.

Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177: 77–80.

Biederman, I., Rabinowitz, J. C., Glass, A. L., and Stacy, E. W., Jr. (1974). On the information extracted from a glance at a scene. *J. Exp. Psychol.*, 103: 597–600.

Blakemore, C. and Nachmias, J. (1971). The orientation specificity of two visual after-effects. *J. Physiol.*, 213: 157–174.

Boynton, G. M. and Finney, E. M. (2003). Orientation-specific adaptation in human visual cortex. *J. Neurosci.*, 23: 8781–8787.

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

Cant, J. S. and Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cereb. Cortex*, 17: 713–731.

Cheng, K. (1986). A purely geometric module in the rats spatial representation. *Cognition*, 23: 149–178.

Cheng, K. and Newcombe, N. S. (2005). Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychon. Bull. Rev.*, 12: 1–23.

Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.*, 18: 193–222.

Diana, R. A., Yonelinas, A. P. and Ranganath, C. (2008). High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. *Hippocampus*, 18: 536–541.

DiCarlo, J. J. and Cox, D. D. (2007). Untangling invariant object recognition. *Trends Cogn. Sci.*, 11: 333–341.

Drucker, D. M. and Aguirre, G. K. (2009). Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cereb. Cortex*, 19(10): 2269–2280.

Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.*, 12: 388–396.

Epstein, R. A. and Higgins, J. S. (2007). Differential parahippocampal and retrosplenial involvement in three types of visual scene recognition. *Cereb. Cortex*, 17: 1680–1693.

Epstein, R. and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392: 598–601.

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23: 115–125.

Epstein, R., DeYoe, E. A., Press, D. Z., Rosen, A. C., and Kanwisher, N. (2001). Neuropsychological evidence for a topographical learning mechanism in parahippocampal cortex. *Cogn. Neuropsychol.*, 18: 481–508.

Epstein, R., Graham, K. S., and Downing, P. E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron*, 37: 865–876.

Epstein, R. A., Higgins, J. S., and Thompson-Schill, S. L. (2005). Learning places from views: variation in scene processing as a function of experience and navigational ability. *J. Cogn. Neurosci.*, 17: 73–83.

Epstein, R. A., Higgins, J. S., Jablonski, K., and Feiler, A. M. (2007a). Visual scene processing in familiar and unfamiliar environments. *J. Neurophysiol.*, 97: 3670–3683.

Epstein, R. A., Parker, W. E., and Feiler, A. M. (2007b). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *J. Neurosci.*, 27: 6141–6149.

Epstein, R. A., Parker, W. E., and Feiler, A. M. (2008). Two kinds of FMRI repetition suppression? Evidence for dissociable neural mechanisms. *J. Neurophysiol.*, 99: 2877–2886.

Evans, K. K. and Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *J. Exp. Psychol. Hum. Percept. Perform.*, 31: 1476–1492.

Fang, F., Murray, S. O., Kersten, D., and He, S. (2005). Orientation-tuned FMRI adaptation in human visual cortex. *J. Neurophysiol.*, 94: 4188–4195.

Fang, F., Murray, S. O., and He, S. (2007). Duration-dependent FMRI adaptation and distributed viewer-centered face representation in human visual cortex. *Cereb. Cortex*, 17: 1402–1411.

Fei-Fei, L., Iyer, A., Koch, C., and Perona, P. (2007). What do we perceive in a glance of a real-world scene? *J. Vis.*, 7: 10.

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, C. R. and King, A. P. (2009). *Memory and the Computational Brain: Why Cognitive Science Will Transform Neuroscience*. Chichester, UK; Malden, MA: Wiley-Blackwell.

Ganel, T., Gonzalez, C. L., Valyear, K. F., Culham, J. C., Goodale, M. A., and Kohler, S. (2006). The relationship between fMRI adaptation and repetition priming. *Neuroimage*, 32: 1432–1440.

Gonsalves, B. D., Kahn, I., Curran, T., Norman, K. A., and Wagner, A. D. (2005). Memory strength and repetition suppression: multimodal imaging of medial temporal cortical contributions to recognition. *Neuron*, 47: 751–761.

Greene, M. R. and Oliva, A. (2009). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.*, 58: 137–176.

Grill-Spector, K. and Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol.*, 107: 293–321.

Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn. Sci.*, 10: 14–23.

Habib, M. and Sirigu, A. (1987). Pure topographical disorientation – a definition and anatomical basis. *Cortex*, 23: 73–85.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293: 2425–2430.

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

Henson, R. N. (2003). Neuroimaging studies of priming. *Prog. Neurobiol.*, 70: 53–81.

Hermer, L. and Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, 370: 57–59.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160: 106–154.

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310: 863–866.

Intraub, H. and Dickinson, C. A. (2008). False memory 1/20th of a second later: what the early onset of boundary extension reveals about perception. *Psychol. Sci.*, 19: 1007–1014.

Intraub, H., Bender, R. S., and Mangels, J. A. (1992). Looking at pictures but remembering scenes. *J. Exp. Psychol.: Learn. Mem. Cogn.*, 18: 180–191.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., and Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proc. Natl. Acad. Sci. USA*, 96: 9379–9384.

