



# HHS Public Access

Author manuscript

*Annu Rev Vis Sci.* Author manuscript; available in PMC 2020 January 29.

Published in final edited form as:

*Annu Rev Vis Sci.* 2019 September 15; 5: 373–397. doi:10.1146/annurev-vision-091718-014809.

## Scene perception in the human brain

Russell A. Epstein<sup>1</sup>, Chris I. Baker<sup>2</sup>

<sup>1</sup>Department of Psychology, University of Pennsylvania, Philadelphia PA 19104 USA,

<sup>2</sup>Section on Learning and Plasticity, Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda MD 20892 USA,

### Abstract

Humans are remarkably adept at perceiving and understanding complex real-world scenes. Uncovering the neural basis of this ability is an important goal of vision science. Neuroimaging studies have identified three cortical regions that respond selectively to scenes: parahippocampal place area (PPA), retrosplenial complex/medial place area (RSC/MPA), and occipital place area (OPA). Here, we review what is known about the visual and functional properties of these brain areas. Scene-selective regions exhibit retinotopic properties and sensitivity to low-level visual features that are characteristic of scenes. They also mediate higher-level representations of layout, objects, and surface properties that allow individual scenes to be recognized and their spatial structure ascertained. Challenges for the future include developing computational models of information processing in scene regions, investigating how these regions support scene perception under ecologically realistic conditions, and understanding how they operate in the context of larger brain networks.

### Keywords

Functional magnetic resonance imaging; visual cortex; visual recognition; spatial navigation; hippocampus; neural networks

### Introduction

What is a scene, that an observer might know it? Henderson and Hollingworth (1999) offer the following definition: “a scene is a semantically coherent (and often namable) view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner.” Under this definition, scene perception can be usefully contrasted to object perception: whereas objects are spatially compact entities that one acts upon, scenes are spatially distributed entities that one acts within (Epstein 2005).

There are several reasons why an organism might care about scenes, and why the visual system might have specialized systems for processing them (Malcolm et al. 2016). Scenes are locations in the world, so one may want to identify a scene as a particular place or a particular kind of place. During navigation, one might want to understand the spatial

structure of a scene so that one can orient oneself relative to it, or plan a path through it. Objects always appear in scenes, therefore it can be useful to analyze the scene in order to know where to search for an object, or to provide clues as to an object's identity. As these considerations indicate, scene perception is ecologically important, and it is not surprising that humans are remarkably good at it: we understand landscapes, cityscapes, and rooms just as readily as we understand faces, bodies, animals, and tools.

The study of scene perception goes back over 50 years, to the seminal contributions of Biederman (1972) and Potter (1975). Over the past 20 years, this psychological work has been complemented by a growing line of neuroscience research, which was initially sparked by the discovery of the parahippocampal place area (PPA), a ventral pathway region that responds strongly in fMRI studies when participants view scenes (landscapes, cityscapes, and rooms), but less strongly when they view objects (faces, bodies, artifacts). Subsequent studies identified two other brain regions that exhibit a scene-selective response: one in the medial parietal/retrosplenial region, and another in the dorsal occipital lobe (Figure 1a). In this review, we describe these three scene-responsive regions, discuss their visual and functional properties, and highlight some recent directions that we think are likely to provide particularly fruitful avenues for future research.

## Delineating the scene network

### Parahippocampal Place Area (PPA)

The preferential response to scenes in posterior parahippocampal/anterior lingual region was first reported by Epstein and Kanwisher (1998), who labelled this region the parahippocampal place area (PPA). Two contemporaneous reports indicated a preferential response to buildings in a similar cortical locus (Aguirre et al. 1998, Ishai et al. 1999). The PPA typically includes portions of the posterior parahippocampal, anterior lingual, and medial fusiform gyri. A recent probabilistic study using cortex-based alignment indicated that the strongest scene selectivity can be reliably localized to the junction of the collateral sulcus and the anterior lingual sulcus (Weiner et al. 2018). Relative to the ventral pathway as a whole, the PPA appears to be part of a medial substream that shows a functional distinction from the more lateral aspects on a number of different dimensions (e.g. medial vs. lateral preference for inanimate vs. animate objects, large vs. small objects, and places/scene vs. faces/bodies) (Konkle & Caramazza 2013, Kravitz et al. 2013). These functional distinctions might be driven by differences in cytoarchitecture as well as differential anatomical connectivity to early visual areas that process (respectively) the periphery vs. center of the visual field (Weiner et al. 2017).

In monkeys, two separate scene regions—the “lateral place patch” (LPP) and the “medial place patch” (MPP)—have been identified in the occipitotemporal sulcus and medial parahippocampal gyrus, respectively, which, together, might form the homologue of the PPA (Kornblith et al. 2013, Nasr et al. 2011). Although only the LPP responds in a scene-preferential manner during fMRI, scene-selective neurons were observed in both regions during neurophysiological recordings, and multiunit classification methods suggest that both regions contain more information about scenes than they do about objects (Kornblith et al., 2013). Scene-selective neurons have also been identified by intracranial recording in

parahippocampal cortex in humans (Mormann et al. 2017). Consistent with neuroimaging and recording studies, direct electrical stimulation of the PPA through intracranial electrodes can elicit hallucinatory images of scenes (Megevand et al. 2014).

### **Retrosplenial Complex/Medial Place Area (RSC/MPA)**

The original 1998 PPA paper noted the existence a second locus of scene-preferential response in some participants, which was described as being in “anterior calcarine cortex”. O’Craven and Kanwisher (2000) obtained more reliable evidence for this second scene-responsive region, which, like the PPA, was active during both perception and mental imagery of scenes. Subsequent studies localized this scene-preferential territory to the medial-parietal/retrosplenial region, along the banks of the parietal-occipital sulcus, and labelled it the retrosplenial complex (RSC). This has led to some confusion, because this functionally-defined scene region is *not* equivalent to retrosplenial cortex, which is a cytoarchitectonically defined region (comprising Brodmann Areas 29 and 30). Indeed, a recent study from one of our labs suggests that the scene-selective territory that is often labelled RSC might not overlap with BA29/30 at all (Silson et al. 2016b). Thus, the more neutral term medial place area (MPA) may be a more useful nomenclature to describe the scene-responsive functional territory. Here, we use the abbreviation RSC/MPA. In monkeys, a possible homologue of RSC/MPA has been identified along the parietal-occipital sulcus, but no recordings have been made from this region and its functional properties beyond its scene-selective response are currently unknown (Kornblith et al. 2013, Nasr et al. 2011).

### **Occipital Place Area (OPA)**

A third scene-responsive locus was first identified in the dorsal occipital lobe using PET (Nakamura et al. 2000) and later confirmed with fMRI (Grill-Spector 2003, Hasson et al. 2003). Initially, this region was labelled TOS, based on its proximity to the transverse occipital sulcus, but subsequent doubts about the accuracy of this anatomical localization led to it being renamed the occipital place area (OPA) (Dilks et al. 2013). A putative homologue has been identified in the dorsal occipital lobe of the macaque, near the prelunate gyrus (Arcaro & Livingstone 2017, Kornblith et al. 2013, Nasr et al. 2011), but like the homologue of MPA/RSC, its functional properties have not been explored.

### **Organization of the scene network**

Functional connectivity studies suggest that the PPA, RSC/MPA, and OPA are strongly connected to each other. Recent analyses of resting state data suggest that this scene “network” might be further fractionated into anterior and posterior subnetworks (Figure 1b): there is a significant amount of functional connectivity between the posterior PPA, posterior RSC/MPA, and OPA on the one hand, and between the anterior PPA, anterior RSC/MPA and adjoining territory in posterior cingulate cortex, and a region adjoining OPA in the caudal inferior parietal lobe on the other (Baldassano et al. 2016a, Nasr et al. 2013, Silson et al. 2016b). The first set of regions might be involved in the perceptual analysis of scenes, whereas the second set of regions might be more involved in spatial processing and memory.

Note that we refer to PPA, RSC/MPA, and OPA as “scene regions” because they respond strongly to scenes and appear to have an important role in scene processing. Our use of this

term in the context of this review does not imply that we believe that these regions are exclusively involved in scene processing. Indeed, there is considerable evidence that PPA and RSC/MPA, at least, are involved in other cognitive functions, such as memory, imagination, and semantic knowledge retrieval (Ranganath & Ritchie, 2012). As vision scientists, we are interested in scene regions because we want to understand scene perception, but researchers with other concerns might consider the literature on scene perception as just one source of evidence about the function of these parts of the brain.

## Mapping visual responses across the scene network

One of the defining characteristics of visual cortex is the presence of retinotopic organization: adjacent points on the cortex respond to stimulation from adjacent points in visual space. For scene-selective regions, a key question is the extent to which they exhibit any retinotopic properties or organization. A second important question is to what extent their activation profiles—including their preferential response to scenes—can be explained in terms of sensitivity to elementary features of the visual images such as spatial frequency, line orientation, and contrast.

### Retinotopic organization of scene-selective cortex

Studies using population receptive field (pRF) mapping have shown that PPA, OPA and RSC/MPA exhibit retinotopic biases in terms of visual field (contralateral, ipsilateral), eccentricity (fovea, periphery) and elevation (upper, lower) (Figure 1a) (Groen et al. 2017). For RSC/MPA, retinotopic properties appear to be restricted to more posterior regions in the parieto-occipital sulcus (Silson et al. 2016b) just anterior to peripheral V1 and dorsal V2 (Elshout et al. 2018, Nasr et al. 2011) and this may explain why initial studies found only weak evidence for retinotopy in this region (MacEvoy & Epstein 2007, Ward et al. 2010).

Responses in all three scene-selective regions are stronger to stimuli presented in the contralateral versus ipsilateral hemifield (MacEvoy & Epstein 2007) and disruption of processing in OPA with transcranial magnetic stimulation (TMS) biases eye movements away from the contralateral visual field (Malcolm et al. 2018). In terms of eccentricity, PPA is embedded within an eccentricity gradient in ventral temporal cortex that varies from foveal responses more laterally (where face-selective cortex is located) to peripheral responses more medially (where PPA is located) (Hasson et al. 2003, Levy et al. 2001). A similar preference for more peripheral parts of the visual field is also observed in OPA (Silson et al. 2015, Silson et al. 2016a) and RSC/MPA (Silson et al. 2016b), and all three scene-selective regions show greater functional connectivity with peripheral (> 5 degrees) than foveal V1 (Baldassano et al. 2016b). Further, consistent with the increase in pRF size with eccentricity in early visual cortex, pRFs in the scene-selective regions are larger than those in more foveally-biased cortex (Grill-Spector et al. 2017, Silson et al. 2015). Finally, in terms of elevation, OPA shows a bias for the lower visual field and PPA for the upper visual field (Silson et al. 2015), but there appears to be no elevation bias in RSC/MPA (Silson et al. 2016b). Similar biases for the upper and lower visual field have also been reported in the monkey, with the LPP showing an upper-field bias and the putative homologue of the OPA in dorsal extrastriate cortex showing a lower field bias (Arcaro & Livingstone 2017).

