

Title: Two distinct scene processing networks connecting vision and memory

Authors: Christopher Baldassano¹, Andre Esteva², Diane M. Beck³, Li Fei-Fei¹

1: Stanford University, Department of Computer Science

2: Stanford University, Department of Electrical Engineering

3: University of Illinois at Urbana-Champaign, Department of Psychology and Beckman Institute

Contact: Christopher Baldassano, chrisb@princeton.edu

Abstract: A number of regions in the human brain are known to be involved in processing natural scenes, but the field has lacked a unifying framework for understanding how these different regions are organized and interact. We provide evidence from functional connectivity and meta-analyses for a new organizational principle, in which scene processing relies on two distinct networks that split the classically defined Parahippocampal Place Area (PPA). The first network consists of the Occipital Place Area (OPA/TOS) and posterior PPA, which contain retinotopic maps and are related primarily to visual features. The second network consists of the caudal Inferior Parietal Lobule (cIPL), Retrosplenial Cortex (RSC), and anterior PPA, which connect to the hippocampus and are involved in a much broader set of tasks involving episodic memory and navigation. This new framework for understanding the neural substrates of scene processing bridges results from many lines of research, and makes specific functional predictions.

22 Introduction

23 Natural scene perception has been shown to rely on a distributed set of cortical regions,
 24 including the parahippocampal place area (PPA) (Epstein & Kanwisher, 1998),
 25 retrosplenial cortex (RSC) (O'Craven & Kanwisher, 2000), and the occipital place area (OPA,
 26 aka transverse occipital sulcus, TOS) (Hasson, Harel, Levy, & Malach, 2003; Nakamura et al.,
 27 2000). More recent work has suggested that the picture is even more complicated, with
 28 multiple subdivisions within PPA and the possible involvement of the parietal lobe
 29 (Baldassano, Beck, & Fei-Fei, 2013). Although there has been substantial progress in
 30 understanding the functional properties of each of these regions and the differences
 31 between them, the field has lacked a coherent framework for summarizing the overall
 32 architecture of the human scene processing system.

33
 34 There is a long history of proposals for partitioning the visual system into separable
 35 components with different functions, such as spatial frequency channels (Campbell &
 36 Robson, 1968), what versus where/how pathways (Kravitz, Saleem, Baker, & Mishkin,
 37 2011; Mishkin, Ungerleider, & Macko, 1983), or magnocellular, parvocellular, and
 38 koniocellular streams (Kaplan, 2004). With respect to natural scene perception, one can
 39 imagine at least two separable functions: processing the specific visual features present in
 40 the current glance of a scene, and connecting that to the stable, high-level knowledge of
 41 where the place exists in the world, what has happened here in the past, and what possible
 42 actions we could take here in the future. For most cognitive and physical tasks we
 43 undertake in real-world places, the specific visual attributes we perceive are just a means

to this end, of recalling and updating information about the physical environment; “the essential feature of a landmark is not its design, but the place it holds in a city’s memory” (Muschamp, 2006). The connection between place and memory has been recognized for thousands of years, reflected in the ancient Greek “method of loci” that strengthens a memory sequence by associating it with physical locations (Yates, 1966).

Some previous work has begun to point to this type of organizing principle among scene perception regions. Mapping functional connectivity differences between pairs of scene-sensitive regions has revealed some consistent distinctions, with some regions more connected to visual cortex and others connected to parietal and medial temporal regions (Baldassano et al., 2013; Nasr, Devaney, & Tootell, 2013). Contrasting activity evoked by perceptual categorization tasks compared to semantic retrieval tasks shows a similar division between visual and higher-level cortex (Fairhall, Anzellotti, Ubaldi, & Caramazza, 2014). These experiments, however, have all been targeted, hypothesis-driven comparisons between regions with similar functional properties. It is unclear whether these divisions are major organizing principles of the brain’s connectivity networks, or simply subtle differences within a single coherent scene-processing network.

To answer this question, we took a data-driven approach to identifying scene-sensitive regions and clustering cortical connectivity. We first aggregate local high-resolution resting-state connectivity information into spatially-coherent parcels (Baldassano, Beck, & Fei-Fei, 2015), in order to increase signal to noise and obtain more interpretable units than individual voxels. We then apply hierarchical clustering to show that there exists a natural

division in posterior human cortex that splits scene-related regions into two separate, bilaterally-symmetric networks. The posterior network includes OPA and the posterior portion of PPA (retinotopic maps PHC1 and PHC2), while the anterior network is composed of the RSC, anterior PPA (aPPA), and the caudal inferior parietal lobule (cIPL). We then show that these two networks differ in their connectivity to the hippocampus, with the anterior network exhibiting much higher resting-state hippocampal coupling (especially to anterior hippocampus), suggesting that memory- and navigation-related functions are primarily restricted to the anterior network. We provide supporting evidence for this functional division from a reverse-inference meta-analysis of previous results from visual, memory, and navigation studies, and an atlas of retinotopic maps.

Based on these results, as well as a review of previous work, we propose that scene processing is fundamentally divided into two collaborating but distinct networks, with one focused on the visual features of a scene image and the other related to contextual retrieval and navigation. Under this framework, scene perception is less the function of a unified set of distributed neural machinery and more of “an ongoing dialogue between the material and symbolic aspects of the past and the continuously unfolding present” (Baker, 2012).

Results

Our primary dataset is a 1.8-billion element resting-state connectivity matrix distributed by the Human Connectome Project (Van Essen et al., 2013), which estimates the timecourse correlation between every pair of locations in the brain at 2mm resolution based on a group of 468 subjects. Since we wish to understand the large-scale structure of visual cortex, it is helpful to abstract away from individual voxels and study the functional and connectivity properties of larger parcels. Rather than imposing a parcellation based on specific regions of interest, we used a data-driven approach to produce spatially-coherent parcels tiling the cortical surface in a way that retains as much information as possible from the full connectivity matrix (Baldassano et al., 2015). This parcellation consists of 172 regions across both hemispheres, each of which contains surface vertices that all have very similar connectivity patterns with the rest of the brain. The connectivity matrix between these 172 parcels captures more than 76% of the variance in the original connectivity matrix, despite being dramatically smaller (by five orders of magnitude). A visualization of this connectivity space is shown in Supplementary Figure 1.

Clustering Parcels into Networks

To determine how these local parcels are organized into distributed networks, we performed hierarchical clustering to group together parcels with high functional connectivity (regardless of their spatial position). These networks are remarkably similar between hemispheres (despite not being constrained to be symmetric), as shown in the 10-network clustering in Figure 1.

Which of these networks are directly related to scene perception? We used data from a standard localizer in a separate group of subjects to define group-level regions of interest for scene-selective regions OPA, PPA, and RSC. We also anatomically identified cIPL as in previous work, since this region has been shown to have functional connections to scene regions (Baldassano et al., 2013).

We found that these scene ROIs fell almost entirely onto two of the connectivity networks. A posterior network (dark blue), overlapping OPA and posterior PPA, covered all of visual cortex outside of an early foveal cluster. An anterior network (magenta), overlapping cIPL, RSC, and anterior PPA, covered a parietal/medial-temporal network that includes anterior temporal and orbitofrontal parcels. This corresponds to a portion of the known default mode regions, with other default mode regions being grouped into a separate network (green); a similar fractionation of the default mode has been proposed previously (Andrews-Hanna, Reidler, Sepulcre, Poulin, & Buckner, 2010).

We note that this bifurcation into posterior and anterior scene networks occurs between neighboring parcels in both dorsal and ventral regions; i.e. there is a rapid change in connectivity profiles between OPA and cIPL and along the posterior-anterior axis of PPA. To statistically evaluate (in individual subjects) the shift in network membership in the

dorsal parcels near OPA/cIPL, we measured functional connectivity between these parcels and a reference parcel in the anterior network. According to our two-network hypothesis, we predict increasing connectivity to this anterior-network reference parcel as we move from posterior to anterior parcels within both the dorsal and ventral stream. In order to avoid influences from local noise correlations, we selected the reference parcel to be the parcel near RSC on the medial surface. Similarly, we measured connectivity between the ventral parcels near PPA and a reference parcel on the opposite lateral surface (the most anterior dorsal parcel, overlapping cIPL).

In both cases, we observed rapid increases in connectivity as we moved posterior to anterior across the network boundaries (Figure 2). Along the dorsal boundary, we see significant increases in connectivity to the RSC parcel when moving from the first to the second parcel (Left: $t_{19}=6.98$, $p<0.001$; Right: $t_{19}=6.35$, $p<0.001$; two-tailed paired t-test), from the second to the third parcel (Left: $t_{19}=7.72$, $p<0.001$; Right: $t_{19}=6.16$, $p<0.001$), and from the third to the fourth parcel (Right: $t_{19}=2.44$, $p=0.025$). We observe a similar significant (though less dramatic) increase in connectivity to the cIPL parcel when moving from the first to the second PPA parcel (Left: $t_{19}=4.21$, $p<0.001$; Right: $t_{19}=2.68$, $p=0.015$) and from the second to the third PPA parcel (Right: $t_{19}=3.03$, $p=0.007$).

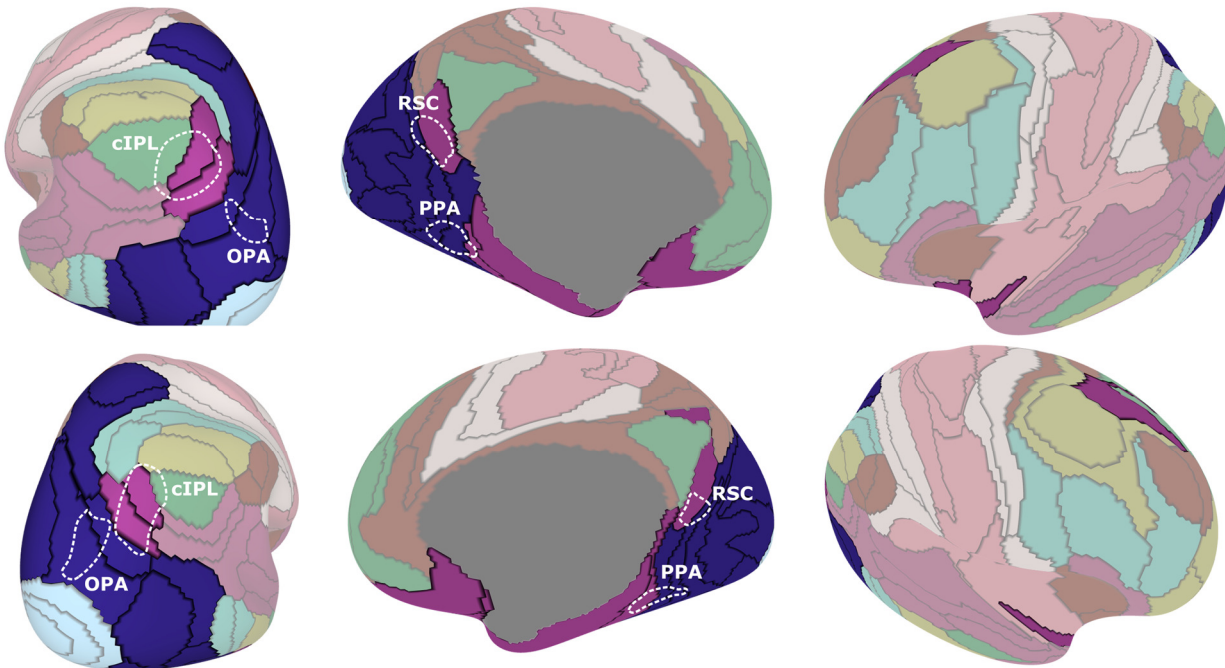


Figure 1: Connectivity clustering of cortical parcels. The cortex was first grouped into 172 local parcels (black lines) such that the surface vertices in each parcel had similar connectivity properties. Performing a second-level hierarchical clustering to group together strongly-connected parcels shows that scene-related regions are split across two networks, which are largely symmetric across left (top row) and right (bottom row) hemispheres. OPA and posterior PPA overlap with a posterior network (dark blue) that covers all of visual cortex outside the foveal confluence, while cIPL, RSC, and anterior PPA overlap with an anterior network (magenta) that covers much of the default mode network.

