



Research report

The occipital place area represents first-person perspective motion information through scenes

Frederik S. Kamps, Vishal Lall and Daniel D. Dilks*

Department of Psychology, Emory University, Atlanta, GA, United States

ARTICLE INFO

Article history:

Received 29 February 2016

Reviewed 8 April 2016

Revised 4 June 2016

Accepted 28 June 2016

Action editor Jason Barton

Published online 15 July 2016

Keywords:

fMRI

OPA

Parahippocampal place area (PPA)

Retrosplenial complex (RSC)

Scene perception

ABSTRACT

Neuroimaging studies have identified multiple scene-selective regions in human cortex, but the precise role each region plays in scene processing is not yet clear. It was recently hypothesized that two regions, the occipital place area (OPA) and the retrosplenial complex (RSC), play a direct role in navigation, while a third region, the parahippocampal place area (PPA), does not. Some evidence suggests a further division of labor even among regions involved in navigation: While RSC is thought to support navigation through the broader environment, OPA may be involved in navigation through the immediately visible environment, although this role for OPA has never been tested. Here we predict that OPA represents first-person perspective motion information through scenes, a critical cue for such “visually-guided navigation”, consistent with the hypothesized role for OPA. Response magnitudes were measured in OPA (as well as RSC and PPA) to i) video clips of first-person perspective motion through scenes (“Dynamic Scenes”), and ii) static images taken from these same movies, rearranged such that first-person perspective motion could not be inferred (“Static Scenes”). As predicted, OPA responded significantly more to the Dynamic than Static Scenes, relative to both RSC and PPA. The selective response in OPA to Dynamic Scenes was not due to domain-general motion sensitivity or to low-level information inherited from early visual regions. Taken together, these findings suggest the novel hypothesis that OPA may support visually-guided navigation, insofar as first-person perspective motion information is useful for such navigation, while RSC and PPA support other aspects of navigation and scene recognition.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Recognizing the visual environment, or “scene”, and using that information to navigate is critical in our everyday lives. Given the ecological importance of scene recognition and navigation, it is perhaps not surprising then that we have dedicated neural machinery for scene processing: the

occipital place area (OPA) (Dilks, Julian, Paunov, & Kanwisher, 2013), the retrosplenial complex (RSC) (Maguire, 2001), and the parahippocampal place area (PPA) (Epstein & Kanwisher, 1998). Beyond establishing the general involvement of these regions in scene processing, however, a fundamental and yet unanswered question remains: What is the precise function of each region in scene processing, and how do these regions

* Corresponding author. Department of Psychology, Emory University, 36 Eagle Row, Atlanta, GA 30322, United States.

E-mail address: dilks@emory.edu (D.D. Dilks).

<http://dx.doi.org/10.1016/j.cortex.2016.06.022>

0010-9452/© 2016 Elsevier Ltd. All rights reserved.

support our crucial ability to recognize and navigate our environment?

Growing evidence indicates that OPA, RSC, and PPA play distinct roles in scene processing. For example, OPA and RSC are sensitive to two essential kinds of information for navigation: sense (i.e., left vs right) and egocentric distance (i.e., near vs far from me) information (Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011; Persichetti & Dilks, 2016). By contrast, PPA is not sensitive to either sense or egocentric distance information. The discovery of such differential sensitivity to navigationally-relevant information across scene-selective cortex has led to the hypothesis that OPA and RSC directly support navigation, while PPA does not (Dilks et al., 2011; Persichetti & Dilks, 2016). Further studies suggest that there may be a division of labor even among those regions involved in navigation, although this hypothesis has never been tested directly. In particular, RSC is thought to represent information about both the immediately visible scene and the broader spatial environment related to that scene (Epstein, 2008; Maguire, 2001), in order to support navigational processes such as landmark-based navigation (Auger, Mullally, & Maguire, 2012; Epstein & Vass, 2015), location and heading retrieval (Epstein, Parker, & Feiler, 2007; Marchette, Vass, Ryan, & Epstein, 2014; Vass & Epstein, 2013), and the formation of environmental survey knowledge (Auger, Zeidman, & Maguire, 2015; Wolbers & Buchel, 2005). By contrast, although little is known about OPA, it was recently proposed that OPA supports visually-guided navigation and obstacle avoidance in the immediately visible scene itself (Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2016).