Consistent with these findings of retinotopic responses, multiple retinotopic maps overlap with scene-selective regions (Figure 1a). Notably, however, there is no one-to-one relationship between scene regions and any given retinotopic map. For example, PPA overlaps to varying proportions with maps termed VO-2, PHC-1 and PHC-2 (Arcaro et al. 2009, Silson et al. 2015). Similarly, OPA overlaps dorsal V3, V3A, V3B, LO-1 and LO-2, with some parts of OPA located outside of any currently identified maps (Bettencourt & Xu 2013, Nasr et al. 2011, Silson et al. 2016a). For RSC/MPA, recent work has suggested the presence of a retinotopic map in a similar cortical location (putatively V2A), although the precise relationship to RSC/MPA is unclear (Elshout et al. 2018). Similar overlap with retinotopic maps has been reported in monkeys, with the LPP overlapping maps OTS-1 and OTS-2, and the dorsal scene-selective region overlapping DP, V3A and part of dorsal V3 (Arcaro & Livingstone 2017).

The retinotopic biases observed in scene regions may have functional implications. The relatively large peripheral pRFs might make the scene regions sensitive to large scale summary statistics or gist of the input that could enable rapid scene recognition (Boucart et al. 2013, Larson & Loschky 2009). The differential biases for the upper and lower visual field in PPA and OPA, respectively, might make them sensitive to different aspects of scenes. For example, large, immovable objects that may serve as landmarks are more likely to occupy the upper visual field and PPA responds strongly to these types of stimuli. In contrast, the surface on which we are moving is likely to occur in the lower visual field, and OPA responses reflect the navigational affordances of scenes (Bonner & Epstein 2017, Bonner & Epstein 2018).

### Visual features modulating responses in scene-selective cortex

Although the scene regions are defined by the comparison of scenes versus non-scene categories, many low-level visual features differ between these categories and have been shown to modulate responses in these regions. For example, the scene regions are more responsive to high than low spatial frequencies (e.g. Rajimehr et al. 2011), and manipulating spatial frequency content changes the information available in multivoxel response patterns (Berman et al. 2017, Watson et al. 2016). Other visual properties that modulate responses in scene regions include rectilinearity (Nasr et al. 2014) and the overall orientation distribution, with stronger responses to images with cardinal rather than oblique orientations (Nasr & Tootell 2012). Even without presenting scenes, differential responses can be observed to minimal stimuli (e.g. geometric shapes) that differ in spatial frequency and rectilinearity (Nasr et al. 2014, Rajimehr et al. 2011). Finally, the temporal dynamics of scene processing also suggest an influence of low level features. Intracranial recordings from PPA show a distinction between scene and non-scene images emerging within 100 ms, which might reflect a rapid feedforward analysis of global scene features (Bastin et al. 2013). In EEG, summary image statistics influence event-related potential amplitudes from 100 ms up to 300 ms post-stimulus, suggesting that low-level visual features also influence later stages of scene processing (Groen et al. 2016, Groen et al. 2013, Harel et al. 2016).

Can the apparent category selectivity of scene regions be explained in terms of sensitivity to low-level visual features? The weight of the evidence tends to argue against this idea. For

example, scene selectivity remains when rectilinearity is controlled for by matching scene and face stimulus sets (Bryan et al. 2016) or when low level featural differences are tightly controlled between stimuli perceived as a scene or not (Schindler & Bartels 2016). When considering the contribution of low level visual features to responses in scene regions, it is important to note that high level aspects of scenes (e.g. category) and low level features (e.g. spatial frequency) are inextricably linked (Groen et al. 2017). For example, forest scenes are typically characterized by high spatial frequency, and cityscapes will have a high degree of rectilinearity. Thus, it is not surprising that features such as spatial frequency and rectilinearity modulate responses. More important is understanding the relationship between low level features and the higher level scene information that is used to guide behavior, a topic we take up later in this paper.

## Information processing in the scene network

Many cognitive functions have been attributed to the scene network, including scene recognition, spatial perception, spatial navigation, and guidance of visual search. To perform such functions, these regions must extract representations of higher-level aspects of scenes, such as spatial layout, scene category, place identity, spatial location and heading (Figure 2). The extraction of such high-level qualities is often thought to be one of the central goals of vision, as this is the kind of information that is necessary for vision to interact with long-term knowledge such as semantic categories or cognitive maps. Here, we survey studies that have probed mid- and high-level representations within the scene network that relate to cognitive function.

### What are the stimuli and tasks that activate the scene regions?

The PPA responds strongly to images that convey information about local spatial layout, even if the depicted environment is devoid of discrete objects (e.g., an empty room, consisting of just walls, ceiling, and floor, or an empty landscape) (Epstein & Kanwisher 1998), and even if the “scene” is nothing more than a layout made of Lego blocks (Epstein et al. 1999) perceived either visually or haptically (Wolbers et al. 2011). Similar effects are observed in RSC/MPA and OPA (Kamps et al. 2016, Wolbers et al. 2011). Consistent with these fMRI results, individual neurons in parahippocampal cortex respond significantly more strongly to images that have an interpretable spatial background than to images that have an uninterpretable background or no background (Mormann et al. 2017). These findings suggest a sensitivity to the spatial structure of the stimulus (e.g. “scene-like” vs. “object-like”). Building on this general notion, fMRI responses in scene regions are sensitive to the size (Park et al. 2015), openness (Henderson et al. 2011), distance (Henderson et al. 2008) and coherence (Epstein & Kanwisher 1998, Kamps et al. 2016) of the depicted space, and to the height of space-defining boundaries (Ferrara & Park 2016).

Additional insight into the functions of scene regions comes from examination of their submaximal activation to non-scene stimuli. In PPA, there is a reliable hierarchy of responses, with high response to images of buildings cut out from the surrounding environment, medium response to man-made objects (vehicles, tools), low response to animals and bodies, and virtually no response to faces. This ordering might reflect



sensitivity to higher order properties of the stimuli. Studies have explored different high-level factors, revealing sensitivity of response that relates to the real-world size of objects (Konkle & Oliva 2012), spatial stability (Mullally & Maguire 2011), distance from the viewer (Amit et al. 2012, Cate et al. 2011), interaction envelope (Bainbridge & Oliva 2015) or the extent to which the objects are associated with specific contextual settings (Bar & Aminoff 2003).

Results such as these have led to theories that the role of PPA, RSC/MPA and OPA in scene perception might be conceptualized more broadly in terms of the processing of landmarks (Auger et al. 2015, Epstein & Vass 2014, Troiani et al. 2014), contexts (Aminoff et al. 2013), or spaces (Bainbridge & Oliva 2015, Mullally & Maguire 2011). A challenge for such interpretations is that some high-level factors might be inherently confounded with low- and mid-level visual features discussed in the previous section that are known to affect response in scene regions (Long et al. 2018). For example, large objects might tend to be more rectilinear, while small objects might tend to be more curved (Konkle & Oliva 2012). However, it is unlikely that low-level features can explain all of these high-level effects. Preferential response to large, stable objects is even observed in blind participants making size judgments in response to auditory cues (He et al. 2013). Moreover, effects of landmark status are observed even when this factor is determined by the experience of the viewer, rather than by the category or shape of the object. For example, the PPA responds more strongly to objects that were previously encountered at decision points along a route compared to objects encountered at less navigationally relevant locations (Janzen & van Turenout 2004, Schinazi & Epstein 2010). A similar effect is observed in RSC/MPA for objects that are encountered at fixed rather than variable locations within a virtual maze (Auger et al. 2015).

Turning back to the activation elicited by scenes, several studies have examined how responses in scene regions vary as a function of real-world familiarity with the depicted location (Epstein et al. 1999, Epstein et al. 2007a, Epstein et al. 2007b). Familiarity effects are especially strong in RSC/MPA, where images of familiar locations can elicit 50% more activation than images of unfamiliar locations. In contrast, familiarity effects in PPA and OPA are weaker and less reliable. These results suggest that RSC/MPA might play a more mnemonic role in scene processing, whereas the function of the PPA and OPA may be more perceptual. Consistent with this view, activation in RSC/MPA when viewing scenes is significantly enhanced if the participants are asked to retrieve spatial information, such as where the scene is located within the broader environment or the facing direction of the depicted view (Epstein et al. 2007b), and activation in RSC/MPA increases in tandem with the participants' acquisition of survey knowledge when learning a new environment (Wolbers & Buchel 2005).

### **Representations related to scene recognition**

Although univariate analyses are useful for identifying the cognitive processes supported by a region, they are less informative about the representations that mediate those processes. In this section and the next, we review studies that have used methods such as multivoxel pattern analysis (MVPA), voxelwise encoding models, and fMRI adaptation to understand

the representations supported by scene regions. The present section considers representations that are useful for scene recognition; the subsequent section considers representations useful for spatial perception and navigation.

A scene can be recognized at several different levels of specificity: (i) as a member of a semantic category (e.g. kitchen, forest, beach); (ii) as a specific location, room, or building (e.g. the kitchen on the fifth floor of Goddard Hall on the University of Pennsylvania campus); (iii) as a specific view of a location, room, or building (e.g. the kitchen viewed from the Northeast). Multivoxel activation patterns in PPA, RSC/MPA and OPA distinguish between scenes of different categories, as do those in the object-selective lateral occipital complex (LOC) and early visual cortex (Epstein & Morgan 2012, Walther et al. 2009). Multivoxel activation patterns in PPA, RSC/MPA and OPA also distinguish between individual landmarks, such as buildings on a college campus (Epstein & Morgan 2012, Morgan et al. 2011). Notably, when the visual similarity is partially controlled for by using very different views (interior vs. exterior) of the same landmark in the classification analysis, it is still possible to decode landmark identity in the PPA, RSC/MPA, and OPA, while classification performance in brain regions outside of the scene network falls to chance (Marchette et al. 2015).

What is the nature of the underlying representations that allow scene categories and individual landmarks to be classified in scene regions using MVPA? The possibilities range from low-level features (e.g. scene categories are distinguishable based on their Fourier amplitude spectra; see Oliva & Torralba 2001) to purely abstract semantic, linguistic, or spatial codes. The fact that it is possible to cross-decode between interior and exterior views of the same landmark suggests some degree of abstraction related to identity, particularly in the anterior PPA, where the cross-decoding was related to the participants' amount of experience with the landmark (Marchette et al. 2015). With regards to abstract representations of scene categories, the evidence is equivocal, with one study finding cross-decoding between visual depictions and verbal descriptions in a wide swath of cortex, including PPA, RSC/MPA, and cIPL (Kumar et al. 2017), but another study finding an absence of cross-decoding between visual and auditory scenes (e.g. between a visual image of a beach and the sounds that one would hear on a beach) in the same regions (Jung et al. 2018). Given the strong evidence that PPA, RSC/MPA, and OPA are—at least in their posterior portions—visually responsive regions, we believe that the category and landmark decoding observed in many MVPA studies is most likely driven primarily by between-category and between-landmark differences in visual or shape features.

Some insight into what those features might be comes from studies that have examined responses to scenes presented in different stimulus formats. Scene category can be cross-decoded between multivoxel patterns elicited by color photographs and line drawings in PPA, RSC/MPA, and OPA (Walther et al. 2009); this result has been replicated using multiunit responses in LPP (Kornblith et al. 2013). In PPA and OPA, such cross-decoding appears to be driven primarily by the statistics of contour junctions, as category information in the multivoxel patterns is eliminated by shifting contours in line drawings relative to each other (thus disrupting junctions), but not by rotating the image (thus disrupting orientation statistics) (Figure 3a)(Choo & Walther 2016). These results suggest that category



distinctions in PPA and OPA are driven in part by category-related differences in the 3-d shapes of scenes, as contour junctions are strong cues to this kind of information. Indeed, as we will discuss in the next section, there is strong evidence that scene regions are sensitive to global shape properties of scenes.