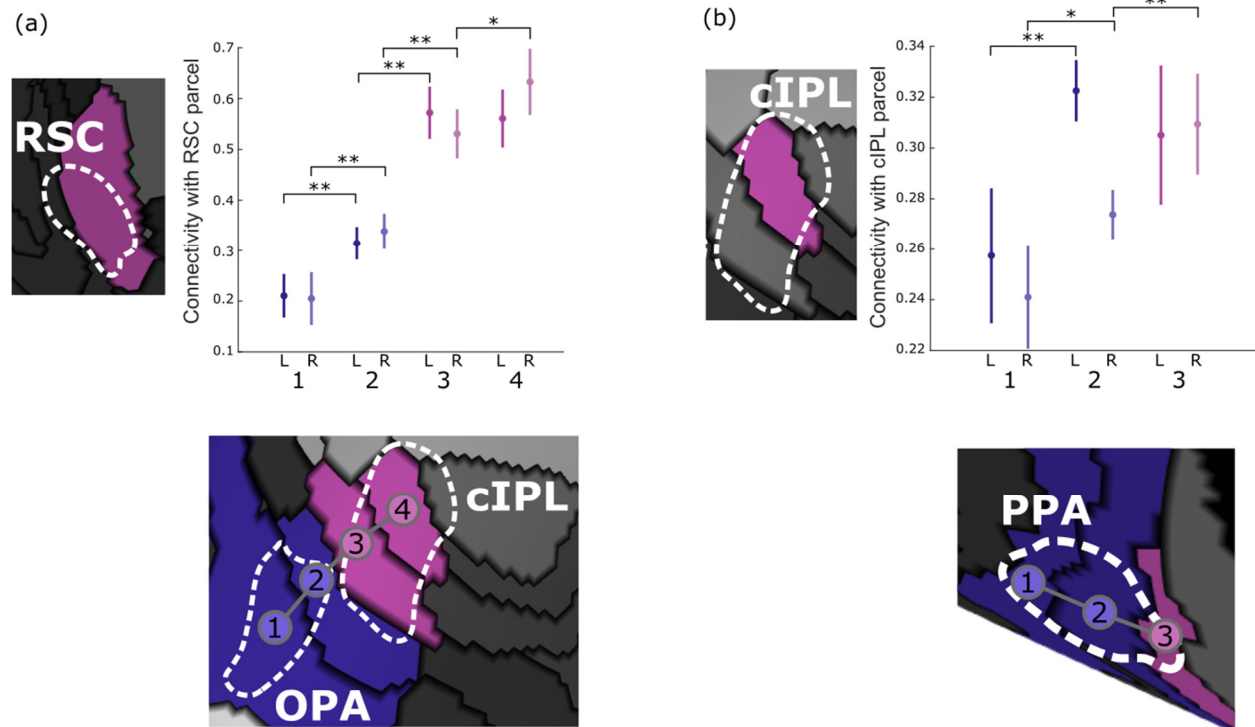


Figure 2: Connectivity shifts across the network border. (a) Dorsal parcels markedly increase their connectivity to the medial RSC parcel as we move anteriorly from parcels overlapping OPA to parcels overlapping cIPL, indicating a sharp transition between networks near the OPA/cIPL border. (b) Ventral parcels also show a shift in network connectivity properties, with increasing connectivity to the anterior cIPL parcel as we move from pPPA to aPPA. Error bars are 95% confidence intervals across subjects, * $p < 0.05$, ** $p < 0.01$.

Connectivity with the Hippocampus

Since the anterior scene-network overlaps with default mode regions, while the posterior scene-network does not, we predict that the anterior network should be more related to memory and navigation tasks that engage the hippocampus. To test this hypothesis, we measured the functional correlation at rest between mean hippocampal activity and the mean activity in each parcel within the posterior and anterior scene networks. As shown in Fig. 3, there is a dramatic difference in hippocampal connectivity for parcels in the posterior network (overlapping with OPA and posterior PPA) compared to the anterior network (overlapping with RSC, cIPL, and anterior PPA). Moving posterior to anterior along the dorsal path, hippocampal connectivity first decreases slightly (first parcel to second parcel, Left: $t_{19}=-3.04$, $p=0.007$; Right: $t_{19}=2.15$, $p<0.04$, two-tailed paired t-test), then increases significantly when moving to the third parcel (Left: $t_{19}=5.62$, $p<0.001$; Right: $t_{19}=3.79$, $p=0.001$) and to the fourth parcel (Left $t_{19}=4.17$, $p<0.001$; Right: $t_{19}=5.74$, $p<0.001$). Along the ventral path, hippocampal connectivity jumps from the first to the second parcel overlapping with PPA (Left: $t_{19}=5.27$, $p<0.001$; Right: $t_{19}=5.76$, $p<0.001$) and from the second to the third parcel (Right: $t_{19}=5.80$, $p<0.001$). This elevated hippocampal connectivity in the anterior network parcels is consistent with our hypothesis that the anterior network is more closely related to navigation and memory.

We also investigated whether this effect was being driven by a subregion of the hippocampus, by correlating the mean timecourse in both scene networks with the timecourses of each posterior-to-anterior coronal slice of the hippocampus. Our results show that the entire hippocampus is more strongly connected to the anterior scene-

network than the posterior scene-network, but this difference is especially large in the anterior hippocampus. To confirm this pattern of results, we divided the hippocampus into posterior and anterior subregions at $y=-21$ (Zeidman, Mullally, & Maguire, 2015) and correlated their mean timecourses with the two scene-network timecourses. This analysis confirmed that the anterior network is more strongly connected to both posterior ($t_{19}=7.66$, $p<0.001$, two-tailed paired t-test) and anterior ($t_{19}=6.58$, $p<0.001$) hippocampus, and that this anterior-network connectivity is larger in anterior hippocampus ($t_{19}=3.29$, $p=0.004$); a repeated-measures ANOVA shows significant main effects of both hippocampal subregion ($F_{1,19}=11.32$, $p=0.003$) and scene network ($F_{1,19}=59.2$, $p<0.001$), and an interaction ($F_{1,19}=7.03$, $p=0.016$). Note that both the anterior and posterior scene networks are closer to posterior hippocampus, ruling out a distance-based explanation for this pattern of results.

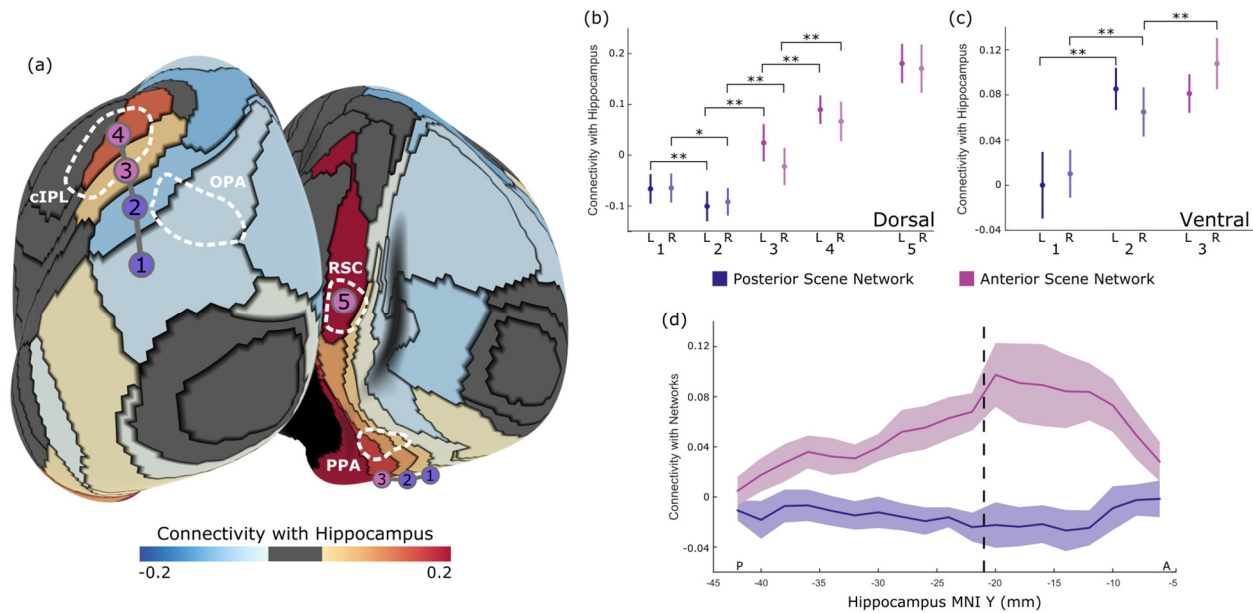


Figure 3: Connectivity between network parcels and the hippocampus. (a) For each parcel in the anterior and posterior scene networks, we computed its resting-state connectivity with the hippocampus, showing a striking increase in hippocampal activity for anterior network parcels overlapping with cIPL, RSC, and anterior PPA (magenta circles) compared to posterior network parcels (blue circles). (b) Along the dorsal network boundary, hippocampal activity first dips slightly and then increases substantially, becoming strongest in the most anterior parcel intersecting cIPL (and is also high in RSC). (c) Ventrally along parcels overlapping with PPA, we observe a similar posterior-anterior gradient in connectivity. (d) Computing the connectivity between each coronal slice of the hippocampus and the two scene networks shows that this increased coupling to the anterior network is present throughout the hippocampus, but is especially pronounced in anterior hippocampus (MNI $y > -21$ mm). Error bars are 95% confidence intervals across subjects, * $p < 0.05$, ** $p < 0.01$.