One critical source of information for such visually-guided navigation is the first-person perspective motion information experienced during locomotion (Gibson, 1950). Thus, here we investigated how OPA represents first-person perspective motion information through scenes. Responses in the OPA (as well as RSC and PPA) were measured using fMRI while participants viewed i) 3-sec video clips of first-person perspective motion through a scene (“Dynamic Scenes”), mimicking the actual visual experience of locomotion, and ii) 3, 1-sec still images taken from these same video clips, rearranged such that first-person perspective motion could not be inferred (“Static Scenes”). We predicted that OPA would respond more to the Dynamic Scenes than the Static Scenes, relative to both RSC and PPA, consistent with the hypothesis that OPA supports visually-guided navigation, since first-person perspective motion information is undoubtedly useful for such navigation, while RSC and PPA are involved in other aspects of navigation and scene recognition.

2. Method

2.1. Participants

Sixteen healthy university students (ages 20–35; mean age = 25.9; SD = 4.3; 7 females) were recruited for this experiment. All participants gave informed consent. All had normal or corrected to normal vision; were right handed (one reported being ambidextrous), as measured by the Edinburgh

Handedness Inventory (mean = .74; SD = .31, where +1 is considered a “pure right hander” and –1 is a “pure left hander”) (Oldfield, 1971); and had no history of neurological or psychiatric conditions. All procedures were approved by the Emory University Institutional Review Board.

2.2. Design

For our primary analysis, we used a region of interest (ROI) approach in which we used one set of runs (Localizer runs, described below) to define the three scene-selective regions (as described previously; Epstein & Kanwisher, 1998), and then used a second, independent set of runs (Experimental runs, described below) to investigate the responses of these regions to Dynamic Scenes and Static Scenes, as well as two control conditions: Dynamic Faces and Static Faces. As a secondary analysis, we performed a group-level analysis exploring responses to the Experimental runs across the entire slice prescription (for a detailed description of this analysis see [Data analysis](#) section).

For the Localizer runs, we used a standard method used in many previous studies to identify ROIs (Epstein & Kanwisher, 1998; Kamps et al., 2016; Kanwisher & Dilks, 2013; Park, Brady, Greene, & Oliva, 2011; Walther, Caddigan, Fei-Fei, & Beck, 2009). Specifically, a blocked design was used in which participants viewed static images of scenes, faces, objects, and scrambled objects. We defined scene-selective ROIs using static images, rather than dynamic movies for two reasons. First, using the standard method of defining scene-selective ROIs with static images allowed us to ensure that we were investigating the same ROIs investigated in many previous studies of cortical scene processing, facilitating the comparison of our results with previous work. Second, the use of dynamic movies to define scene-selective ROIs could potentially bias responses in regions that are selective to dynamic information in scenes, inflating the size of the “dynamic” effect. The same argument, of course, could be used for the static images (i.e., potentially biasing responses in regions that are selective to static information in scenes, again inflating the size of the “dynamic” effect). However, note that in either case, the choice of dynamic or static stimuli to define scene-selective ROIs would result in a main effect of motion (i.e., a greater response to Dynamic Scenes than Static Scenes in all three scene-selective regions, or vice versa), not an interaction of motion by ROI (i.e., a greater response in OPA to Dynamic Scenes than Static Scenes, relative to PPA and RSC), as predicted. Each participant completed 3 runs, with the exception of two participants who only completed 2 runs due to time constraints. Each run was 336 sec long and consisted of 4 blocks per stimulus category. For each run, the order of the first eight blocks was pseudorandomized (e.g., faces, faces, objects, scenes, objects, scrambled objects, scenes, scrambled objects), and the order of the remaining eight blocks was the palindrome of the first eight (e.g., scrambled objects, scenes, scrambled objects, objects, scenes, objects, faces, faces). Each block contained 20 images from the same category for a total of 16 sec blocks. Each image was presented for 300 msec, followed by a 500 msec interstimulus interval (ISI), and subtended 8 × 8 degrees of visual angle. We also included five

16 sec fixation blocks: one at the beginning, and one every four blocks thereafter. Participants performed a one-back task, responding every time the same image was presented twice in a row.