Surface properties, such as color, texture and material, might provide another source of information about the category or identity of a scene. For example, a desert is identifiable based on the presence of fine grained yellow sand, and a specific building might be made of wood, concrete, or brick. Scenes with different wall textures elicit distinguishable multivoxel patterns in PPA, even when they are artificial rooms with the same spatial geometry (Park & Park 2017). Moreover, the PPA shows fMRI adaptation across scenes and surfaces with the same surface texture and across 2-d object arrays that have different contours but share the same visual statistics (e.g. two different piles of strawberries) (Cant & Xu 2012). These texture effects are less reliable or unreliable in RSC/MPA and OPA. Consistent with these neuroimaging results, the firing rates of cells in the LPP and MPP are modulated by differences in texture in addition to being modulated by differences in viewing angle and distance (Kornblith et al. 2013).

A third source of information about the category or identity of a scene comes from the objects that the scene contains. A beach umbrella, for example, likely indicates that one is looking at a beach, while a computer monitor suggests that one is looking at an office. For this reason, it is perhaps not surprising that many studies have reported decoding of scene category in object-selective LOC (Choo & Walther 2016, Jung et al. 2018, Walther et al. 2009). The multivoxel patterns elicited by scenes in LOC are linearly related to the multivoxel patterns elicited by their constituent objects, suggesting that LOC “constructs” scene representations out of object representations (MacEvoy & Epstein 2011). Moreover, in a study that used an encoding model approach, Stansbury and colleagues (Stansbury et al 2013) demonstrated that it was possible to predict the voxelwise responses to scenes throughout high-level visual cortex, including PPA, RSC/MPA, OPA, and LOC, based on the scenes’ membership in artificial “categories” that were defined based on within-scene objects. When viewing artificial scenes consisting of walls and a single focal object, PPA encodes information about both the object and the shape of the room (e.g. open vs. closed), while LOC encodes information about the object alone and RSC/MPA encodes information about the shape alone (Harel et al. 2012). This suggests that PPA might form a unified representation of a scene that incorporates object information obtained from LOC and information about the shape of space obtained from RSC/MPA (and possibly OPA—see next section).

Beyond category and scene/landmark identity, a number of studies have examined coding of individual scene views. These studies have typically used fMRI adaptation, which tends to be a more fine-grained tool for examining representations than MVPA (Drucker & Aguirre 2009, Epstein & Morgan 2012, Hatfield et al. 2016). These studies have revealed adaptation effects in PPA, RSC/MPA, and OPA that are specific to individual views (Epstein et al. 2003, Epstein et al. 2007a), consistent with the sensitivity to viewpoint observed in LPP and MPP neurons (Kornblith et al. 2013), as well as adaptation effects that generalize across views (Epstein et al. 2007a, Morgan et al. 2011). In PPA, cross-adaptation is observed

between views of the same scene taken at different distances (Persichetti & Dilks 2016) and also between mirror-reversed images of the same scene (Dilks et al. 2011), suggesting this region is somewhat indifferent to manipulations that preserve the identity and the intrinsic spatial structure of the scene, consistent with a putative role in recognition. In contrast, OPA and RSC/MPA show recovery from adaptation in these cases, suggesting that they are more sensitive to the spatial relationship between the scene and the viewer. Other experiments, however, have found a greater degree of viewpoint-invariance in RSC/MPA compared to PPA (Park & Chun 2009). Some of these apparently conflicting results might be attributed to the use of different adaptation paradigms (Epstein et al. 2008), as the relationship between adaptation effects and underlying neural representations is not well understood (Epstein & Morgan 2012, Hatfield et al. 2016).

### Representations related to spatial perception and navigation

Scenes are—by definition—spaces. Thus, it is not surprising that many studies have focused on how the PPA, RSC/MPA, and OPA encode spatial information. These studies have identified representations of the spatial structure of the local scene (“vista space”) and also representations of the broader space extending beyond the current perceptual horizon (“environmental space”).

A key concept is the spatial layout of scenes. Broadly speaking, layout is the spatial organization of the elements of the scene. An important component of layout is the arrangement of fixed surfaces such as walls and ground planes, which defines the geometric shape of the scene. A long line of research in animal behavior and developmental psychology suggests that geometric information is crucial for spatial orientation (Cheng & Newcombe 2005, Gallistel 1990, Julian et al. 2018b, Lee 2017) and behavioral evidence suggests the shape of the scene is one of several global features used for scene recognition (Greene & Oliva 2009). As noted previously, there is considerable evidence that scene regions respond strongly to the presence of geometry-defining boundary surfaces (Epstein & Kanwisher 1998, Ferrara & Park 2016).

MVPA studies have shown that scene regions represent the spatial layout of scenes. In PPA, for example, scene-evoked patterns are determined primarily by the shape of the scene (open vs. closed) and the distance to the scene surfaces (near vs. far), rather than the content of the scene (urban vs. natural) (Figure 3b)(Kravitz et al. 2011, Park et al. 2011). Patterns in RSC/MPA exhibit similar sensitivity to the shape of the scene (Harel et al. 2012, Park et al. 2011) and to the size of the depicted space (large vs. small) (Park et al. 2015), and shape coding is also observed in OPA (Kravitz et al. 2011). In a recent study, Lescroart and Gallant (2019) found that voxelwise responses in all three regions could be predicted based on the histogram of surfaces distances and orientations within a scene (Figure 3c). This 3-d structural model explained unique variance in the responses that could not be explained by low-level models based on 2-d orientation and spatial frequency. Consistent with results from earlier studies, the first principal component of the 3-d structural model weights reflected the distances to the surfaces in the scene, while the second PC reflected the degree of openness.

What about more specific representations of environmental shape, such as whether a room is square, rectangular, round, or L-shaped? Surprisingly, this issue has not been investigated with MVPA, though a recent adaptation study found evidence that the PPA is sensitive to length and angle changes in line drawings of scenes, which are potential cues to scene geometry (Dillon et al. 2018). Additional relevant evidence comes from a study that used MVPA to identify representations of the navigational affordances of scenes—where one can move to, and where one’s movement is blocked (Bonner & Epstein 2017). The study used both artificial scenes, for which environmental shape was completely controlled and the affordances were defined by the locations of exits, and real-world scenes, for which pathways for movement were determined by a combination of features including environmental boundaries. In both cases, multivoxel patterns in OPA distinguished between scenes based on the direction (left vs. center vs. right) that one could move to egress the scene (Figure 3d). These results are consistent with other findings that implicate OPA in the processing of the spatial structure of scenes (Dilks et al. 2011, Julian et al. 2016, Persichetti & Dilks 2018).

The spatial structure visible in the local scene is just a part of the spatial structure of the broader environment. Among the three scene regions, RSC/MPA appears to be most centrally involved in relating the local scene (vista space) to its surroundings (environmental space) (Byrne et al. 2007, Epstein 2008, Julian et al. 2018b, Vann et al. 2009). Crucial to solving this problem is the ability to represent allocentric (i.e. world-referenced) spatial quantities such as heading (the direction that one is currently facing) and location (one’s position in the world). Several MVPA and adaptation studies have found evidence that RSC/MPA represents heading and location when participants are given tasks that require them to recover spatial information from memory, either in response to scenes or in response to verbal prompts (Baumann & Mattingley 2010, Marchette et al. 2014, Robertson et al. 2016, Shine et al. 2016, Vass & Epstein 2013, Vass & Epstein 2017). The precise quantity observed (heading, location, or both) in RSC/MPA varies across studies, consistent with neurophysiological results that suggest the existence of a flexible and multidimensional spatial code that might manifest itself in fMRI responses in multiple ways depending on the details of the stimuli, environment, and task (Alexander & Nitz 2015). Notably, one study found that location and heading codes in RSC/MPA were anchored to the shape of the local space as defined by fixed boundary elements (Marchette et al. 2014). Such scene-referenced allocentric codes might be essential for mediating between the local scene and the broader environmental space. For further discussion of the role of scene regions in environmental spatial coding, see previous reviews (Epstein et al. 2017, Julian et al. 2018b).

### **Establishing a causal role in function**

The conclusions above are drawn primarily from neuroimaging studies, but they are supported by studies using causal methods. There is an extensive neuropsychological literature examining the effects of neurological insult to the parahippocampal and retrosplenial cortices, which has been reviewed elsewhere (Aguirre & D’Esposito 1999, Epstein 2008, Maguire 2001). Broadly speaking, damage to parahippocampal cortex leads to impairments in recognizing places and landmarks, while damage to the medial parietal region encompassing RSC/MPA leads to deficits in the ability to use scenes and landmarks

to retrieve a heading and localize oneself in space. To our knowledge, there are no reports of patients impaired at recognizing scenes at the categorical level, although this is perhaps not surprising given that scene category can be ascertained through both scene-based and object-based cues, the latter being processed outside of PPA and RSC/MPA. Of note, one patient with extensive LOC damage but preserved PPA retained the ability to recognize scenes at the categorical level even though her object recognition ability was profoundly impaired (Steeves et al. 2004).

The neuropsychological literature on OPA is less well established. Damage to the inferior intraparietal sulcus, adjoining but distinct from OPA, is associated with Balint's syndrome, one of whose primary symptoms is simultanagnosia, an inability to attend simultaneously to multiple elements within a scene (Bettencourt & Xu 2013, Xu & Chun 2009). Neurological insults localized to OPA, on the other hand, have not been reported. However, because OPA is close to the skull, it is an ideal target for transcranial magnetic stimulation (TMS), a technique that allows researchers to create a "virtual lesion" that temporarily disrupts normal information processing. TMS of OPA leads to impairments in the ability to recognize the categories of scenes (Ganaden et al. 2013) and discriminate scenes based on their spatial layout (Dilks et al. 2013). Moreover, a recent TMS study suggested that OPA may be especially involved in the perception of environmental boundaries (Julian et al. 2016). Taken together with the neuroimaging results, these findings suggest that OPA may process visual features that are essential for both scene recognition and spatial perception. We expect that results from studies using causal methods will continue to be important, especially insofar as they guide and constrain our interpretation of data obtained with correlational methods like fMRI and neural recordings.

## The future of scene research: New approaches and conceptualizations

### Computational modelling

As we have emphasized throughout this review, the existence of inherent correlations between low level visual features (e.g. edges, contrast, color), mid-level features (e.g. contour junctions) and high-level abstract properties (e.g. specific place, semantic features) can make it challenging to attribute observed responses to any particular type of representation. To gain traction on this issue, one approach is to test explicit, computable models of the properties that might underlie representations. To date, model-based approaches have been used to investigate representations of navigational affordances (Bonner & Epstein 2018) and scene category (Groen et al, 2018), and have modelled properties such as objects (Stansbury et al. 2013), and surface distances and orientations (Lescroart & Gallant 2019). Here we will focus on issues that future work will need to consider.