Comparison to Meta-analyses and Retinotopic Atlas

The connectivity results described thus far suggest a functional division for scene-related regions, with some belonging to a posterior network and others belonging to an anterior network. To assess the functional significance of these two networks, we ran two reverse-inference meta-analyses using the *NeuroSynth* tool (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011). This system automatically extracts activation coordinates from many fMRI studies (greater than 10,000 at the time of writing); given a particular set of studies, it can identify voxels that are *more likely* to be activated in this set of studies *relative to the full set of studies*. These voxels are therefore preferentially active in the query set compared to general fMRI experiments. Based on the areas involved, we hypothesize that the posterior network processes the current visual properties of the scene, whereas the anterior network incorporates episodic memories and contextual aspects of the scene. Thus, in Fig. 4a, we compare meta-analyses for the query “scene” (47 studies) with the query “episodic memory OR navigation OR past future” (125 studies). Along the parahippocampal gyrus, we find that the visual scene activations tend to be posterior to the memory activations, and that the transition point corresponds almost exactly to the division between our two networks. Dorsally, we also observe a separation between the reverse inference maps, with scene and memory activations falling into our two separate networks. Overall, voxels significant only in the scene meta-analysis were concentrated in the posterior network (66% in posterior network, 18% in anterior, 16% in other) while voxels significant only in the memory/navigation meta-analysis were spread more widely across the cortex, but were concentrated more in the anterior than posterior network (16% posterior, 42% anterior, 42% other). Voxels significant in both the scene and memory/navigation meta-

analyses tended to fall near the border between the two networks and divided roughly equally between them (44% posterior, 53% anterior, 4% other).

Another prediction of our framework is that voxels whose activity is tied to specific locations in the visual field (i.e. retinotopic) should, as clearly visual voxels, be present only in the posterior scene network. In Figure 4b, we compared our networks to a group-level probabilistic atlas of retinotopic visual field maps (Wang, Mruczek, Arcaro, & Kastner, 2014). The vast majority of the probability mass in this atlas is concentrated in the posterior network. In early visual cortex (V1, V2, V3, hV4) all non-foveal portions of the visual field maps fall in the posterior network (80% posterior, 0% anterior, 20% other). Ventrally, the posterior network covers VO1/2 (100% posterior, 0% anterior, 0% other), PHC1 (98% posterior, 2% anterior, 0% other), and the peak of the probability distribution for PHC2, which also extends slightly across the anterior network border (78% posterior, 22% anterior, 0% other). Laterally and dorsally, the posterior network includes most of the LO1/2 and TO1/2 maps (82% posterior, 0% anterior, 17% other), V3a and V3b (96% posterior, 0% anterior, 3% other), and IPS0-IPS5 (68% posterior, 4% anterior, 28% other), with SPL1 being the only map falling substantially outside the networks we consider (18% posterior, 2% anterior, 80% other).

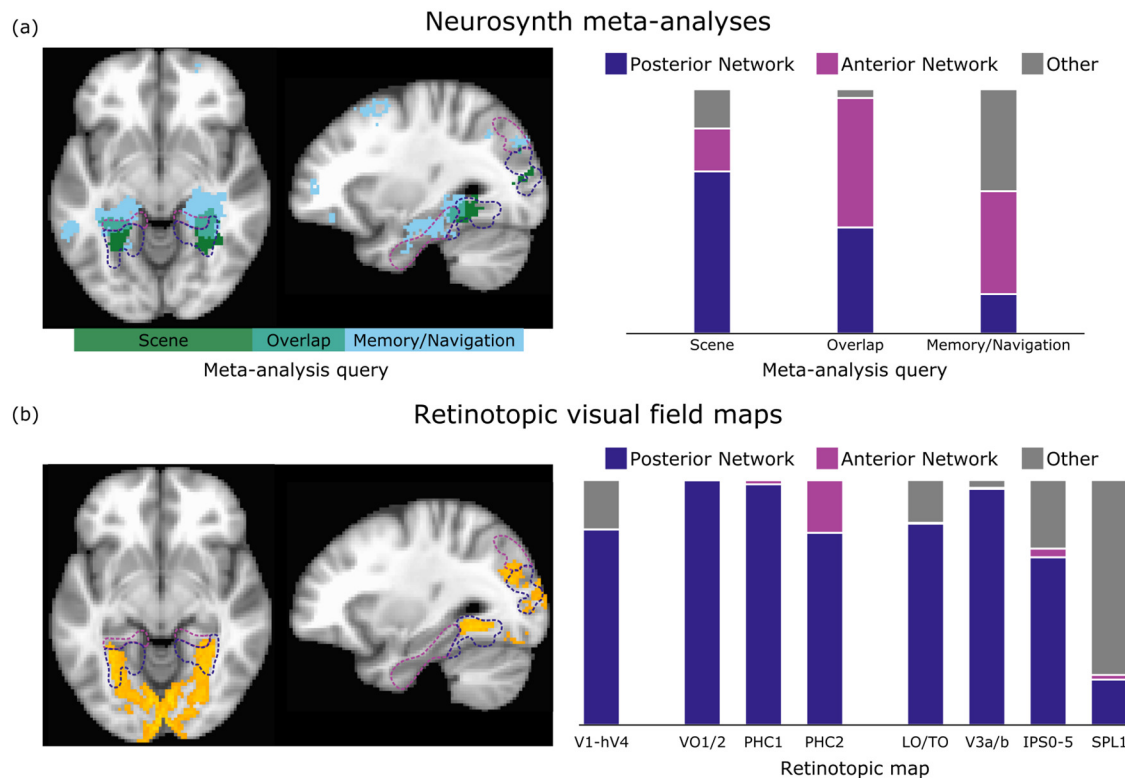


Figure 4: **Overlap of posterior and anterior scene networks with previous work.** (a)

Two meta-analyses conducted using *NeuroSynth* identified overlapping but distinct reverse-inference maps corresponding to studies of visual scenes and to studies of higher-level memory and navigation tasks. These maps separate into our two scene networks, with visual scenes activating voxels in the posterior network and memory/navigation tasks activating voxels in the anterior network, as shown on example axial ($z=-8$) and sagittal ($x=-30$) slices. $FDR < 0.01$, cluster size 80 voxels (640 mm^3). (b) Voxels having a greater than 50% chance of belong to a retinotopic map (orange) overlap with much of the posterior scene network, but end near the border of the anterior scene network. Breaking up the contributions of individual regions, we find that the probability mass of the topographic maps falls primarily within the posterior network, with only PHC2 showing a small overlap with the anterior network (probabilistically at the group level).

Discussion

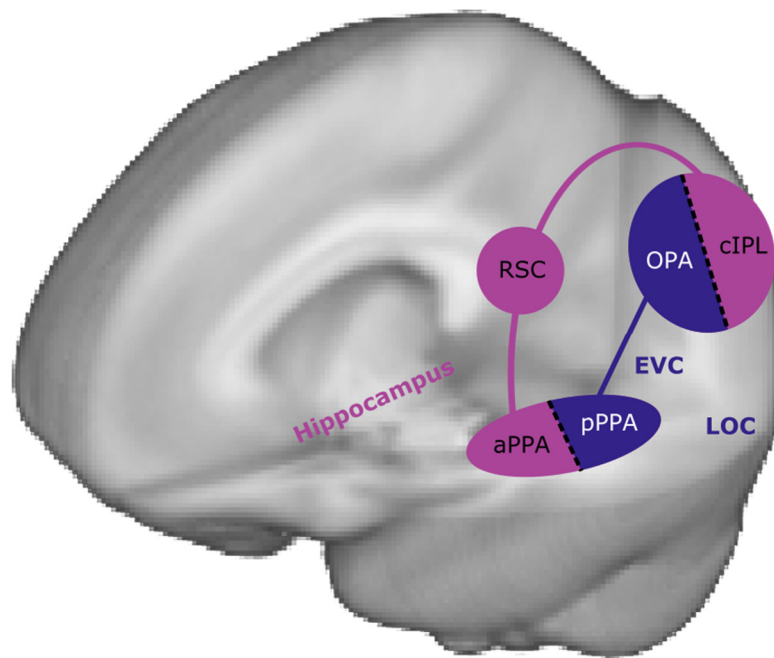


Figure 5: **Two-network model of scene perception.** Our results provide strong evidence for dividing scene-sensitive regions into two separate networks. We argue that OPA and posterior PPA (PHC1/2) process the current visual features of a scene (in concert with other visual areas, such early visual cortex, EVC, and lateral occipital cortex, LOC), while cIPL, RSC, and anterior PPA perform higher-level context and navigation tasks (drawing on long-term memory structures such as the hippocampus).

By combining a variety of data sources, we have shown converging evidence for a functional division of scene-processing regions into two separate networks (summarized in Figure 5). The posterior visual network covers retinotopically-organized regions including OPA and posterior PPA (pPPA), while an anterior memory-related network connects cIPL, RSC, and anterior PPA (aPPA). This division emerges from a purely data-driven network clustering, suggesting that this is a core organizing principle of the visual system.

Subdivisions of the PPA

The division of the PPA into multiple anterior-posterior subregions with differing connectivity properties replicates our previous work (Baldassano et al., 2013) on an entirely different large-scale dataset, and shows that there is a strong connection between connectivity changes in PPA and the boundaries of retinotopic field maps. There is now a growing literature on anterior versus posterior PPA, including not only connectivity differences (Nasr et al., 2013) but also the response to low-level (Baldassano, Beck, & Fei-Fei, 2016; Baldassano, Fei-Fei, & Beck, 2016; Nasr, Echavarria, & Tootell, 2014; E. H. Silson, Chan, Reynolds, Kravitz, & Baker, 2015; Watson, Hymers, Hartley, & Andrews, 2016) and high-level (Aminoff & Tarr, 2015; Linsley & Macevoy, 2014; Marchette, Vass, Ryan, & Epstein, 2015; Park, Konkle, & Oliva, 2014) scene features, as well as stimulation studies (Rafique, Solomon-Harris, & Steeves, 2015). Our results place this division into a larger context, and demonstrate that the connectivity differences within PPA are not just an isolated property of this region but a general organizing principle for scene-processing regions.

The visual network

The visual network shows a close correspondence with the full set of retinotopic maps identified in previous studies (Brewer & Barton, 2012; Huang & Sereno, 2013; Wang et al., 2014). Previous measurements in individual subjects have also shown strong overlap between OPA and retinotopic maps, especially V3b and LO2 (Bettencourt & Xu, 2013; Nasr et al., 2011; Edward H Silson, Groen, Kravitz, & Baker, 2016), and between pPPA and VO2, PHC1, and PHC2 (Arcaro, McMains, Singer, & Kastner, 2009). The only portion of cortex with known retinotopic maps that is not clustered in this network is the shared foveal representation of early visual areas, which segregates into its own cluster, consistent with other work showing a peripheral eccentricity bias in the scene network (Baldassano, Fei-Fei, et al., 2016; Goesaert & Op de Beeck, 2010; Huang & Sereno, 2013; Malach, Levy, & Hasson, 2002).