Kamitani, Y. and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neurosci.*, 8: 679–685.

Kastner, S., De Weerd, P., Desimone, R., and Ungerleider, L. C. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, 282: 108–111.

Kim, M., Ducros, M., Carlson, T., Ronen, I., He, S., Ugurbil, K., and Kim, D.-S. (2006). Anatomical correlates of the functional organization in the human occipitotemporal cortex. *Magn. Reson. Imaging*, 24: 583–590.

Kourtzi, Z. and Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, 293: 1506–1509.

Maccotta, L. and Buckner, R. L. (2004). Evidence for neural effects of repetition that directly correlate with behavioral priming. *J. Cogn. Neurosci.*, 16: 1625–1632.

MacEvoy, S. P. and Epstein, R. A. (2007). Position selectivity in scene- and object-responsive occipitotemporal regions. *J. Neurophysiol.*, 98: 2089–2098.

MacEvoy, S. P. and Epstein, R. A. (2009a). Building scenes from objects: a distributed pattern perspective. In *Neuroscience Meeting Planner*, Program No. 262.29. Chicago, IL: Society for Neuroscience. Online.

MacEvoy, S. P. and Epstein, R. A. (2009b). Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.,* 19: 943–947.

MacEvoy, S. P., Tucker, T. R., and Fitzpatrick, D. (2009). A precise form of divisive suppression supports population coding in the primary visual cortex. *Nature Neurosci.*, 12: 637–645.

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 vortex. *Proc. Natl. Acad. Sci. USA*, 92: 8135–8139.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

Mendez, M. F. and Cherrier, M. M. (2003). Agnosia for scenes in topographagnosia. *Neuropsychologia*, 41: 1387–1395.

Miller, E. K., Li, L., and Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term-memory task. *J. Neurosci.*, 13: 1460–1478.

Morgan, L. K., MacEvoy, S. P., Aguirre, G. K., and Epstein, R. A. (2009). Decoding scene categories and individual landmarks from cortical response patterns. In *2009 Neuroscience Meeting Planner*, Program No. 262.8. Chicago, IL: Society for Neuroscience. Online.

Muller, J. R., Metha, A. B., Krauskopf, J., and Lennie, P. (1999). Rapid adaptation in visual cortex to the structure of images. *Science*, 285: 1405–1408.

Pallis, C. A. (1955). Impaired identification of faces and places with agnosia for colours – report of a case due to cerebral embolism. *J. Neurol. Neurosurg. Psychiatry*, 18: 218–224.

Park, S. and Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage*, 47: 1747–1756.

Peelen, M. V., Fei-Fei, L., and Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, 460: 94–97.

Potter, M. C. (1975). Meaning in visual search. *Science*, 187: 965–966.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *J. Exp. Psychol.: Hum. Learn. Mem.*, 2: 509–522.

Potter, M. C. and Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *J. Exp. Psychol.* 81: 10–15.

Renninger, L. W. and Malik, J. (2004). When is scene identification just texture recognition? *Vis. Res.*, 44: 2301–2311.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.*, 8: 368–373.

Sasaki, Y., Rajimehr, R., Kim, B. W., Ekstrom, L. B., Vanduffel, W., and Tootell, R. B. (2006). The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. *Neuron*, 51: 661–670.

Sawamura, H., Orban, G. A., and Vogels, R. (2006). Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the FMRI adaptation paradigm. *Neuron*, 49: 307–318.

Schyns, P. G. and Oliva, A. (1994). From blobs to boundary edges: evidence for time- and spatial-scale-dependent scene recognition. *Psychol. Sci.*, 5: 195–200.

Shelton, A. L. and Pippitt, H. A. (2007). Fixed versus dynamic orientations in environmental learning from ground-level and aerial perspectives. *Psychol. Res.*, 71: 333–346.

Simons, D. J. and Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends Cogn. Sci.*, 9: 16–20.

Steeves, J. K., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., and Goodale, M. A. (2004). Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *J. Cogn. Neurosci.*, 16: 955–965.

Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, 262: 685–688.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381: 520–522.

Tsao, D. Y. and Livingstone, M. S. (2008). Mechanisms of face perception. *Annu. Rev. Neurosci.*, 31: 411–437.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B., and Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science*, 311: 670–674.

Vuilleumier, P., Henson, R. N., Driver, J., and Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature Neurosci.*, 5: 491–499.

Walther, D. B., Caddigan, E., Fei-Fei, L., and Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.*, 29: 10573–10581.

Wig, G. S., Grafton, S. T., Demos, K. E., and Kelley, W. M. (2005). Reductions in neural activity underlie behavioral components of repetition priming. *Nature Neurosci.*, 8: 1228–1233.

Wiggs, C. L. and Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.*, 8: 227–233.

Yamane, Y., Carlson, E. T., Bowman, K. C., Wang, Z., and Connor, C. E. (2008). A neural code for three-dimensional object shape in macaque inferotemporal cortex. *Nature Neurosci.*, 11(11), 1352–1360.

Zoccolan, D., Cox, D. D., and DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *J. Neurosci.*, 25: 8150–8164.