First, it is important to compare multiple models since many different models may account for some aspects of the measured responses. One such approach is to use variance partitioning to establish the unique response variance accounted for by each model. For example, using a voxel-wise encoding model approach, Lescroart and colleagues (2015) found that separate models based on the spatial frequency, subjective distance and the categories of objects present could each explain some variance in fMRI responses in scene

regions to a large set of scenes. Importantly, however, when variance partitioning was used to determine the extent to which the different models were accounting for unique or shared components of the response variance, it was revealed that all three models were largely explaining the same variance. A researcher who had explored only one of these models might have been tempted to conclude that it was “correct” when in fact other theoretically distinct models provide equally good accounts of the data.

Second, to optimize the ability to compare different models, studies should consider selecting stimuli that reduce the covariation among model features (see Lescroart et al, 2015 for discussion). One possible approach is to restrict variation of features by testing highly constrained or artificial stimuli. However, this approach runs the risk of generating findings that do not generalize to the broader range of natural scene stimuli. Alternatively, natural stimuli can be sampled in a way that minimizes the covariation across the stimulus features in the models. Such an approach was used to compare object, deep neural network (DNN) and functional models of scene processing (Groen et al. 2018) revealing that responses to scenes in all three scene regions were best explained by the DNN feature model (Figure 4a).

Third, while it is important to test multiple models, there are a huge number of possible models that could be tested and the statistical power to distinguish between model contributions diminishes as the number of models increases. The critical question then becomes how to select the specific models tested. Groen et al. (2018) addressed this issue by selecting models based on a prior large-scale behavioral study that tested many more models (Greene et al, 2016). Alternatively, Lescroart and Gallant (2019) adopted the approach of separately comparing three 3D structural models and three 2D visual feature models and then comparing the best 3D model with the best 2D model.

Finally, DNNs have become a popular model of visual processing given their high levels of performance on object classification tasks and the correspondence between representation in different layers of the DNNs and stages of the ventral visual processing pathway. Representational similarities within a network trained on scene classification (Zhou et al. 2014) have been shown to correspond with representational similarities in the brain observed with both MEG (Cichy et al. 2017) and fMRI (Bonner & Epstein 2018, Groen et al. 2018). Such findings suggest that the DNN is a good model for scene processing. However, it can be challenging to understand the internal operations and theoretical principles that account for the DNN-brain correspondences. One possible approach is to run *in silico* experiments to assess how information in the DNN is affected by systematically varying the inputs. A recent study using this approach found that the computation of affordance related activation patterns depended on the presence in the image of high spatial frequencies and cardinal orientations in the lower visual field (Bonner & Epstein 2018). Moreover, units in the intermediate levels of the DNN that represented affordances were selectively responsive to boundary-defining junctions and extended surfaces (Figure 4b). These observations provide insight into how the DNN transforms low-level visual information into higher-level features. While most work thus far has focused on DNNs pre-trained for recognition, future work should investigate how the features represented in DNNs vary as a function of both task and training images.

## Scene perception in the real world

Another important challenge for future research is understanding the contributions of scene regions to scene perception in cases where both the stimuli and tasks are ecologically realistic. The vast majority of neuroimaging experiments on scene perception might be characterized as “holiday snapshot perception”—a well-composed photograph is flashed on the screen in foveal vision, and the subject attempts to understand the photograph as quickly as possible. Certainly there are some situations – for example, when coming through a doorway—in which a scene suddenly becomes visible and must be rapidly interpreted. However, it is more frequently the case that we are immersed in an environment for an extended period of time, and our comprehension of the scene must develop across multiple fixations and changes in body and head position. In such real-world situations, the identity and category of the surroundings does not change at a rapid rate, but the spatial relationship between the observer and the scene is in constant flux. This contrasts somewhat with the case of object perception, where a new object comes into view at each fixation and must be rapidly identified based on a new bolus of high-resolution information from the center of the visual field.

So, how are scenes perceived in the real world? Some insight comes from naturalistic viewing experiments. When participants are asked to simply watch a movie in the scanner, PPA responds most strongly to movie frames depicting street scenes, landscapes, rooms, or corridors (Hasson et al. 2004). This finding shows that the preferential response to scenes in the PPA is not an artifact of the one-image-at-a-time presentation paradigm that is used in most laboratory experiments. Moreover, it suggests the possibility that scene areas tend to be most active when we attend broadly to the environment rather than focusing in on a specific object (Treisman 2006). It would be of great interest to relate activity in scene regions to the dynamics of eye movements, attention, and full-body kinematics during active navigation (e.g. Matthis et al. 2018). fMRI may not be the optimal method for investigating this issue, given the inherent motion restrictions of fMRI scanners, but other cognitive neuroscience methods, such as mobile electroencephalography, functional near-infrared spectroscopy, and electrocorticography, might provide useful data.

A related problem is understanding how scene regions contribute to ecologically relevant real-world behaviors. Consider the case of navigation—getting from one location to another in the real world. It is broadly accepted that there are at least two ways that this can be achieved (Chersi & Burgess 2015). First, a navigator can use a response-based strategy, in which she implements a series of actions, each of which is triggered by the presence of a specific landmark (“turn left at the church, walk down the street about 500 feet”). Second, she can use a cognitive-map based strategy, which involves keeping track of one’s location and heading in a consistent spatial coordinate frame. Response-based strategies involve the striatum, particularly the caudate nucleus, whereas map-based strategies involve the hippocampus and entorhinal cortex. A recent report indicates a functional connection between the PPA and the caudate (Nasr & Rosas 2016), which suggests the possibility that the PPA might provide landmark information to this structure during response-based route following. In cognitive-map based navigation, PPA and RSC/MPA are typically activated (Hartley et al. 2003), and we previously outlined evidence that they might play



complementary functional roles, with the PPA primarily concerned with processing place-related cues that indicate which hippocampal map needs to be retrieved, and RSC/MPA primarily concerned with processing cues that allow the navigator to localize and orient themselves relative to the retrieved map (Epstein et al. 2017, Julian et al. 2018b, Julian et al. 2015). Beyond these preliminary observations, we believe that the role of scene regions in realistic navigation remains to be explored.

Two other important real-world behaviors that often recruit scene processing are visual search and object recognition. Whether it is looking for our keys on a crowded counter, or looking for a mailbox on a crowded street, scene properties can be essential for guiding visual search (Torralba et al. 2006). The neural operations that implement this guidance are not well understood, but one neuroimaging study found that when subjects searched for cued objects in unfamiliar scenes, multivoxel patterns in RSC/MPA, LOC, and the intraparietal sulcus coded the side of the scene that the object was most likely to appear (Preston et al. 2013). In familiar scenes, on the other hand, work using the contextual cueing paradigm suggests that the hippocampus is crucial for guiding search for objects that appear consistently in the same position (Chun & Phelps 1999, Greene et al. 2007). Scene information may also have an influence on object recognition (Biederman et al. 1982, Davenport & Potter 2004) especially when the percept of the object is unclear (Oliva & Torralba 2007). In such cases, contextual signals from the PPA and RSC/MPA may work in concert with top-down signals from medial prefrontal cortex to constrain object recognition (Bar 2004, Brandman & Peelen 2017). These contextual signals may provide information about which objects are typically found in a given scene and their most likely locations.

### **Beyond the “classical” scene network**

A third promising line of new research explores scene representations outside of the “classical” scene network, most notably in the hippocampus. Contrary to the common view that this structure exclusively supports memory, alternative theories propose that it plays a central role in scene construction—that is, the bringing together of elements from memory or the imagination into a coherent spatial framework (Maguire & Mullally 2013)—or scene perception (Graham et al. 2010). Consistent with these ideas, damage to the hippocampus leads to impairment on tasks involving scenes, even when there is minimal mnemonic demand (Lee et al. 2005), and reduces boundary extension, which is often taken as a behavioral marker of scene processing (Mullally et al. 2012; although see Kim et al. 2015). The anterior medial hippocampus, especially the subiculum and pre/para-subiculum, is more activated when imagining scenes than when imagining objects (Dalton et al. 2018) and more active in a perceptual oddity task when scenes rather than objects are the items to be compared (Hodgetts et al. 2017). Moreover, a recent study found that the hippocampus was more active when viewing intact scenes than when viewing scrambled scenes during the standard 1-back perceptual matching task that is often used by vision scientists to localize scene regions (Hodgetts et al. 2016) (see also Zeidman et al. 2015).

A key unresolved question is whether scene representations in the hippocampus bear any resemblance to scene representation in the PPA, RSC/MPA, and OPA. One salient difference between these regions is that there is no evidence for a retinotopic map of visual space in the

hippocampus. However, there are longstanding reports of “spatial view cells” in this region (Rolls & Wirth 2018), and recent work suggests that entorhinal cortex—the primary input structure to the hippocampus—uses grid cells to represent visual space in a reference frame that is anchored to the geometry of the visual display (Julian et al. 2018a; see also Killian et al. 2012, Nau et al. 2018). Thus, one possibility is that visual scene representations in the PPA, RSC/MPA, and OPA are transformed into more purely spatial scene representations in entorhinal cortex and the hippocampus that would be akin to a cognitive map of the scene (Epstein et al. 2017, Julian et al. 2018b, Nau et al. 2018). In any case, the role of the MTL in scene processing deserves further investigation, and in general an important issue for the future is understanding how scene regions operate within the context of larger brain networks.

Such studies may provide insight into an interesting question: why did the human visual system evolve to have three scene regions? As discussed previously, PPA and OPA are differentially sensitive to stimulation in the LVF vs. UVF, so it is reasonable to hypothesize that differences in visual input might have driven the development of separate scene regions within the ventral vs. dorsal visual pathway. Equally relevant in determining the anatomical and functional organization of the scene system, however, might be the need to communicate with downstream target systems such as the hippocampus. The medial temporal lobe memory system is evolutionarily conserved across humans, monkeys, and rodents (Burwell et al. 1995). In rodents, major inputs to this system are provided by postrhinal and retrosplenial cortices (Furtak et al. 2007, Yoder et al. 2011) which are believed to be the homologues of parahippocampal and retrosplenial cortices in primates. Thus, PPA and RSC/MPA might have developed in their specific anatomical locations in order to take advantage of these two pre-existing points of connection to the entorhinal cortex and hippocampus.

## Conclusions

Over the past 20 years, we have learned much about scene processing in the human brain. Of central importance are the scene-selective brain regions PPA, RSC/MPA, and OPA, although other brain regions, such as object-selective LOC and possibly the hippocampus, also play a role. Our review reveals the complex nature of processing within these regions: sensitivity to low-level features is observed, but also sensitivity to higher-order properties of the stimulus; scene regions mediate visual recognition, but also spatial navigation; there is some differentiation of function between the three scene regions, but the observed specializations are by no means absolute. Despite these important insights, there is still much to learn, and future work will need to develop computational models, explore scene perception under ecologically realistic conditions, and understand how scene regions interact with larger brain networks. Given the importance of these questions and the richness of the data, both neural and behavioral, we anticipate that scene perception will continue to be a central topic of investigation in vision science in the years to come.

## Acknowledgements

We thank Edward Silson and Susan Wardle for comments. RAE is supported by NIH grants EY-022350 and EY-027047. CIB is supported by the Intramural Research Program of the National Institute of Mental Health (ZIA-MH-002909).