OPA and posterior PPA have been shown to be closely related to the visual content of a stimulus. Even low-level manipulations of spatial frequency (Kauffmann, Ramanoël, Guyader, Chauvin, & Peyrin, 2015; Rajimehr, Devaney, Bilenko, Young, & Tootell, 2011; Watson et al., 2016) or rectilinearity (Nasr et al., 2014) can drive responses in these regions. Higher-level visual features also drive response patterns in these regions (Bryan, Julian, & Epstein, 2016), and they are hypothesized to be involved in extracting visual environmental features that can be used for navigation (Julian, Ryan, Hamilton, & Epstein, 2016; Marchette et al., 2015). However, neither OPA nor posterior PPA show reliable familiarity effects (Epstein et al., 2007; see further discussion below).

The functional distinction between pPPA and OPA is currently unclear. Previous work has speculated about the purpose of the apparent ventral and dorsal “duplication” of regions sensitive to large landmarks, proposing that it may be related to different output goals (e.g. action planning in OPA, object recognition in pPPA) (Konkle & Caramazza, 2013), or to different input connections (e.g. lower visual field processing in OPA, upper visual field processing in pPPA) (Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013; E. H. Silson et al., 2015). OPA and pPPA may also use information from different visual eccentricities, with OPA processing less peripheral, relatively high-resolution environmental features and pPPA processing more peripheral, large-scale geometry and context (Baldassano, Fei-Fei, et al., 2016).

The memory and navigation network

The network of parahippocampal, retrosplenial, and posterior parietal regions we identify has been emerged independently in many different fields of neuroimaging, outside of scene perception. Meta-analyses of internally-directed tasks such as theory of mind, autobiographical memory, and prospection have identified this as a core, re-occurring network (Kim, 2010; Spreng, Mar, & Kim, 2009; component C10 of Yeo et al., 2014). It comprises a subset of the broader default mode regions, but functional and anatomical evidence suggests that it is a distinct, coherent subnetwork (Andrews-Hanna et al., 2010; Andrews-Hanna, Smallwood, & Spreng, 2014; Yeo et al., 2011). The broad set of tasks which recruit this network have been summarized in various ways, such as “scene construction” (Hassabis & Maguire, 2007), “mnemonic scene construction” (Andrews-Hanna et al., 2010), “long-timescale integration” (Hasson, Chen, & Honey, 2015), or

“relational processing” (Eichenbaum & Cohen, 2014). A review of memory studies referred to this network as the posterior medial (PM) memory system, and proposed that it is involved in any task requiring “situation models” relating entities, actions, and outcomes (Ranganath & Ritchey, 2012).

The network has strong functional connections to the hippocampus, which has been implicated in a broad set of cognitive tasks involving “cognitive maps” for organizing declarative memories, spatial routes, and even social dimensions (Eichenbaum & Cohen, 2014; Schiller et al., 2015). During perception, the hippocampus binds together visual elements of an image (Olsen, Moses, Riggs, & Ryan, 2012; Warren, Duff, Jensen, Tranel, & Cohen, 2012; Zeidman et al., 2015), which is especially important for scene stimuli (Graham et al., 2006; Hodgetts, Shine, Lawrence, Downing, & Graham, 2016; Lee, Buckley, et al., 2005; Lee, Bussey, et al., 2005), and then stores this representation into long-term memory (Ryan & Cohen, 2004). As we become familiar with an environment, the hippocampus builds a map of the spatial relationships between visual landmarks, which is critical for navigation (Morgan, Macevoy, Aguirre, & Epstein, 2011). Recalling or even imagining scenes also engages the hippocampus, especially anterior hippocampus, which may serve to integrate memory and spatial information (Zeidman & Maguire, 2016). Our results argue that only the anterior scene regions are directly involved in building hippocampal representations of the environment, and in retrieving relevant memories and navigational information for a presented or imagined scene.

The specific functions of the individual components of this network have also been studied in a number of contexts. RSC appears to be most directly involved in orienting the viewer to the structure of the environment (both within and beyond the borders of the presented image) for the purpose of navigational planning; it encodes both absolute location and facing direction (Epstein & Vass, 2014; Marchette, Vass, Ryan, & Epstein, 2014; Vass & Epstein, 2013), integrates across views presented in a panoramic sequence (Park & Chun, 2009), and shows strong familiarity effects (Epstein et al., 2007; Epstein, Parker, & Feiler, 2007). This is consistent with rodent neurophysiological studies, which have identified head direction cells in this region (Chen, Lin, Green, Barnes, & McNaughton, 1994). RSC is not sensitive to low-level rectilinear features in non-scene images such as objects or textures (Nasr et al., 2014), though it does show some preference for rectilinear features in images of 3D scenes (Nasr et al., 2014; Watson et al., 2016).

The specific properties of anterior PPA have been less well-studied, since it was not recognized as a separate region within the PPA until recently. It has been shown to be driven more by high-level category information than spatial frequency content (Watson et al., 2016), to represent real-world locations (even from perceptually-distinct views) (Marchette et al., 2015), to encode object co-occurrences (Aminoff & Tarr, 2015), and to represent real-world physical scene size (Park et al., 2014). Its representation of scene spaciousness draws on prior knowledge about the typical size of different scene categories, since it is affected by the presence of diagnostic objects (Linsley & Macevoy, 2014).

The cIPL (also referred to as pIPL, PGp, or the angular gyrus) has been proposed as a “cross-modal hub” (Andrews-Hanna et al., 2014) that connects visual information with other sensory modalities as well as knowledge of the past. It is more intimately associated with visual cortex than most lateral parietal regions, since it has strong anatomical connections to higher-level visual regions in humans and macaques (Caspers et al., 2011), and has a neurotransmitter receptor distribution similar to V3v and distinct from the rest of the IPL (Caspers et al., 2012). It has been mostly ignored in the scene perception literature, primarily because it is not strongly responsive to standard scene localizers that show sequences of unfamiliar and unrelated scene images. For example, a study showing familiarity effects in cIPL described this location only as “near TOS” (Epstein et al., 2007). The cIPL appears commonly, however, in studies involving personally familiar places, which are associated with a wealth of memory, context, and navigational information. It is involved in memory for visual scene images (Elman, Rosner, Cohn-Sheehy, Cerreta, & Shimamura, 2013; Montaldi, Spencer, Roberts, & Mayes, 2006; Takashima et al., 2006; van Assche, Kebets, Vuilleumier, & Assal, 2014), learning navigational routes (Bray, Arnold, Levy, & Iaria, 2014; Burgess, Maguire, Spiers, & O’Keefe, 2001), and even imagining past events or future events in familiar places (Hassabis, Kumaran, & Maguire, 2007; Szpunar, Chan, & McDermott, 2009). It can integrate information across space (Livne & Bar, 2016) and time (Lerner, Honey, Silbert, & Hasson, 2011; Vilberg & Rugg, 2012), and has been shown in lesion studies to be critical for orientation and navigation (Kravitz et al., 2011). Our connectivity results and meta-analysis suggest that cIPL may play a prominent role in connecting visual scenes to the real-world location they depict.

Contrasting the two networks

Although our work is the first to propose the visual versus context networks as a general framework for scene perception, several previous studies have shown differential effects within these two networks. Contrasting the functional connectivity patterns of RSC vs. OPA or LOC (Nasr et al., 2013) or anterior vs. posterior PPA (Baldassano et al., 2013) show a division between the two networks, consistent with our results. Contrasting scene-specific activity with general (image or word) memory retrieval showed an anterior vs. posterior distinction in PPA and cIPL/OPA, with only more anterior regions (aPPA and cIPL, along with RSC) responding to content-independent retrieval tasks (Fairhall et al., 2014; Johnson & Rugg, 2007). Our two-network division is also consistent with the “dual intertwined rings” model, which argues for a high-level division of cortex into a sensory ring and an association ring, the second of which is distributed but connected into a continuous ring through fiber tracts (Mesmoudi et al., 2013).

Open questions

The anterior/posterior pairing of aPPA/pPPA and cIPL/OPA raises the question of whether there is a similar anterior/posterior division in RSC. There is some evidence to suggest that this is the case: wide-field retinotopic mapping using natural scenes shows a partial retinotopic organization in RSC (Huang & Sereno, 2013), and RSC’s response to visual rectilinear features appears to be limited to the posterior portion (Nasr et al., 2014). However a study of retinotopic coding in scene-selective regions failed to find any consistent topographic organization to RSC responses (Ward, MacEvoy, & Epstein, 2010), and previous analyses of the functional properties of anterior versus posterior RSC have

not found any significant differences (Park et al., 2014). More precise functional mapping of the regions near the parieto-occipital sulcus will be required to see if a retinotopic subregion is present in this scene-sensitive region as well.

Another interesting question is how spatial reference frames differ between and within the two networks. Given its retinotopic fieldmaps, the visual network presumably represents scene information relative to the current eye position; previous work has argued that this reference frame is truly retina-centered and not egocentric (Golomb & Kanwisher, 2012; Ward et al., 2010). The context network, however, likely transforms information between multiple reference frames. Models of spatial memory suggest that medial temporal lobe (possibly including aPPA) utilizes an allocentric representation, while the posterior parietal lobe (possibly including cIPL) is based on an egocentric reference frame, and that the two are connected via a transformation circuit in RSC that combines allocentric location and head direction (Byrne, Becker, & Burgess, 2007; Vann, Aggleton, & Maguire, 2009). There is some recent evidence for this model in human neuroimaging: posterior parietal cortex codes the direction of attention in an egocentric reference frame (even for positions outside the field of view) (Schindler & Bartels, 2013), and RSC contains both position and head direction information (anchored to the local environment) (Marchette et al., 2014). This raises the possibility that another critical role of cIPL could be to transform retinotopic visual information into a stable egocentric scene over the course of multiple eye movements. The properties of aPPA, however, are much less clear; it seems unlikely that it would utilize an entirely different coordinate system than neighboring PHC1/2, and some

aspects of the scene encoded in aPPA such as object co-occurrence (Aminoff & Tarr, 2015) do not seem tied to any particular coordinate system.

Conclusion

Based on data-driven connectivity analyses and analysis of previous literature, we have proposed a unifying framework for understanding the neural systems involved in processing both visual and non-visual properties of natural scenes. This new two-network classification system makes explicit the relationships between known scene-sensitive regions, re-emphasizes the importance of the functional subdivision within the PPA, and incorporates posterior parietal cortex as a primary component of the scene-understanding system. Our proposal, that much of the scene-processing network relates more to contextual and navigational information than to specific visual features, suggests that experiments with unfamiliar natural scene images will give only a partial picture of the neural processes evoked in real-world places. Experiencing our visual environment requires a dynamic cooperation between distinct cortical systems, to extract information from the current view of a scene and then integrate it with our understanding of the world and determine our place in it.