For the Experimental runs, the Dynamic Scene stimuli consisted of 60, 3-sec video clips depicting first-person perspective motion, as would be experienced during locomotion through a scene (Fig. 1). Specifically, the video clips were filmed by walking at a typical pace through 8 different places (e.g., a parking garage, a hallway, etc.) with the camera (a Sony HDR XC260V HandyCam with a field of view of 90.3×58.9 degrees) held at eye level. The video clips subtended 23×15.33 degrees of visual angle. The Static Scene stimuli were created by taking stills from each Dynamic Scene video clip at 1-, 2- and 3-sec time points, resulting in 180 images. These still images were presented in groups of three images taken from the same place, and each image was presented for 1 sec with no ISI, thus equating the presentation time of the static images with the duration of the movie clips from which they were made. Importantly, the still images were presented in a random sequence such that first-person perspective motion could not be inferred. Like the video clips, the still images subtended 23×15.33 degrees of visual angle. Next, to test the specificity of any observed differences between Dynamic Scene and Static Scene conditions, we also included Dynamic

Face and Static Face conditions (Fig. 1). The Dynamic Face stimuli were the same as those used in Pitcher, Dilks, Saxe, Triantafyllou, and Kanwisher (2011), and depicted only the faces of 7 children against a black background as they smiled, laughed, and looked around while interacting with off-screen toys or adults. The Static Face stimuli were created and presented using the exact same procedure and parameters described for the Static Scene condition above.

Participants completed 3 “dynamic” runs (i.e., blocks of Dynamic Scene and Dynamic Face conditions) and 3 “static” runs (i.e., blocks of Static Scene and Static Face conditions). The dynamic and static runs were interspersed within participant, and the order of runs was counterbalanced across participants. Separate runs of dynamic and static stimuli were used for two reasons. First, the exact same design had been used previously to investigate dynamic face information representation across face-selective regions (Pitcher et al., 2011), which allowed us to compare our findings in the face conditions to those of Pitcher and colleagues, validating our paradigm. Second, we wanted to prevent the possibility of “contamination” of motion information from the Dynamic Scenes to the Static Scenes, as could be the case if they were presented in the same run, reducing any differences we might observe between the two conditions. Each run was 315 sec long and consisted of 8 blocks of each