## Literature Cited

- Aguirre GK, D'Esposito M. 1999 Topographical disorientation: a synthesis and taxonomy. *Brain* 122: 1613–28 [PubMed: 10468502]
- Aguirre GK, Zarahn E, D'Esposito M. 1998 An area within human ventral cortex sensitive to “building” stimuli: Evidence and implications. *Neuron* 21: 373–83 [PubMed: 9728918]
- Alexander AS, Nitz DA. 2015 Retrosplenial cortex maps the conjunction of internal and external spaces. *Nat Neurosci* 18: 1143–51. [PubMed: 26147532]
- Aminoff EM, Kveraga K, Bar M. 2013 The role of the parahippocampal cortex in cognition. *Trends Cogn Sci* 17: 379–90 [PubMed: 23850264]
- Amit E, Mehoudar E, Trope Y, Yovel G. 2012 Do object-category selective regions in the ventral visual stream represent perceived distance information? *Brain Cogn* 80: 201–13 [PubMed: 22902306]
- Arcaro MJ, Livingstone MS. 2017 Retinotopic organization of scene areas in macaque inferior temporal cortex. *J Neurosci* 37: 7373–89 [PubMed: 28674177]
- Arcaro MJ, McMains SA, Singer BD, Kastner S. 2009 Retinotopic organization of human ventral visual cortex. *J Neurosci* 29: 10638–52 [PubMed: 19710316]
- Auger SD, Zeidman P, Maguire EA. 2015 A central role for the retrosplenial cortex in de novo environmental learning. *Elife* 4: e09031
- Bainbridge WA, Oliva A. 2015 Interaction envelope: Local spatial representations of objects at all scales in scene-selective regions. *Neuroimage* 122: 408–16 [PubMed: 26236029]
- Baldassano C, Esteva A, Fei-Fei L, Beck DM. 2016a Two distinct scene-processing networks connecting vision and memory. *eNeuro3: ENEURO*. 0178–16.2016
- Baldassano C, Fei-Fei L, Beck DM. 2016b Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *J Vis* 16: 9
- Bar M. 2004 Visual objects in context. *Nat Rev Neurosci* 5: 617–29 [PubMed: 15263892]
- Bar M, Aminoff E. 2003 Cortical analysis of visual context. *Neuron* 38: 347–58 [PubMed: 12718867]
- Bastin J, Vidal JR, Bouvier S, Perrone-Bertolotti M, Benis D, et al. 2013 Temporal Components in the Parahippocampal Place Area Revealed by Human Intracerebral Recordings. *J Neurosci* 33: 10123–31 [PubMed: 23761907]
- Baumann O, Mattingley JB. 2010 Medial Parietal Cortex Encodes Perceived Heading Direction in Humans. *J Neurosci* 30: 12897–901 [PubMed: 20881108]
- Berman D, Golomb JD, Walther DB. 2017 Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. *Plos One* 12: e0189828 [PubMed: 29272283]
- Bettencourt KC, Xu YD. 2013 The Role of Transverse Occipital Sulcus in Scene Perception and Its Relationship to Object Individuation in Inferior Intraparietal Sulcus. *J Cogn Neurosci* 25: 1711–22 [PubMed: 23662863]
- Biederman I 1972 Perceiving real-world scenes. *Science* 177: 77–80 [PubMed: 5041781]
- Biederman I, Mezzanotte RJ, Rabinowitz JC. 1982 Scene perception: detecting and judging objects undergoing relational violations. *Cogn Psychol* 14: 143–77 [PubMed: 7083801]
- Burwell RD, Witter MP, Amaral DG. 1995 Perirhinal and postrhinal cortices of the rat: A review of the neuroanatomical literature and comparison with findings from the monkey brain. *Hippocampus* 5: 390–408. [PubMed: 8773253]
- Bonner MF, Epstein RA. 2017 Coding of navigational affordances in the human visual system. *Proc Natl Acad Sci U S A* 114: 4793–98 [PubMed: 28416669]
- Bonner MF, Epstein RA. 2018 Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLoS Comput Biol* 14: e1006111 [PubMed: 29684011]

- Boucart M, Moroni C, Thibaut M, Szaffarczyk S, Greene M. 2013 Scene categorization at large visual eccentricities. *Vision Res* 86: 35–42 [PubMed: 23597581]
- Brandman T, Peelen XV. 2017 Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. *J Neurosci* 37: 7700–10 [PubMed: 28687603]
- Bryan PB, Julian JB, Epstein RA. 2016 Rectilinear Edge Selectivity Is Insufficient to Explain the Category Selectivity of the Parahippocampal Place Area. *Front Hum Neurosci* 10: 137 [PubMed: 27064591]
- Byrne P, Becker S, Burgess N. 2007 Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychol Rev* 114: 340–75 [PubMed: 17500630]
- Cant JS, Xu Y. 2012 Object ensemble processing in human anterior-medial ventral visual cortex. *J Neurosci* 32: 7685–700 [PubMed: 22649247]
- Cate AD, Goodale MA, Kohler S. 2011 The role of apparent size in building- and object-specific regions of ventral visual cortex. *Brain Res* 1388: 109–22 [PubMed: 21329676]
- Cheng K, Newcombe NS. 2005 Is there a geometric module for spatial orientation? Squaring theory and evidence. *Psychon Bull Rev* 12: 1–23 [PubMed: 15945200]
- Chersi F, Burgess N. 2015 The Cognitive Architecture of Spatial Navigation: Hippocampal and Striatal Contributions. *Neuron* 88: 64–77 [PubMed: 26447573]
- Choo H, Walther DB. 2016 Contour junctions underlie neural representations of scene categories in high-level human visual cortex. *Neuroimage* 135: 32–44 [PubMed: 27118087]
- Chun MM, Phelps EA. 1999 Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. *Nat Neurosci* 2: 844–7 [PubMed: 10461225]
- Cichy RM, Khosla A, Pantazis D, Oliva A. 2017 Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *Neuroimage* 153: 346–58 [PubMed: 27039703]
- Dalton MA, Zeidman P, McCormick C, Maguire EA. 2018 Differentiable processing of objects, associations, and scenes within the hippocampus. *J Neurosci* 38: 8146–59 [PubMed: 30082418]
- Davenport JL, Potter MC. 2004 Scene consistency in object and background perception. *Psychol Sci* 15: 559–64 [PubMed: 15271002]
- Dilks DD, Julian JB, Kubilius J, Spelke ES, Kanwisher N. 2011 Mirror-image sensitivity and invariance in object and scene processing pathways. *J Neurosci* 31: 11305–12 [PubMed: 21813690]
- Dilks DD, Julian JB, Paunov AM, Kanwisher N. 2013 The occipital place area (OPA) is causally and selectively involved in scene perception. *J Neurosci* 33: 1331–36 [PubMed: 23345209]
- Dillon MR, Persichetti AS, Spelke ES, Dilks DD. 2018 Places in the brain: Bridging layout and object geometry in scene-selective cortex. *Cereb Cortex* 28: 2365–74 [PubMed: 28633321]
- Drucker DM, Aguirre GK. 2009 Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cereb Cortex* 19: 2269–80 [PubMed: 19176637]
- Elshout JA, van den Berg AV, Haak KV. 2018 Human V2A: A map of the peripheral visual hemifield with functional connections to scene-selective cortex. *J Vis* 18: 22
- Epstein R, Graham KS, Downing PE. 2003 Viewpoint-specific scene representations in human parahippocampal cortex. *Neuron* 37: 865–76 [PubMed: 12628176]
- Epstein R, Harris A, Stanley D, Kanwisher N. 1999 The parahippocampal place area: Recognition, navigation, or encoding? *Neuron* 23: 115–25 [PubMed: 10402198]
- Epstein R, Kanwisher N. 1998 A cortical representation of the local visual environment. *Nature* 392: 598–601 [PubMed: 9560155]
- Epstein RA. 2005 The cortical basis of visual scene processing. *Vis Cogn* 12: 954–78
- Epstein RA. 2008 Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn Sci* 12: 388–96 [PubMed: 18760955]
- Epstein RA, Higgins JS, Jablonski K, Feiler AM. 2007a Visual scene processing in familiar and unfamiliar environments. *J Neurophysiol* 97: 3670–83 [PubMed: 17376855]
- Epstein RA, Morgan LK. 2012 Neural responses to visual scenes reveals inconsistencies between fMRI adaptation and multivoxel pattern analysis. *Neuropsychologia* 50: 530–43 [PubMed: 22001314]

- Epstein RA, Parker WE, Feiler AM. 2007b Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *J Neurosci* 27: 6141–9 [PubMed: 17553986]
- Epstein RA, Parker WE, Feiler AM. 2008 Two kinds of fMRI repetition suppression? Evidence for dissociable neural mechanisms. *J Neurophysiol* 99: 2877–86 [PubMed: 18400954]
- Epstein RA, Patai EZ, Julian JB, Spiers HJ. 2017 The cognitive map in humans: spatial navigation and beyond. *Nat Neurosci* 20: 1504–13 [PubMed: 29073650]
- Epstein RA, Vass LK. 2014 Neural systems for landmark-based wayfinding in humans. *Philos Trans R Soc London B Biol Sci* 369: 20120533 [PubMed: 24366141]
- Ferrara K, Park S. 2016 Neural representation of scene boundaries. *Neuropsychologia* 89: 180–90 [PubMed: 27181883]
- Furtak SC, Wei S, Agster KL, Burwell RD. 2007 Functional neuroanatomy of the parahippocampal region in the rat: The perirhinal and postrhinal cortices. *Hippocampus* 17: 709–22. [PubMed: 17604355]
- Gallistel CR. 1990 *The Organization of Learning*. Cambridge, MA: MIT Press
- Ganaden RE, Mullin CR, Steeves JK. 2013 Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing. *J Cogn Neurosci* 25: 961–68 [PubMed: 23410031]
- Graham KS, Barense MD, Lee AC. 2010 Going beyond LTM in the MTL: a synthesis of neuropsychological and neuroimaging findings on the role of the medial temporal lobe in memory and perception. *Neuropsychologia* 48: 831–53 [PubMed: 20074580]
- Greene AJ, Gross WL, Elsinger CL, Rao SM. 2007 Hippocampal differentiation without recognition: an fMRI analysis of the contextual cueing task. *Learn Mem* 14: 548–53 [PubMed: 17690338]
- Greene MR, Oliva A. 2009 Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn Psychol* 58: 137–76 [PubMed: 18762289]
- Grill-Spector K. 2003 The neural basis of object perception. *Curr Opin Neurobiol* 13: 159–66 [PubMed: 12744968]
- Grill-Spector K, Weiner KS, Kay K, Gomez J. 2017 The functional neuroanatomy of human face perception. *Annu Rev Vis Sci* 3: 167–96 [PubMed: 28715955]
- Groen IIA, Ghebreab S, Lamme VAF, Scholte HS. 2016 The time course of natural scene perception with reduced attention. *J Neurophysiol* 115: 931–46 [PubMed: 26609116]
- Groen IIA, Ghebreab S, Prins H, Lamme VAF, Scholte HS. 2013 From image statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *J Neurosci* 33: 18814–24 [PubMed: 24285888]
- Groen IIA, Greene MR, Baldassano C, Li FF, Beck DM, Baker CI. 2018 Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *Elife* 7: e32962 [PubMed: 29513219]
- Groen IIA, Silson EH, Baker CI. 2017 Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philos Trans R Soc London B Biol Sci* 372: 20160102 [PubMed: 28044013]
- Harel A, Groen IIA, Kravitz DJ, Deouell LY, Baker CI. 2016 The temporal dynamics of scene processing: A multifaceted EEG investigation. *Eneuro* 3: e0139–16.2016
- Harel A, Kravitz DJ, Baker CI. 2012 Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cereb Cortex* 23: 947–957 [PubMed: 22473894]
- Hartley T, Maguire EA, Spiers HJ, Burgess N. 2003 The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. *Neuron* 37: 877–88 [PubMed: 12628177]
- Hasson U, Harel M, Levy I, Malach R. 2003 Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron* 37: 1027–41 [PubMed: 12670430]
- Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R. 2004 Intersubject synchronization of cortical activity during natural vision. *Science* 303: 1634–40 [PubMed: 15016991]
- Hatfield M, McCloskey M, Park S. 2016 Neural representation of object orientation: A dissociation between MVPA and Repetition Suppression. *Neuroimage* 139: 136–48 [PubMed: 27236084]