Materials and Methods

Imaging Data

The majority of the data used in this study was obtained from the Human Connectome Project (HCP), which provides detailed documentation on the experimental and acquisition parameters for these datasets (Van Essen et al., 2013). We provide an overview of these datasets below.

The group-level functional connectivity data were derived from the 468-subject group-PCA eigenmaps, distributed with the June 2014 “500 Subjects” HCP data release. Resting-state fMRI data were acquired over four sessions (14 min, 33 seconds each) while subjects fixed on a bright cross-hair on a dark background, using a multiband sequence to achieve a TR of 720ms at 2.0mm isotropic resolution (59412 surface vertices). These timecourses were cleaned using FMRIB's ICA-based Xnoiseifier (FIX) (Salimi-Khorshidi et al., 2014), and then the top 4500 eigenvectors for each voxel were estimated across all subjects using Group-PCA (Smith, Hyvärinen, Varoquaux, Miller, & Beckmann, 2014). This data was used to perform the parcellation and network clustering, and to generate whole-brain maps (i.e. Figs 1, 3a)

For the first 20 subjects within the “500 Subjects” release with complete data (subj ids 101006, 101107, 101309, 102008, 102311, 103111, 104820, 105014, 106521, 107321, 107422, 108121, 108323, 108525, 108828, 109123, 109325, 111413, 113922, 120515), we created individual subject resting-state datasets by demeaning and concatenating their

four resting-state sessions. This data was used to statistically measure the robustness of connectivity differences observed in the group-level data (i.e. Figs 2, 3b-d). We also obtained these subjects' data from the HCP Working Memory experiment, in which they observed blocks of stimuli consisting of faces, places, tools, or body parts. We collapsed across the two memory tasks being performed by participants (target-detection or 2-back detection).

To identify group-level scene localizers, we used data from a separate set of 24 subjects scanned at Stanford University (see below). Each subject viewed blocks of stimuli from six categories: child faces, adult faces, indoor scenes, outdoor scenes, objects (abstract sculptures with no semantic meaning), and scrambled objects. Functional data were acquired with an in-plane resolution of 1.56mm, slice thickness of 3mm (with 1 mm gap), and a TR of 2s; a high-resolution (1mm isotropic) SPGR structural scan was also acquired to allow for transformation to MNI space. Full details of the localizer stimuli and acquisition parameters are given in our previous work (Baldassano et al., 2013).

The cIPL was defined using the Eickhoff-Zilles PGp probabilistic cytoarchitectonic map (Eickhoff et al., 2005) as described in our previous work (Baldassano et al., 2013). The hippocampus was divided into anterior and posterior subregions at MNI y=-21, consistent with previous studies (Zeidman et al., 2015).

Subjects

Scene localizer data was collected from 24 subjects (6 female, ages 22-32, including one of the authors). Subjects were in good health with no past history of psychiatric or neurological diseases, and with normal or corrected-to-normal vision. The experimental protocol was approved by the Institutional Review Board of Stanford University, and all subjects gave their written informed consent.

Resting-state Parcellation

We generated a voxel-level functional connectivity matrix by correlating the group-level eigenmaps for every pair of voxels and applying the arctangent function. We parcellated this 59412 by 59412 matrix into contiguous regions, using a generative probabilistic model (Baldassano et al., 2015). This method finds a parcellation of the cortex such that the connectivity properties within each parcel are as uniform as possible, making multiple passes over the dataset to fine-tune the parcel borders. We set the scaling hyperparameter $\sigma_0^2 = 3000$ to produce a manageable number of parcels, but our results are similar for a wide range of settings for σ_0^2 (see Supplementary Fig. 2).

Scene localizers

To identify PPA, RSC, and OPA, we deconvolved the localizer data from the 24 Stanford subjects using the standard block hemodynamic model in AFNI (Cox, 1996), with faces, scenes, objects, and scrambled objects as regressors. The Scenes > Objects t-statistic was used to define PPA (top 300 voxels near the parahippocampal gyrus), RSC (top 200 voxels

near retrosplenial cortex), and OPA (top 200 voxels near the transverse occipital sulcus). The ROI masks were then transformed to MNI space, summed across all subjects, and mapped to the closest vertices on the group cortical surface. The cluster denoting highest overlap between subjects was then manually annotated.

Parcel-to-parcel and hippocampal functional connectivity

The 468-subject eigenmaps distributed by the HCP are approximately equal to performing a singular value decomposition on the concatenated timecourses of all 468 subjects, and then retaining the right singular values scaled by their eigenvalues (Smith et al., 2014). This allows us to treat these eigenmaps as pseudo-timecourses, since dot products (and thus correlations) between eigenmaps approximate the dot products between the original voxel timecourses. Given a parcellation, we computed the group-level connectivity between a pair of regions by taking the mean over all eigenmaps in each region, then correlating these mean eigenmaps and applying the Fisher z-transform (hyperbolic arctangent). We computed subject-level connectivity in the same way, using the resting-state timecourse for each voxel rather than the eigenmap.

Connectivity between cortical parcels and the hippocampus was computed similarly, using eigenmaps (for group data) or timecourses (for subject data) extracted from the hippocampal volume data distributed by the HCP. In order to focus on hippocampal connectivity differences among parcels, we used the mean gray timecourse regression (MGTR) version of the group data and regressed out the global timecourse from the subject data.

Multidimensional scaling and Network Clustering

The 172 by 172 parcel functional connectivity matrix was converted into a distance matrix by subtracting every entry from the maximum entry. Classical multidimensional scaling was applied to the distance matrix, and the first three dimensions were used to assign voxels RGB colors (with each color channel scaled to span the full range of 0 to 255 along each axis) and to plot parcels in a 3D space. We performed the same operation on each subject-level matrix as well, and then aligned each subject's 3D pointcloud to the group pointcloud using a procrustes transform. Hierarchical ward clustering (unconstrained by parcel position) was also applied to the distance matrix to compute a hard clustering into 10 networks.

Meta-analysis and retinotopic field maps

Two reverse-inference meta-analyses were performed using the *NeuroSynth* website (Yarkoni et al., 2011). *NeuroSynth* is a set of open-source python tools for automatically extracting data from fMRI studies and computing activation likelihood maps, and the website hosts these tools (and associated datasets) for public use. Supplying a keyword query will identify all studies in which that query appears frequently, and then analyzes the activations reported in these queried studies. In addition to standard "forward inference" maps giving the probability $p(\text{activation}|\text{query})$ that a voxel will be activated in these studies, *NeuroSynth* generates "reverse inference" maps giving the probability $p(\text{query}|\text{activation})$ that a voxel activation came specifically from this query set. Voxels appearing the reverse inference map therefore appear more often in the query set relative

to the full set of (>10,000) fMRI studies in the database. This accounts for base rate differences in how often activation is observed in different brain regions. Our meta-analyses can be viewed online at <http://neurosynth.org/analyses/custom/dda0e003-efd0-4cfa/> and <http://neurosynth.org/analyses/custom/9e6df59d-02df-4357/>. A volumetric group-level probabilistic atlas (Wang et al., 2014) was used to define retinotopic field maps. We computed the total probability mass of each map that fell within one of our two networks or in other regions of the cortex, and then normalized the sum of the three values to 100%. For visualization, the probability that a voxel belongs to any field map was computed as $1 - \prod_i (1 - p_i)$ where p_i is the probability that the voxel falls within field map i .

Parcel scene selectivity

In order to validate that our group-level functional scene localizer (from the Stanford subjects) properly identified scene-related parcels in the HCP subjects, we used data from Working Memory experiment performed by the 20 HCP subjects. We first used a hemodynamic model to associate timepoints within the working memory experiment with specific stimulus categories. We labeled timepoints as corresponding to bodies, faces, places, or tools by constructing a boxcar timecourse denoting when each stimulus category was being displayed, convolving these indicators with the standard SPM hemodynamic response function provided with AFNI (Cox, 1996), rescaling the maximum value to 1, then re-thresholding to a binary indicator. Effectively, this produced a shift of the stimulus blocks by 5.55s to account for hemodynamic delay. The fMRI timecourses were cleaned by regressing out movement (6 degree-of-freedom translation/rotation and derivatives) and

constant, linear, and quadratic trends from each run, then normalizing each voxel to have unit variance. Voxel timecourses were then averaged within each parcel, yielding a vector of average parcel activities for each timepoint. The mean parcel activation to scenes was compared to the mean parcel activation to all other categories, and this difference was statistically tested across the 20 subjects for each parcel. The significance threshold was corrected for multiple comparisons using the same false discovery rate (FDR) procedure implemented in AFNI's 3dFDR (Cox, 1996). The results are shown in Supplementary Figure 3.

Acknowledgements

Funding was provided by a National Science Foundation Graduate Research Fellowship (to CB) under grant number DGE-0645962, and Office of Naval Research Multidisciplinary University Research Initiative (to DMB and LF) grant number N000141410671. Data were provided in part by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We thank the Richard M. Lucas Center for Imaging, the Center for Cognitive and Neurobiological Imaging, and Michael Arcaro for helpful discussions.