- He C, Peelen MV, Han Z, Lin N, Caramazza A, Bi Y. 2013 Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *Neuroimage* 79: 1–9 [PubMed: 23624496]
- Henderson JM, Hollingworth A. 1999 High-level scene perception. *Annu Rev Psychol* 50: 243–71 [PubMed: 10074679]
- Henderson JM, Larson CL, Zhu DC. 2008 Full scenes produce more activation than close-up scenes and scene-diagnostic objects in parahippocampal and retrosplenial cortex: an fMRI study. *Brain Cogn* 66: 40–9 [PubMed: 17606317]
- Henderson JM, Zhu DC, Larson CL. 2011 Functions of parahippocampal place area and retrosplenial cortex in real-world scene analysis: An fMRI study. *Vis Cogn* 19: 910–27
- Hodgetts CJ, Shine JP, Lawrence AD, Downing PE, Graham KS. 2016 Evidencing a place for the hippocampus within the core scene processing network. *Hum Brain Mapp* 37: 3779–94 [PubMed: 27257784]
- Hodgetts CJ, Voets NL, Thomas AG, Clare S, Lawrence AD, Graham KS. 2017 Ultra-High-field fMRI reveals a role for the subiculum in scene perceptual discrimination. *J Neurosci* 37: 3150–59 [PubMed: 28213445]
- Ishai A, Ungerleider LG, Martin A, Schouten HL, Haxby JV. 1999 Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci U S A* 96: 9379–84 [PubMed: 10430951]
- Janzen G, van Turenout M. 2004 Selective neural representation of objects relevant for navigation. *Nat Neurosci* 7: 673–7 [PubMed: 15146191]
- Julian JB, Keinath AT, Frazzetta G, Epstein RA. 2018a Human entorhinal cortex represents visual space using a boundary-anchored grid. *Nat Neurosci* 21: 191–4 [PubMed: 29311745]
- Julian JB, Keinath AT, Marchette SA, Epstein RA. 2018b The neurocognitive basis of spatial reorientation. *Curr Biol* 28: R1059–R73 [PubMed: 30205055]
- Julian JB, Keinath AT, Muzzio IA, Epstein RA. 2015 Place recognition and heading retrieval are mediated by dissociable cognitive systems in mice. *Proc Natl Acad Sci U S A* 112: 6503–08 [PubMed: 25941390]
- Julian JB, Ryan J, Hamilton RH, Epstein RA. 2016 The Occipital Place Area is causally involved in representing environmental boundaries during navigation. *Curr Biol* 26: 1104–9 [PubMed: 27020742]
- Jung Y, Larsen B, Walther DB. 2018 Modality-independent coding of scene categories in prefrontal cortex. *J Neurosci* 38: 5969–81 [PubMed: 29858483]
- Kamps FS, Julian JB, Kubilius J, Kanwisher N, Dilks DD. 2016 The occipital place area represents the local elements of scenes. *NeuroImage* 132: 417–24 [PubMed: 26931815]
- Killian NJ, Jutras MJ, Buffalo EA. 2012 A map of visual space in the primate entorhinal cortex. *Nature* 491: 761–64 [PubMed: 23103863]
- Kim S, Dede AJ, Hopkins RO, Squire LR. 2015 Memory, scene construction, and the human hippocampus. *Proc Natl Acad Sci U S A* 112: 4767–72 [PubMed: 25825712]
- Konkle T, Caramazza A. 2013 Tripartite organization of the ventral stream by animacy and object size. *J Neurosci* 33: 10235–42 [PubMed: 23785139]
- Konkle T, Oliva A. 2012 A real-world size organization of object responses in occipitotemporal cortex. *Neuron* 74: 1114–24 [PubMed: 22726840]
- Kornblith S, Cheng XQ, Ohayon S, Tsao DY. 2013 A network for scene processing in the macaque temporal lobe. *Neuron* 79: 766–81 [PubMed: 23891401]
- Kravitz DJ, Peng CS, Baker CI. 2011 Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J Neurosci* 31: 7322–33 [PubMed: 21593316]
- Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013 The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci* 17: 26–49 [PubMed: 23265839]
- Kumar M, Federmeier KD, Fei-Fei L, Beck DM. 2017 Evidence for similar patterns of neural activity elicited by picture- and word-based representations of natural scenes. *Neuroimage* 155: 422–36 [PubMed: 28343000]



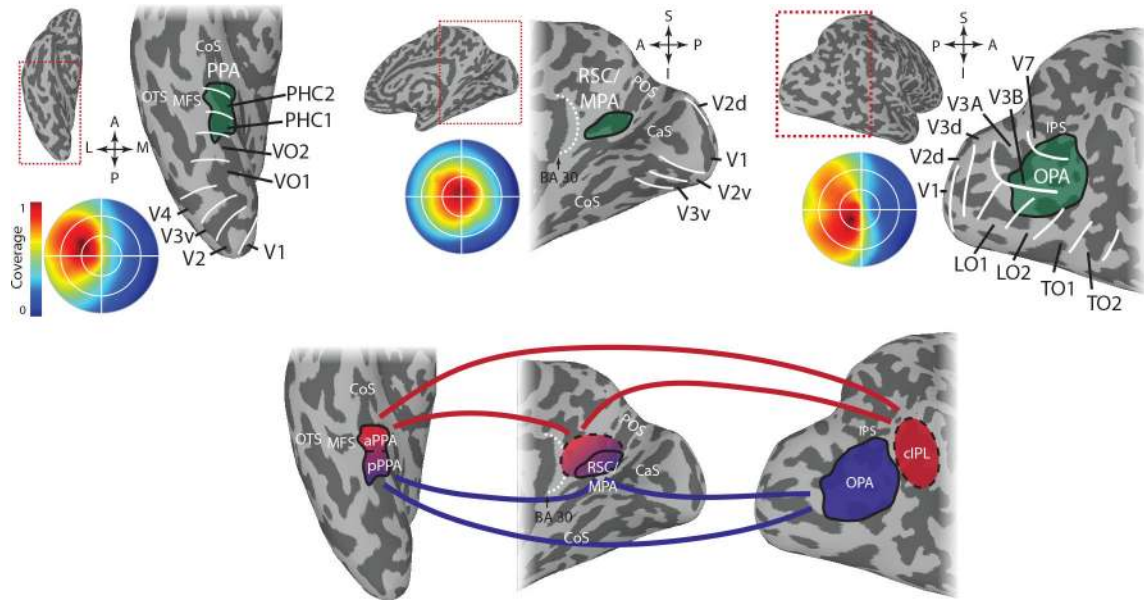
- Larson AM, Loschky LC. 2009 The contributions of central versus peripheral vision to scene gist recognition. *J Vis* 9: 6, 1–16
- Lee AC, Bussey TJ, Murray EA, Saksida LM, Epstein RA, et al. 2005 Perceptual deficits in amnesia: challenging the medial temporal lobe ‘mnemonic’ view. *Neuropsychologia* 43: 1–11 [PubMed: 15488899]
- Lee SA. 2017 The boundary-based view of spatial cognition: a synthesis. *Curr Opin Behav Sci* 16: 58–65
- Lescroart MD, Gallant JL. 2019 Human scene-selective areas represent 3D configurations of surfaces. *Neuron* 1: 178–192.e7
- Lescroart MD, Stansbury DE, Gallant JL. 2015 Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Front Comput Neurosci* 9: 135 [PubMed: 26594164]
- Levy I, Hasson U, Avidan G, Hendler T, Malach R. 2001 Center-periphery organization of human object areas. *Nat Neurosci* 4: 533–39 [PubMed: 11319563]
- Long B, Yu CP, Konkle T. 2018 Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proc Natl Acad Sci U S A* 115: E9015–E24 [PubMed: 30171168]
- MacEvoy SP, Epstein RA. 2007 Position selectivity in scene- and object-responsive occipitotemporal regions. *J Neurophysiol* 98: 2089–98 [PubMed: 17652421]
- MacEvoy SP, Epstein RA. 2011 Constructing scenes from objects in human occipitotemporal cortex. *Nat Neurosci* 14: 1323–9 [PubMed: 21892156]
- Maguire EA. 2001 The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scand J Psychol* 42: 225–38 [PubMed: 11501737]
- Maguire EA, Mullally SL. 2013 The hippocampus: a manifesto for change. *J Exp Psychol Gen* 142: 1180–9 [PubMed: 23855494]
- Malcolm GL, Groen II, Baker CI. 2016 Making sense of real-world scenes. *Trends Cogn Sci* 20: 843–56 [PubMed: 27769727]
- Malcolm GL, Silson EH, Henry JR, Baker CI. 2018 Transcranial magnetic stimulation to the Occipital Place Area biases gaze during scene viewing. *Front Hum Neurosci* 12: 189 [PubMed: 29867413]
- Marchette SA, Vass LK, Ryan J, Epstein RA. 2014 Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nat Neurosci* 17: 1598–1606 [PubMed: 25282616]
- Marchette SA, Vass LK, Ryan J, Epstein RA. 2015 Outside looking in: Landmark generalization in the human navigational system. *J Neurosci* 35: 14896–908 [PubMed: 26538658]
- Matthis JS, Yates JL, Hayhoe MM. 2018 Gaze and the control of foot placement when walking in natural terrain. *Curr Biol* 28: 1224–33.e5 [PubMed: 29657116]
- Megevand P, Groppe DM, Goldfinger MS, Hwang ST, Kingsley PB, et al. 2014 Seeing scenes: Topographic visual hallucinations evoked by direct electrical stimulation of the Parahippocampal Place Area. *J Neurosci* 34: 5399–405 [PubMed: 24741031]
- Morgan LK, Macevoy SP, Aguirre GK, Epstein RA. 2011 Distances between real-world locations are represented in the human hippocampus. *J Neurosci* 31: 1238–45 [PubMed: 21273408]
- Mormann F, Kornblith S, Cerf M, Ison MJ, Kraskov A, et al. 2017 Scene-selective coding by single neurons in the human parahippocampal cortex. *Proc Natl Acad Sci U S A* 114: 1153–58 [PubMed: 28096381]
- Mullally SL, Intraub H, Maguire EA. 2012 Attenuated boundary extension produces a paradoxical memory advantage in amnesic patients. *Curr Biol* 22: 261–8 [PubMed: 22264610]
- Mullally SL, Maguire EA. 2011 A new role for the parahippocampal cortex in representing space. *J Neurosci* 31: 7441–9 [PubMed: 21593327]
- Nakamura K, Kawashima R, Sato N, Nakamura A, Sugiura M, et al. 2000 Functional delineation of the human occipito-temporal areas related to face and scene processing - A PET study. *Brain* 123: 1903–12 [PubMed: 10960054]
- Nasr S, Devaney KJ, Tootell RBH. 2013 Spatial encoding and underlying circuitry in scene-selective cortex. *Neuroimage* 83: 892–900 [PubMed: 23872156]

- Nasr S, Echavarria CE, Tootell RBH. 2014 Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *J Neurosci* 34: 6721–35 [PubMed: 24828628]
- Nasr S, Liu N, Devaney KJ, Yue X, Rajimehr R, et al. 2011 Scene-selective cortical regions in human and nonhuman primates. *J Neurosci* 31: 13771–85 [PubMed: 21957240]
- Nasr S, Rosas HD. 2016 Impact of visual corticostriatal loop disruption on neural processing within the Parahippocampal Place Area. *J Neurosci* 36: 10456–71 [PubMed: 27707978]
- Nasr S, Tootell RB. 2012 A cardinal orientation bias in scene-selective visual cortex. *J Neurosci* 32: 14921–6 [PubMed: 23100415]
- Nau M, Navarro Schroder T, Bellmund JLS, Doeller CF. 2018 Hexadirectional coding of visual space in human entorhinal cortex. *Nat Neurosci* 21: 188–90 [PubMed: 29311746]
- O’Craven KM, Kanwisher N. 2000 Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *J Cogn Neurosci* 12: 1013–23 [PubMed: 11177421]
- Oliva A, Torralba A. 2001 Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vis* 42: 145–75
- Oliva A, Torralba A. 2007 The role of context in object recognition. *Trends Cogn Sci* 11: 520–7 [PubMed: 18024143]
- Park J, Park S. 2017 Conjoint representation of texture ensemble and location in the parahippocampal place area. *J Neurophysiol* 117: 1595–607 [PubMed: 28123006]
- Park S, Brady TF, Greene MR, Oliva A. 2011 Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *J Neurosci* 31: 1333–40 [PubMed: 21273418]
- Park S, Chun MM. 2009 Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. *Neuroimage* 47: 1747–56 [PubMed: 19398014]
- Park S, Konkle T, Oliva A. 2015 Parametric Coding of the Size and Clutter of Natural Scenes in the Human Brain. *Cereb Cortex* 25: 1792–805 [PubMed: 24436318]
- Persichetti AS, Dilks DD. 2016 Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex* 77: 155–63 [PubMed: 26963085]
- Persichetti AS, Dilks DD. 2018 Dissociable neural systems for recognizing places and navigating through them. *J Neurosci* 38: 10295–304. [PubMed: 30348675]
- Potter MC. 1975 Meaning in visual search. *Science* 187: 965–6 [PubMed: 1145183]
- Preston TJ, Guo F, Das K, Giesbrecht B, Eckstein MP. 2013 Neural representations of contextual guidance in visual search of real-world scenes. *J Neurosci* 33: 7846–55 [PubMed: 23637176]
- Rajimehr R, Devaney KJ, Bilenko NY, Young JC, Tootell RB. 2011 The “parahippocampal place area” responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol* 9: e1000608 [PubMed: 21483719]
- Ranganath C, Ritchey M. 2012 Two cortical systems for memory-guided behavior. *Nat Rev Neurosci* 13: 713–26. [PubMed: 22992647]
- Robertson CE, Hermann KL, Mynick A, Kravitz DJ, Kanwisher N. 2016 Neural representations integrate the current field of view with the remembered 360 panorama in scene-selective cortex. *Curr Biol* 26: 2463–68 [PubMed: 27618266]
- Rolls ET, Wirth S. 2018 Spatial representations in the primate hippocampus, and their functions in memory and navigation. *Prog Neurobiol* 171: 90–113 [PubMed: 30219248]
- Schinazi VR, Epstein RA. 2010 Neural correlates of real-world route learning. *Neuroimage* 53: 725–35 [PubMed: 20603219]
- Schindler A, Bartels A. 2016 Visual high-level regions respond to high-level stimulus content in the absence of low-level confounds. *Neuroimage* 132: 520–25 [PubMed: 26975552]
- Shine JP, Valdés-Herrera JP, Hegarty M, Wolbers T. 2016 The human retrosplenial cortex and thalamus code head direction in a global reference frame. *J Neurosci* 36: 6371–81 [PubMed: 27307227]
- Silson EH, Chan AW-Y, Reynolds RC, Kravitz DJ, Baker CI. 2015 A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *J Neurosci* 35: 11921–35 [PubMed: 26311774]

- Silson EH, Groen IIA, Kravitz DJ, Baker CI. 2016a Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *J Vis* 16: 14, 1–21
- Silson EH, Steel AD, Baker CI. 2016b Scene-selectivity and retinotopy in medial parietal cortex. *Front Hum Neurosci* 10: 412 [PubMed: 27588001]
- Stansbury DE, Naselaris T, Gallant JL. 2013 Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron* 79: 1025–34 [PubMed: 23932491]
- Steeves JKE, Humphrey GK, Culham JC, Menon RS, Milner AD, Goodale MA. 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: 955–65 [PubMed: 15298783]
- Torralba A, Oliva A, Castelhano MS, Henderson JM. 2006 Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol Rev* 113: 766–86 [PubMed: 17014302]
- Treisman A. 2006 How the deployment of attention determines what we see. *Vis Cogn* 14: 411–43 [PubMed: 17387378]
- Troiani V, Stigliani A, Smith ME, Epstein RA. 2014 Multiple object properties drive scene-selective regions. *Cereb Cortex* 24: 883–97 [PubMed: 23211209]
- Vann SD, Aggleton JP, Maguire EA. 2009 What does the retrosplenial cortex do? *Nat Rev Neurosci* 10: 792–802 [PubMed: 19812579]
- Vass LK, Epstein RA. 2013 Abstract representations of location and facing direction in the human brain. *J Neurosci* 33: 6133–42 [PubMed: 23554494]
- Vass LK, Epstein RA. 2017 Common neural representations for visually guided reorientation and spatial imagery. *Cereb Cortex* 27: 1457–71 [PubMed: 26759482]
- Walther DB, Caddigan E, Fei-Fei L, Beck DM. 2009 Natural scene categories revealed in distributed patterns of activity in the human brain. *J Neurosci* 29: 10573–81 [PubMed: 19710310]
- Ward EJ, MacEvoy SP, Epstein RA. 2010 Eye-centered encoding of visual space in scene-selective regions. *J Vis* 10: 6
- Watson DM, Hymers M, Hartley T, Andrews TJ. 2016 Patterns of neural response in scene-selective regions of the human brain are affected by low-level manipulations of spatial frequency. *Neuroimage* 124: 107–17 [PubMed: 26341028]
- Weiner KS, Barnett MA, Lorenz S, Caspers J, Stigliani A, et al. 2017 The cytoarchitecture of domain-specific regions in human high-level visual cortex. *Cereb Cortex* 27: 146–61 [PubMed: 27909003]
- Weiner KS, Barnett MA, Witthoft N, Golarai G, Stigliani A, et al. 2018 Defining the most probable location of the parahippocampal place area using cortex-based alignment and cross-validation. *Neuroimage* 170: 373–84 [PubMed: 28435097]
- Wolbers T, Büchel C. 2005 Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *J Neurosci* 25: 3333–40 [PubMed: 15800188]
- Wolbers T, Klatzky RL, Loomis JM, Wutte MG, Giudice NA. 2011 Modality-independent coding of spatial layout in the human brain. *Curr Biol* 21: 984–89 [PubMed: 21620708]
- Xu Y, Chun MM. 2009 Selecting and perceiving multiple visual objects. *Trends Cogn Sci* 13: 167–74 [PubMed: 19269882]
- Yoder RM, Clark BJ, Taube JS. 2011 Origins of landmark encoding in the brain. *Trends Neurosci* 34: 561–71. [PubMed: 21982585]
- Zeidman P, Mullally SL, Maguire EA. 2015 Constructing, perceiving, and maintaining scenes: Hippocampal activity and connectivity. *Cereb Cortex* 25: 3836–55 [PubMed: 25405941]
- Zhou B, Lapedriza A, Xiao J, Torralba A, Oliva A. 2014 Learning deep features for scene recognition using places database. *Advances in neural information processing systems* 27

**Summary Points:**

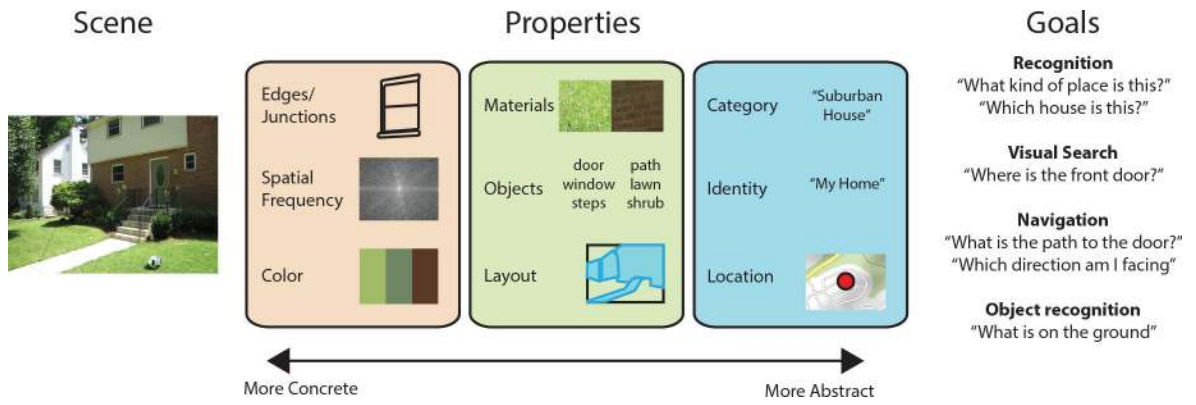
1. fMRI studies have identified three brain regions that respond selectively to scenes: the PPA, RSC/MPA, and OPA. Homologues of these three scene-responsive brain regions have been found in macaques.
2. Retinotopic responses are observed in PPA, posterior RSC/MPA, and OPA. PPA is more sensitive to stimulation in the upper visual field, whereas OPA is more sensitive to stimulation in the lower visual field.
3. Scene regions exhibit preferences for low-level visual features that are characteristics of scenes, such as high spatial frequencies, rectilinear junctions, and edges at cardinal orientations. However, this sensitivity to low-level features does not appear to be sufficient, on its own, to account for their scene-selective response.
4. Scene regions also exhibit responses that relate to higher-order structure in the stimulus. They respond strongly to the presence of environmental boundaries that define the spatial layout of scenes. Their response to single objects is modulated by spatial factors such as the real-world size, spatial stability, and navigational relevance of the objects.
5. Scene regions discriminate between scenes at multiple levels: as members of different scene categories (“beach”), as unique places or landmarks, or as individual views. Underlying these discriminations are representations of spatial layout (in PPA, RSC/MPA, and OPA), surface properties (in PPA), and within-scene objects (in PPA and LOC).
6. Scene regions also encode spatial properties that are useful for navigation. PPA and OPA are primarily concerned with analyzing local spatial structure of scenes (“vista space”), whereas RSC/MPA encodes quantities like facing direction and location that are crucial for understanding the relationship between the local scene and the broader environment.
7. Future work should develop explicit computational models of information processing in scene regions, investigate scene processing using realistic stimuli in the context of ecologically important tasks, and explore the interaction between scene regions and other parts of the brain, including the hippocampus.