References

Aminoff, E. M., & Tarr, M. J. (2015). Associative Processing Is Inherent in Scene Perception. *Plos One*, 10(6), e0128840. <http://doi.org/10.1371/journal.pone.0128840>

- Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, 65(4), 550–62. <http://doi.org/10.1016/j.neuron.2010.02.005>
- Andrews-Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences*, 1316, 29–52. <http://doi.org/10.1111/nyas.12360>
- Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience*, 29(34), 10638–10652. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/19710316>
- Baker, K. (2012). Identity, Memory and Place. *The Word Hoard*, 1(1), Article 4. Retrieved from <http://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1003&context=wordhoard>
- Baldassano, C., Beck, D. M., & Fei-Fei, L. (2013). Differential connectivity within the Parahippocampal Place Area. *NeuroImage*, 75(0), 228–237. <http://doi.org/10.1016/j.neuroimage.2013.02.073>
- Baldassano, C., Beck, D. M., & Fei-Fei, L. (2015). Parcellating connectivity in spatial maps. *PeerJ*, 3(e784). <http://doi.org/10.7717/peerj.784>
- Baldassano, C., Beck, D. M., & Fei-Fei, L. (2016). Human–Object Interactions Are More than the Sum of Their Parts. *Cerebral Cortex*, bhw077. <http://doi.org/10.1093/cercor/bhw077>
- Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *Journal of Vision*, 16(2), 9. <http://doi.org/10.1167/16.2.9>
- Bettencourt, K. C., & Xu, Y. (2013). The role of transverse occipital sulcus in scene perception and its relationship to object individuation in inferior intraparietal sulcus. *Journal of Cognitive Neuroscience*, 25, 1711–1722. <http://doi.org/10.1162/jocn>
- Bray, S., Arnold, A. E. G. F., Levy, R. M., & Iaria, G. (2014). Spatial and temporal functional connectivity changes between resting and attentive states. *Human Brain Mapping*, 00(September). <http://doi.org/10.1002/hbm.22646>
- Brewer, A., & Barton, B. (2012). Visual field map organization in human visual cortex. In S. Molotchnikoff & J. Rouat (Eds.), *Visual Cortex-Current Status and Perspectives*. Retrieved from http://www.researchgate.net/publication/234038943_Visual_field_map_organization_in_human_visual_cortex/file/504635279e1482e2f1.pdf
- Bryan, P. B., Julian, J. B., & Epstein, R. A. (2016). Rectilinear Edge Selectivity Is Insufficient to Explain the Category Selectivity of the Parahippocampal Place Area. *Frontiers in Human Neuroscience*, 10(March), 1–12. <http://doi.org/10.3389/fnhum.2016.00137>
- Burgess, N., Maguire, E. a, Spiers, H. J., & O'Keefe, J. (2001). A temporoparietal and prefrontal network for retrieving the spatial context of lifelike events. *NeuroImage*, 14(2), 439–53. <http://doi.org/10.1006/nimg.2001.0806>
- Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future:

- a neural model of spatial memory and imagery. *Psychological Review*, 114(2), 340–75.
<http://doi.org/10.1037/0033-295X.114.2.340>
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology*, 197(3), 551–66. Retrieved from
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1351748&tool=pmcentrez&rendertype=abstract>
- Caspers, S., Eickhoff, S. B., Rick, T., von Kapri, A., Kuhlen, T., Huang, R., ... Zilles, K. (2011). Probabilistic fibre tract analysis of cytoarchitectonically defined human inferior parietal lobule areas reveals similarities to macaques. *NeuroImage*, 58(2), 362–80.
<http://doi.org/10.1016/j.neuroimage.2011.06.027>
- Caspers, S., Schleicher, A., Bacha-Trams, M., Palomero-Gallagher, N., Amunts, K., & Zilles, K. (2012). Organization of the Human Inferior Parietal Lobule Based on Receptor Architectonics. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–14.
<http://doi.org/10.1093/cercor/bhs048>
- Chen, L. L., Lin, L. H., Green, E. J., Barnes, C. A., & McNaughton, B. L. (1994). Head-direction cells in the rat posterior cortex. I. Anatomical distribution and behavioral modulation. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, 101(1), 8–23. <http://doi.org/10.1007/BF00243213>
- Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, 29(3), 162–73. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8812068>
- Eichenbaum, H., & Cohen, N. J. (2014). Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? *Neuron*, 83(4), 764–770.
<http://doi.org/10.1016/j.neuron.2014.07.032>
- Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, 25(4), 1325–35.
<http://doi.org/10.1016/j.neuroimage.2004.12.034>
- Elman, J. A., Rosner, Z. A., Cohn-Sheehy, B. I., Cerreta, A. G., & Shimamura, A. P. (2013). Dynamic changes in parietal activation during encoding: Implications for human learning and memory. *NeuroImage*, 82, 44–52.
<http://doi.org/10.1016/j.neuroimage.2013.05.113>
- Epstein, R. a, Parker, W. E., & Feiler, A. M. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 27(23), 6141–6149.
<http://doi.org/10.1523/JNEUROSCI.0799-07.2007>
- Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology*, 97(5), 3670–83.
<http://doi.org/10.1152/jn.00003.2007>
- Epstein, R. A., & Vass, L. K. (2014). Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1635).

- 709 Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual
710 environment. *Nature*, 392(6676), 598–601. <http://doi.org/10.1038/33402>
- 711 Fairhall, S. L., Anzellotti, S., Ubaldi, S., & Caramazza, A. (2014). Person- and place-selective
712 neural substrates for entity-specific semantic access. *Cerebral Cortex (New York, N.Y. :
713 1991)*, 24(7), 1687–96. <http://doi.org/10.1093/cercor/bht039>
- 714 Goesaert, E., & Op de Beeck, H. P. (2010). Continuous mapping of the cortical object vision
715 pathway using traveling waves in object space. *NeuroImage*, 49(4), 3248–56.
716 <http://doi.org/10.1016/j.neuroimage.2009.11.036>
- 717 Golomb, J. D., & Kanwisher, N. (2012). Higher level visual cortex represents retinotopic, not
718 spatiotopic, object location. *Cerebral Cortex*, 22(December), 2794–2810.
719 <http://doi.org/10.1093/cercor/bhr357>
- 720 Graham, K. S., Scahill, V. L., Hornberger, M., Barense, M. D., Lee, A. C. H., Bussey, T. J., &
721 Saksida, L. M. (2006). Abnormal categorization and perceptual learning in patients
722 with hippocampal damage. *The Journal of Neuroscience : The Official Journal of the
723 Society for Neuroscience*, 26(29), 7547–54. [http://doi.org/10.1523/JNEUROSCI.1535-
06.2006](http://doi.org/10.1523/JNEUROSCI.1535-
724 06.2006)
- 725 Hassabis, D., Kumaran, D., & Maguire, E. a. (2007). Using imagination to understand the
726 neural basis of episodic memory. *The Journal of Neuroscience : The Official Journal of
727 the Society for Neuroscience*, 27(52), 14365–74.
728 <http://doi.org/10.1523/JNEUROSCI.4549-07.2007>
- 729 Hassabis, D., & Maguire, E. a. (2007). Deconstructing episodic memory with construction.
730 *Trends in Cognitive Sciences*, 11(7), 299–306.
731 <http://doi.org/10.1016/j.tics.2007.05.001>
- 732 Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: Memory as an
733 integral component of information processing. *Trends in Cognitive Sciences*, 19(6),
734 304–313. <http://doi.org/10.1016/j.tics.2015.04.006>
- 735 Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry
736 organization of human occipito-temporal object areas. *Neuron*, 37, 1027–1041.
737 [http://doi.org/10.1016/S0896-6273\(03\)00144-2](http://doi.org/10.1016/S0896-6273(03)00144-2)
- 738 Hodgetts, C. J., Shine, J. P., Lawrence, A. D., Downing, P. E., & Graham, K. S. (2016).
739 Evidencing a place for the hippocampus within the core scene processing network.
740 *Human Brain Mapping*, 00. <http://doi.org/10.1002/hbm.23275>
- 741 Huang, R.-S., & Sereno, M. I. (2013). Bottom-up Retinotopic Organization Supports Top-
742 down Mental Imagery. *The Open Neuroimaging Journal*, 7(858), 58–67.
743 <http://doi.org/10.2174/1874440001307010058>
- 744 Johnson, J. D., & Rugg, M. D. (2007). Recollection and the reinstatement of encoding-related
745 cortical activity. *Cerebral Cortex (New York, N.Y. : 1991)*, 17(11), 2507–15.
746 <http://doi.org/10.1093/cercor/bhl156>
- 747 Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The Occipital Place Area Is
748 Causally Involved in Representing Environmental Boundaries during Navigation.
749 *Current Biology*, 26(8), 1104–1109. <http://doi.org/10.1016/j.cub.2016.02.066>

- Kaplan, E. (2004). The M, P, and K Pathways of the Primate Visual System. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neuroscience* (pp. 481–494). Cambridge, MA: MIT Press.
- Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, 112, 86–95. <http://doi.org/10.1016/j.neuroimage.2015.02.058>
- Kim, H. (2010). Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. *NeuroImage*, 50(4), 1648–57. <http://doi.org/10.1016/j.neuroimage.2010.01.051>
- Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 33(25), 10235–42. <http://doi.org/10.1523/JNEUROSCI.0983-13.2013>
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4), 217–230. <http://doi.org/10.1038/nrn3008>
- Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, 17(1), 26–49. <http://doi.org/10.1016/j.tics.2012.10.011>
- Lee, A. C. H., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., ... Graham, K. S. (2005). Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus*, 15(6), 782–797. <http://doi.org/10.1002/hipo.20101>
- Lee, A. C. H., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., ... Graham, K. S. (2005). Perceptual deficits in amnesia: Challenging the medial temporal lobe “mnemonic” view. *Neuropsychologia*, 43(1), 1–11. <http://doi.org/10.1016/j.neuropsychologia.2004.07.017>
- Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 31(8), 2906–15. <http://doi.org/10.1523/JNEUROSCI.3684-10.2011>
- Linsley, D., & Macevoy, S. P. (2014). Encoding-Stage Crosstalk Between Object- and Spatial Property-Based Scene Processing Pathways. *Cerebral Cortex (New York, N.Y. : 1991)*. <http://doi.org/10.1093/cercor/bhu034>
- Livne, T., & Bar, M. (2016). Cortical Integration of Contextual Information across Objects. *Journal of Cognitive Neuroscience*, 25(9), 1–11. http://doi.org/10.1162/jocn_a_00944
- Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4), 176–184. [http://doi.org/10.1016/S1364-6613\(02\)01870-3](http://doi.org/10.1016/S1364-6613(02)01870-3)
- Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2014). Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. *Nature Neuroscience*, 17(11), 1598–1606. <http://doi.org/10.1038/nn.3834>
- Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside Looking In: Landmark

791 Generalization in the Human Navigational System. *Journal of Neuroscience*, 35(44),
792 14896–14908. <http://doi.org/10.1523/JNEUROSCI.2270-15.2015>

793 Mesmoudi, S., Perlberg, V., Rudrauf, D., Messe, A., Pinsard, B., Hasboun, D., ... Burnod, Y.
794 (2013). Resting state networks' corticotomy: the dual intertwined rings architecture.
795 *PloS One*, 8(7), e67444. <http://doi.org/10.1371/journal.pone.0067444>

796 Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two
797 cortical pathways. *Trends in Neurosciences*, 6, 414–417. [http://doi.org/10.1016/0166-2236\(83\)90190-X](http://doi.org/10.1016/0166-2236(83)90190-X)

799 Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that
800 mediates familiarity memory. *Hippocampus*, 16(5), 504–20.
801 <http://doi.org/10.1002/hipo.20178>

802 Morgan, L. K., Macevoy, S. P., Aguirre, G. K., & Epstein, R. A. (2011). Distances between real-
803 world locations are represented in the human hippocampus. *The Journal of*
804 *Neuroscience*, 31(4), 1238–1245. <http://doi.org/10.1523/JNEUROSCI.4667-10.2011>

805 Muschamp, H. (2006, January 8). The Secret History of 2 Columbus Circle. *New York Times*.

806 Nakamura, K., Kawashima, R., Sato, N., Nakamura, a, Sugiura, M., Kato, T., ... Zilles, K. (2000).
807 Functional delineation of the human occipito-temporal areas related to face and scene
808 processing. A PET study. *Brain : A Journal of Neurology*, 123 (Pt 9, 1903–1912.
809 <http://doi.org/10.1093/brain/123.9.1903>

810 Nasr, S., Devaney, K. J., & Tootell, R. B. H. (2013). Spatial encoding and underlying circuitry
811 in scene-selective cortex. *NeuroImage*, 83, 892–900.
812 <http://doi.org/10.1016/j.neuroimage.2013.07.030>

813 Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking outside the box: rectilinear
814 shapes selectively activate scene-selective cortex. *The Journal of Neuroscience : The*
815 *Official Journal of the Society for Neuroscience*, 34(20), 6721–35.
816 <http://doi.org/10.1523/JNEUROSCI.4802-13.2014>

817 Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. H.
818 (2011). Scene-selective cortical regions in human and nonhuman primates. *The*
819 *Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 31(39),
820 13771–85. <http://doi.org/10.1523/JNEUROSCI.2792-11.2011>

821 O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates
822 corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, 12,
823 1013–1023. <http://doi.org/10.1162/08989290051137549>

824 Olsen, R. K., Moses, S. N., Riggs, L., & Ryan, J. D. (2012). The hippocampus supports multiple
825 cognitive processes through relational binding and comparison. *Frontiers in Human*
826 *Neuroscience*, 6(May), 146. <http://doi.org/10.3389/fnhum.2012.00146>

827 Park, S., & Chun, M. M. (2009). Different roles of the parahippocampal place area (PPA) and
828 retrosplenial cortex (RSC) in panoramic scene perception. *NeuroImage*, 47(4), 1747–
829 56. <http://doi.org/10.1016/j.neuroimage.2009.04.058>

830 Park, S., Konkle, T., & Oliva, A. (2014). Parametric Coding of the Size and Clutter of Natural
831 Scenes in the Human Brain. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–14.

832 <http://doi.org/10.1093/cercor/bht418>

833 Rafique, S. A., Solomon-Harris, L. M., & Steeves, J. K. E. (2015). TMS to object cortex affects
834 both object and scene remote networks while TMS to scene cortex only affects scene
835 networks. *Neuropsychologia*, 79, 86–96.
836 <http://doi.org/10.1016/j.neuropsychologia.2015.10.027>

837 Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. H. (2011). The
838 “parahippocampal place area” responds preferentially to high spatial frequencies in
839 humans and monkeys. *PLoS Biology*, 9(4), e1000608.
840 <http://doi.org/10.1371/journal.pbio.1000608>

841 Ranganath, C., & Ritchey, M. (2012). Two cortical systems for memory-guided behaviour.
842 *Nature Reviews. Neuroscience*, 13(10), 713–26. <http://doi.org/10.1038/nrn3338>

843 Ryan, J. D., & Cohen, N. J. (2004). Processing and short-term retention of relational
844 information in amnesia. *Neuropsychologia*, 42(4), 497–511.
845 <http://doi.org/10.1016/j.neuropsychologia.2003.08.011>

846 Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., & Smith, S. M.
847 (2014). Automatic denoising of functional MRI data: combining independent
848 component analysis and hierarchical fusion of classifiers. *NeuroImage*, 90, 449–68.
849 <http://doi.org/10.1016/j.neuroimage.2013.11.046>

850 Schiller, D., Eichenbaum, H., Buffalo, E. A., Davachi, L., Foster, D. J., Leutgeb, S., & Ranganath,
851 C. (2015). Memory and Space: Towards an Understanding of the Cognitive Map.
852 *Journal of Neuroscience*, 35(41), 13904–13911.
853 <http://doi.org/10.1523/JNEUROSCI.2618-15.2015>

854 Schindler, A., & Bartels, A. (2013). Parietal cortex codes for egocentric space beyond the
855 field of view. *Current Biology : CB*, 23(2), 177–82.
856 <http://doi.org/10.1016/j.cub.2012.11.060>

857 Silson, E. H., Chan, a. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A Retinotopic
858 Basis for the Division of High-Level Scene Processing between Lateral and Ventral
859 Human Occipitotemporal Cortex. *Journal of Neuroscience*, 35(34), 11921–11935.
860 <http://doi.org/10.1523/JNEUROSCI.0137-15.2015>

861 Silson, E. H., Groen, I. I. A., Kravitz, D. J., & Baker, C. I. (2016). Evaluating the correspondence
862 between face-, scene-, and object-selectivity and retinotopic organization within
863 lateral occipitotemporal cortex. *Journal of Vision*, 16(6), 1–21.
864 <http://doi.org/10.1167/16.6.14>

865 Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L., & Beckmann, C. F. (2014). Group-PCA
866 for very large fMRI datasets. *NeuroImage*.
867 <http://doi.org/10.1016/j.neuroimage.2014.07.051>

868 Spreng, R., Mar, R., & Kim, A. (2009). The common neural basis of autobiographical memory,
869 prospection, navigation, theory of mind, and the default mode: a quantitative meta-
870 analysis. *Journal of Cognitive Neuroscience*, 7, 489–510. Retrieved from
871 <http://www.mitpressjournals.org/doi/abs/10.1162/jocn.2008.21029>

872 Szpunar, K. K., Chan, J. C. K., & McDermott, K. B. (2009). Contextual processing in episodic

873 future thought. *Cerebral Cortex* (New York, N.Y. : 1991), 19(7), 1539–48.
874 <http://doi.org/10.1093/cercor/bhn191>

875 Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., ...
876 Fernández, G. (2006). Declarative memory consolidation in humans: a prospective
877 functional magnetic resonance imaging study. *Proceedings of the National Academy of*
878 *Sciences of the United States of America*, 103(3), 756–61.
879 <http://doi.org/10.1073/pnas.0507774103>

880 van Assche, M., Kebets, V., Vuilleumier, P., & Assal, F. (2014). Functional Dissociations
881 Within Posterior Parietal Cortex During Scene Integration and Viewpoint Changes.
882 *Cereb Cortex*, bhu215–. <http://doi.org/10.1093/cercor/bhu215>

883 Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013).
884 The WU-Minn Human Connectome Project: an overview. *NeuroImage*, 80, 62–79.
885 <http://doi.org/10.1016/j.neuroimage.2013.05.041>

886 Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do?
887 *Nature Reviews Neuroscience*, 10(11), 792–802. <http://doi.org/10.1038/nrn2733>

888 Vass, L. K., & Epstein, R. a. (2013). Abstract representations of location and facing direction
889 in the human brain. *The Journal of Neuroscience : The Official Journal of the Society for*
890 *Neuroscience*, 33(14), 6133–42. <http://doi.org/10.1523/JNEUROSCI.3873-12.2013>

891 Vilberg, K. L., & Rugg, M. D. (2012). The neural correlates of recollection: transient versus
892 sustained fMRI effects. *The Journal of Neuroscience : The Official Journal of the Society*
893 *for Neuroscience*, 32(45), 15679–87. [http://doi.org/10.1523/JNEUROSCI.3065-](http://doi.org/10.1523/JNEUROSCI.3065-12.2012)
894 12.2012

895 Wang, L., Mruczek, R. E. B., Arcaro, M. J., & Kastner, S. (2014). Probabilistic Maps of Visual
896 Topography in Human Cortex. *Cerebral Cortex*, bhu277–.
897 <http://doi.org/10.1093/cercor/bhu277>

898 Ward, E. J., MacEvoy, S. P., & Epstein, R. A. (2010). Eye-centered encoding of visual space in
899 scene-selective regions. *Journal of Vision*, 10, 6. <http://doi.org/10.1167/10.14.6>

900 Warren, D. E., Duff, M. C., Jensen, U., Tranel, D., & Cohen, N. J. (2012). Hiding in plain view:
901 Lesions of the medial temporal lobe impair online representation. *Hippocampus*, 22(7),
902 1577–1588. <http://doi.org/10.1002/hipo.21000>

903 Watson, D. M., Hymers, M., Hartley, T., & Andrews, T. J. (2016). Patterns of neural response
904 in scene-selective regions of the human brain are affected by low-level manipulations
905 of spatial frequency. *NeuroImage*, 124, 107–117.
906 <http://doi.org/10.1016/j.neuroimage.2015.08.058>

907 Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale
908 automated synthesis of human functional neuroimaging data. *Nature Methods*, 8(8),
909 665–670. <http://doi.org/10.1038/nmeth.1635>

910 Yates, F. A. (1966). *The Art of Memory*. Chicago, IL: University of Chicago Press.

911 Yeo, B. T. T., Krienen, F. M., Eickhoff, S. B., Yaakub, S. N., Fox, P. T., Buckner, R. L., ... Chee, M.
912 W. L. (2014). Functional Specialization and Flexibility in Human Association Cortex.
913 *Cerebral Cortex*, bhu217–. <http://doi.org/10.1093/cercor/bhu217>

914 Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ...
915 Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by
916 intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3), 1125–65.
917 <http://doi.org/10.1152/jn.00338.2011>

918 Zeidman, P., & Maguire, E. A. (2016). Anterior hippocampus: the anatomy of perception,
919 imagination and episodic memory. *Nature Reviews Neuroscience*, 17(3), 173–182.
920 <http://doi.org/10.1038/nrn.2015.24>

921 Zeidman, P., Mullally, S. L., & Maguire, E. A. (2015). Constructing, Perceiving, and
922 Maintaining Scenes: Hippocampal Activity and Connectivity. *Cerebral Cortex (New York,*
923 *N.Y. : 1991)*, 25(10), 3836–55. <http://doi.org/10.1093/cercor/bhu266>