**Figure 1. Scene-selective cortical regions.**

A) Group average data showing the location of the scene-selective cortical regions with respect to anatomy and retinotopically-defined areas. The circular insets show the portion of the visual field eliciting the strongest response within each scene region based on population receptive field mapping. PPA (left) is located in and around the collateral sulcus on the medial part of the ventral temporal cortex. It overlaps with retinotopically defined regions PHC-1, PHC-2 and VO-2 and responds most strongly to stimuli in the contralateral upper visual field. RSC/MPA (middle) is located in medial parietal cortex in and around the ventral portion of the parieto-occipital sulcus. It responds most strongly to stimuli in the contralateral visual field with no clear bias to the upper or lower visual field. OPA (right) is located near the transverse occipital sulcus in occipito-parietal cortex. It overlaps most prominently with V3B and LO2, but also with V3A, V7/IPS0, and LO1, and responds most strongly to stimuli in the contralateral lower visual field.

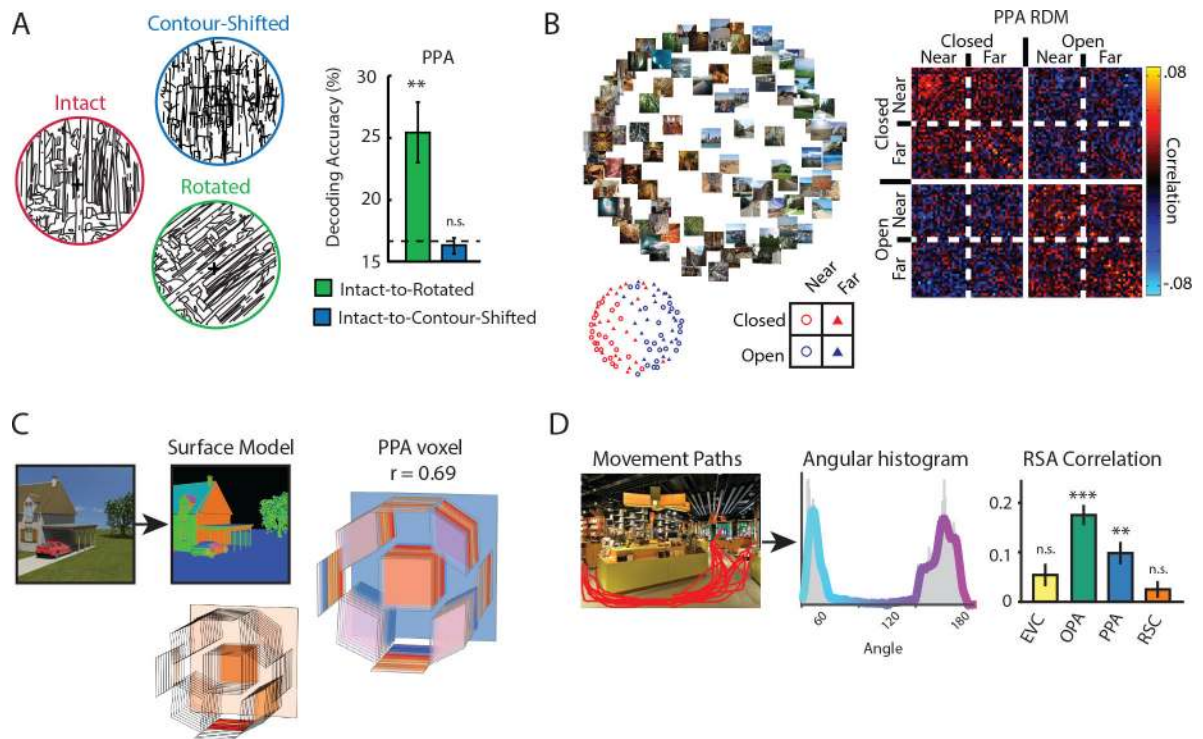
B) Relationship between the different regions based on functional connectivity. There is strong functional connectivity between all three regions, but posterior PPA (pPPA) shows stronger connectivity to OPA and posterior parts of RSC/MPA, while anterior PPA (aPPA) shows stronger connectivity with the caudal inferior parietal lobe (cIPL), a region anterior to OPA, and anterior parts of RSC/MPA as well as adjoining regions in posterior cingulate cortex. This pattern of functional connectivity might reflect separate networks for perceptual (pPPA, OPA, posterior RSC/MPA) and memory-based (aPPA, cIPL, anterior RSC/MPA) processing. COS – collateral sulcus, OTS – occipitotemporal sulcus, MFS – mid-fusiform sulcus, CaS – calcarine sulcus, POS – parieto-occipital sulcus, IPS – intraparietal sulcus



**Figure 2. Scene perception depends on both multi-level properties of the image and the observer’s goals.**

The visual system analyzes many properties of scenes, ranging from low-level features (e.g. edges, color) to mid-level elements (e.g. layout, objects) to high-level semantic and spatial properties (e.g. scene category). The results of these analyses can be used in the service of several different behavioral goals. Note that although we group properties into three levels for exposition, these levels are notional, as properties at different levels are inherently correlated, and consequently there may not be a strict low-to-high hierarchy of processing.





**Figure 3. Representations of the spatial structure of scenes**

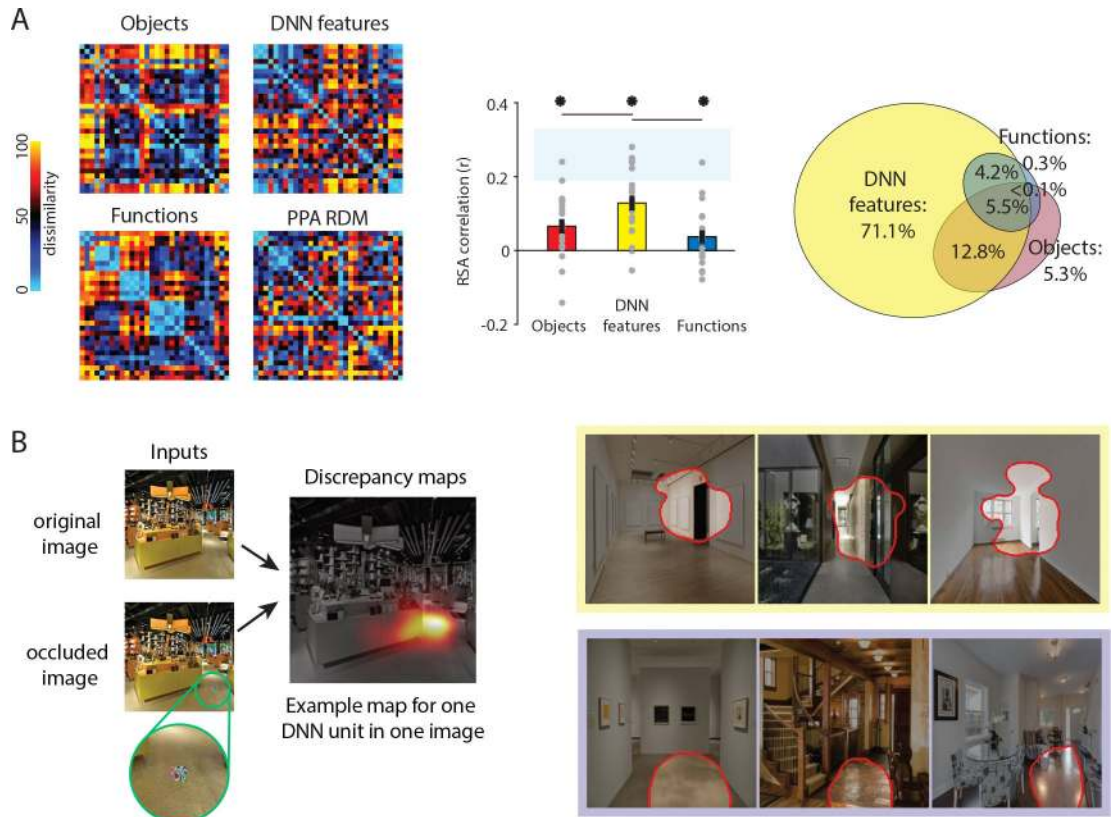
A) Results from an fMRI study showing that scene representations in PPA rely on contour junctions, an important cue for the three-dimensional arrangement of scene surfaces.

Multivoxel patterns were measured for line drawings depicting 6 scene categories. Category could be cross-decoded between original (intact) and rotated line drawings, but not between original (intact) and contour-shifted line drawings. The first manipulation preserves the contour junctions in the stimulus, while the second manipulation destroys them. Similar results were obtained in OPA.

B) Results from an fMRI study showing that scene representations in PPA are organized by spatial structure. Multivoxel activation patterns in the PPA were measured for 96 scenes. Multidimensional scaling of these data (left) reveals grouping of scenes based on layout (open vs. closed). The representational dissimilarity matrix (right) shows a clear distinction between open and closed scenes.

C) Results from an fMRI study showing that individual voxels in PPA respond to scenes based on layout-defining surfaces. Artificial scenes were modelled in terms of a histogram of surfaces at different tilt/slant and depth. Responses of voxels in scene regions could be predicted based on this model. Right shows a PPA voxel that exhibits a complex sensitivity, including strong response to fronto-parallel surfaces at intermediate distances.

D) The navigational affordances (i.e. pathways for movement) of scenes were evaluated by a set of raters, and then quantified in terms of an angular histogram. Representational similarities between multivoxel patterns in OPA (and, to a lesser extent, PPA) were related to these affordances.



**Figure 4. Computational approaches to understanding scene perception in the brain.**

A) Multivoxel fMRI patterns in PPA were obtained for 30 scene categories, and the resulting representational dissimilarity matrix (RDM) was compared to RDMs for three possible models of scene processing. Dissimilarity in the Objects model was based on the objects present within each scene; dissimilarity in the DNN features model was based on activation in a deep neural network trained on object classification; and dissimilarity in the Functions model was based on types of actions (e.g. walking, vacuuming) that could be carried out in each scene. Categories (e.g. bus depot, putting green, volcano, pier) were chosen to maximally differentiate between the three models. Middle panel shows that the RDMs for all three models correlate with the PPA RDM, with the strongest correlation for the DNN feature model. Right panel shows the results of variance partitioning, showing that much of the PPA variance explained by the Object and Functions models is shared with the DNN features model, which explains the most unique variance. Total response variance accounted for by all three models was 14.8%.

B) In an *in silico* experiment, the response profiles of individual DNN units were assessed by comparing response to an unaltered image with response to the same image overlaid with a small occluder. A discrepancy map showing the portion of the image that the unit responds to was created by varying the location of the occluder. On the right are discrepancy maps of three scenes for two DNN units that were previously shown to convey information about navigational affordances. The top unit appears to respond to features related to doorways; the bottom unit appears to respond to open spaces along the ground plane.