924

925

926

927

928

929

930

931

932

Supplementary Figures

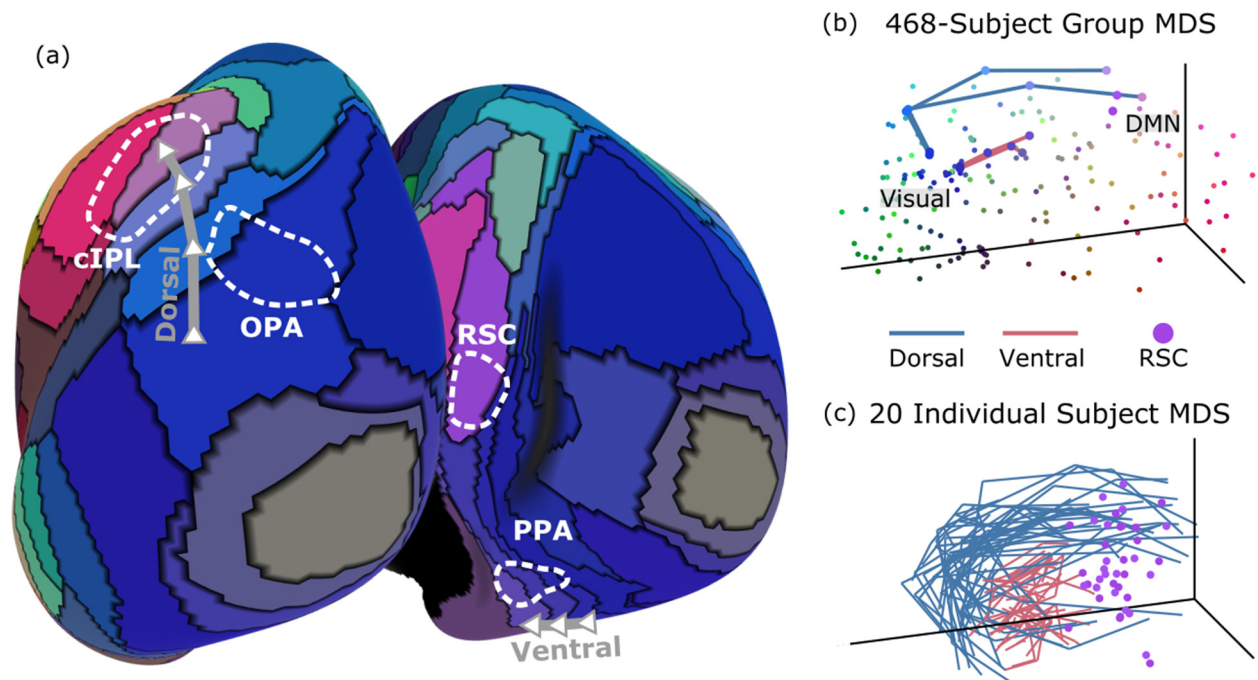
933

Two distinct scene processing networks connecting vision and memory

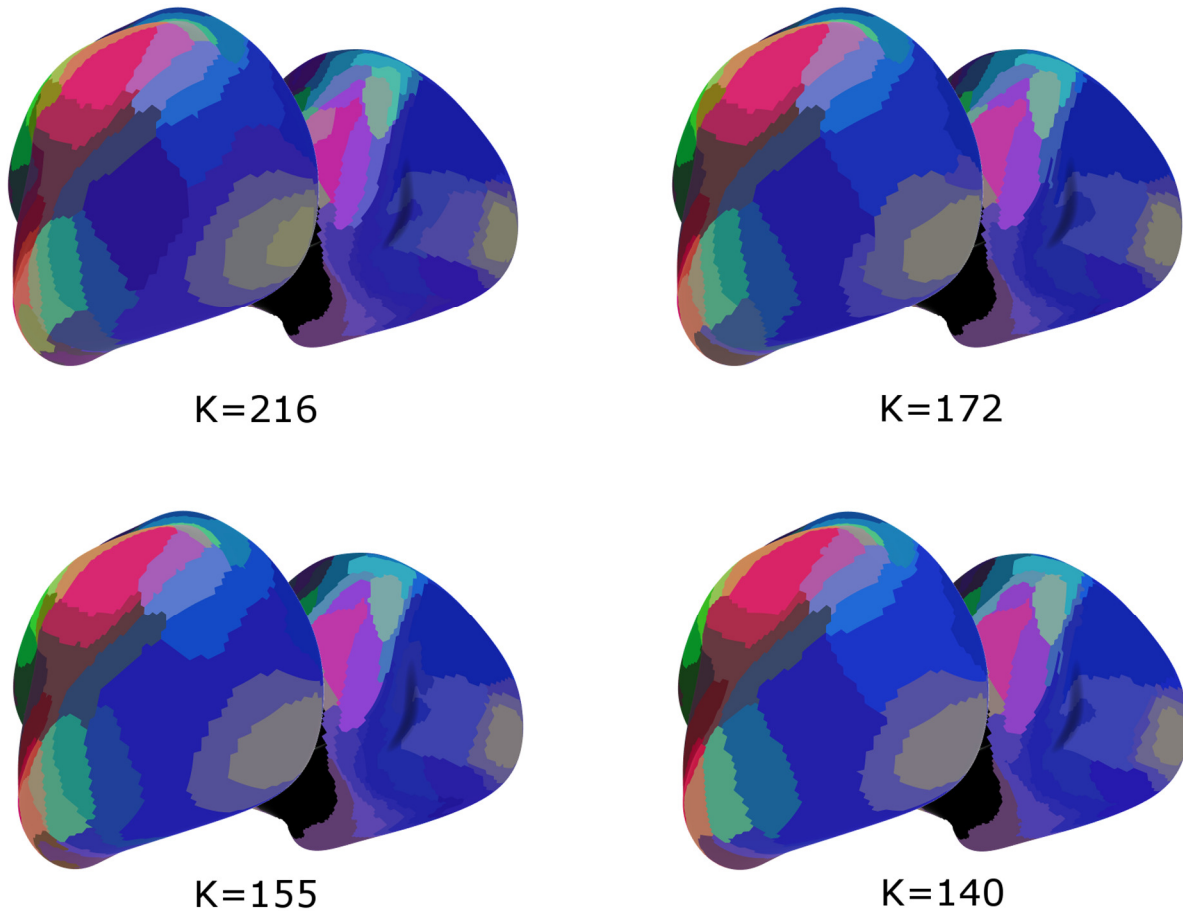
934

Christopher Baldassano, Andre Esteva, Diane M. Beck, Li Fei-Fei

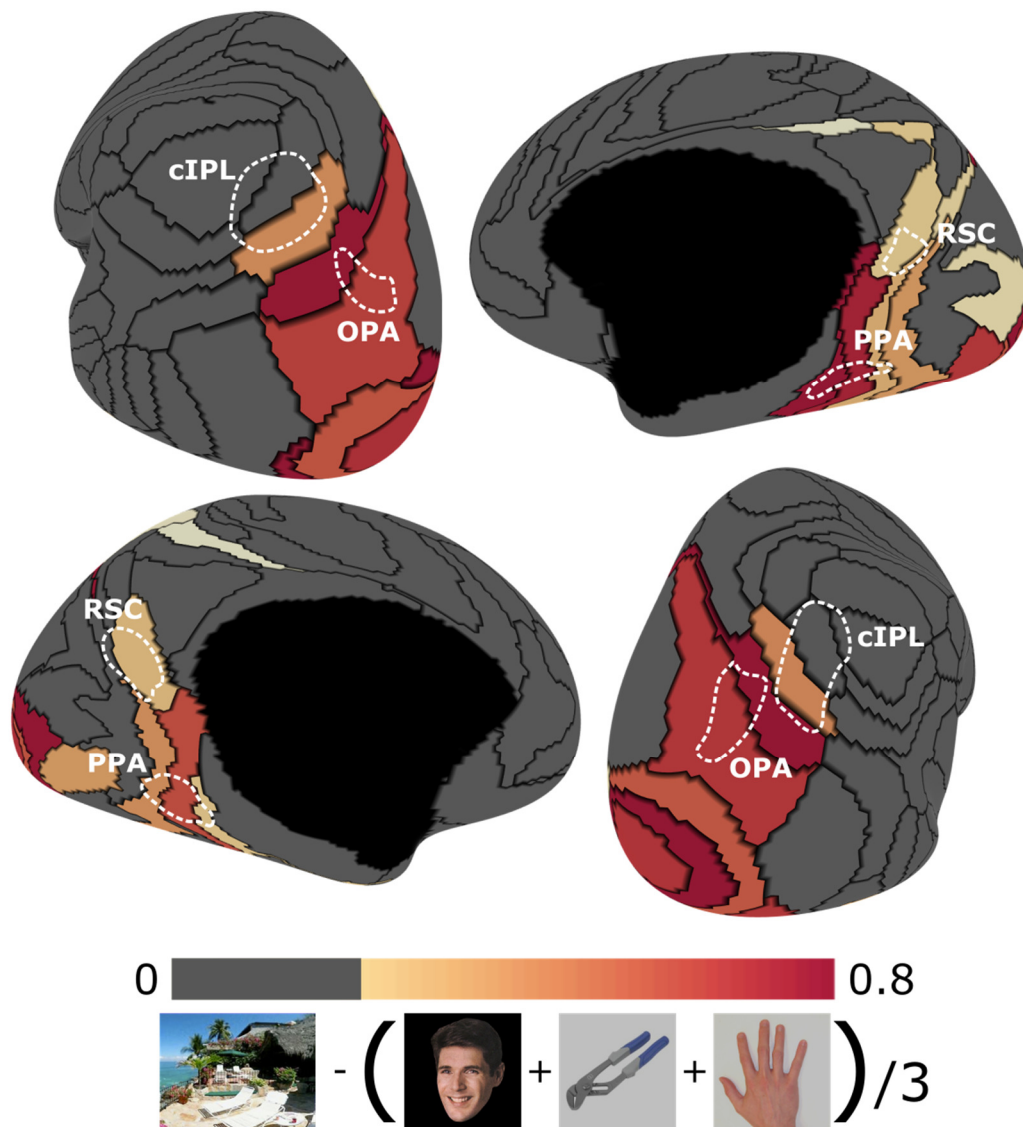
935



Supplementary Figure 1: Multidimensional scaling of parcel connectivity matrix. We can use classical multidimensional scaling (MDS) to embed parcels into a three-dimensional space RGB space. Distances in this space approximate the functional connectivity strength between parcels, such that strongly-connected parcels are close together in the embedding space and have similar colors. (a) There is a notable border in both dorsal (OPA to cIPL) and ventral (posterior to anterior PPA) cortex, where moving a short spatial distance along the cortical surface produces significant changes in functional connectivity properties. (b-c) These same paths are visualized in the MDS embedding space, both for the group and individual subjects. The most posterior regions show strong connectivity to other parcels in visual cortex, while the most anterior regions are instead more related to default mode regions.



Supplementary Figure 2: Varying the number of cortical parcels. The number of parcels chosen by our clustering algorithm (Baldassano et al., 2015) depends on a hyperparameter σ_0^2 which controls the level of desired homogeneity within each cluster. Rather than using $\sigma_0^2 = 3000$ as in the main text, we can reduce σ_0^2 to 2000 (yielding 216 parcels) or increase σ_0^2 to 4000 (yielding 155 parcels) or 5000 (yielding 140 parcels). Performing the same MDS embedding as in Supplementary Figure 1 on each of these parcellations yields results that are qualitatively very similar, with the same transition between posterior and anterior scene regions.



Supplementary Figure 3: Testing for scene-sensitivity in HCP subjects. In the main text we use functional ROIs (OPA, PPA, RSC) defined using a standard localizer in an independent group of subjects. To validate that the parcels overlapping these ROIs are scene selective, we measured their mean response to scenes minus their response to other categories during the HCP working memory experiment. The results validate that these parcels are in fact scene selective in the HCP subjects. Note that early visual regions also show scene selectivity, likely because peripheral contrast was not well controlled in these stimuli. FDR<0.05.