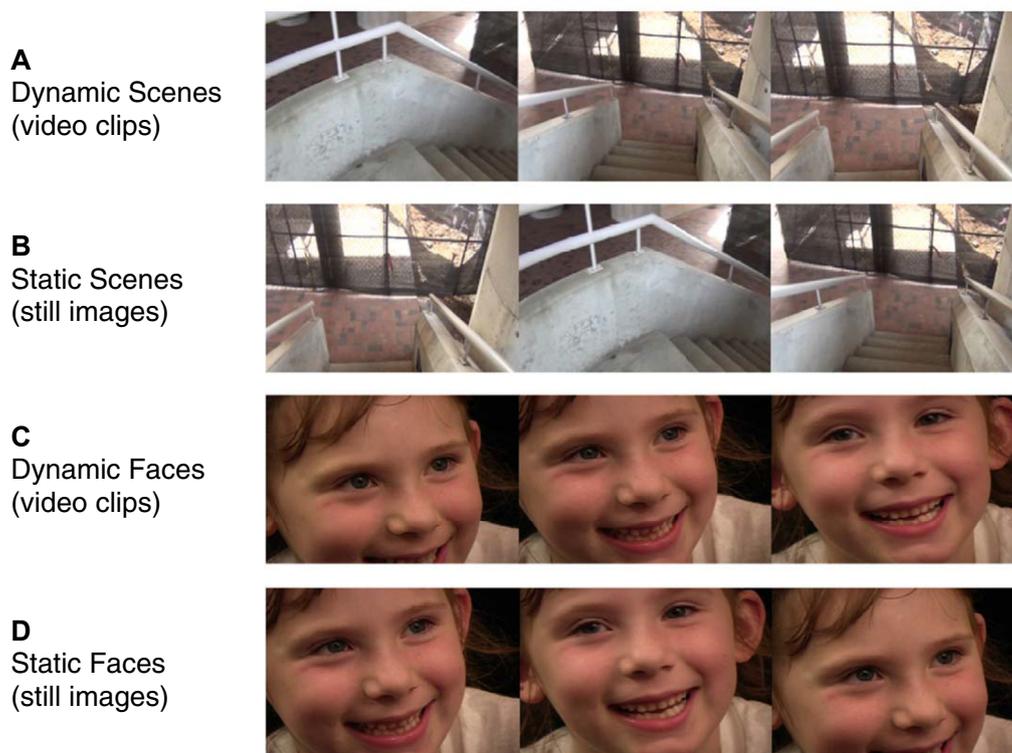


Fig. 1 – Example stimuli used in the experimental scans. The conditions included A) Dynamic Scenes, which consisted of 3-sec video clips of first-person perspective motion through a scene; B) Static Scenes, which consisted of 3 1-sec stills taken from the Dynamic Scenes condition and presented in a random order, such that first-person perspective motion could not be inferred; C) Dynamic Faces, which consisted of 3-sec video clips of only the faces of children against a black background as they interacted with off-screen adults or toys; and D) Static Faces, which consisted of 3 1-sec stills taken from the Dynamic Faces and presented in a random order.

condition. For each run, the order of the first eight blocks was pseudorandomized (e.g., faces, scenes, scenes, faces, scenes, scenes, faces, faces), and the order of the remaining eight blocks was the palindrome of the first eight (e.g., faces, faces, scenes, scenes, faces, scenes, scenes, faces). In the dynamic runs, each block consisted of 5, 3-sec movies of Dynamic Scenes or Dynamic Faces, totaling 15 sec per block. In the static runs, each block consisted of 5 sets of 3-sec images of Static Scenes or Static Faces, totaling 15 sec per block. We also included five 15 sec fixation blocks: one at the beginning, and one every four blocks thereafter. During the Experimental runs, participants were instructed to passively view the stimuli.

2.3. fMRI scanning

All scanning was performed on a 3T Siemens Trio scanner in the Facility for Education and Research in Neuroscience at Emory University. Functional images were acquired using a 32-channel head matrix coil and a gradient-echo single-shot echoplanar imaging sequence (28 slices, TR = 2 sec, TE = 30 msec, voxel size = $1.5 \times 1.5 \times 2.5$ mm, and a .25 interslice gap). For all scans, slices were oriented approximately between perpendicular and parallel to the calcarine sulcus, covering all of the occipital and temporal lobes, as well as the lower portion of the parietal lobe. Whole-brain, high-resolution anatomical images were also acquired for each participant for purposes of registration and anatomical localization (see [Data analysis](#)).

2.4. Data analysis

fMRI data analysis was conducted using the FSL software (FMRIB's Software Library; www.fmrib.ox.ac.uk/fsl) (Smith et al., 2004) and the FreeSurfer Functional Analysis Stream (FS-FAST; <http://surfer.nmr.mgh.harvard.edu/>). ROI analysis was conducted using the FS-FAST ROI toolbox. Before statistical analysis, images were motion corrected (Cox & Jesmanowicz, 1999). Data were then detrended and fit using a double gamma function. All data were spatially smoothed with a 5-mm kernel. After preprocessing, scene-selective regions OPA, RSC, and PPA were bilaterally defined in each participant (using data from the independent localizer scans) as those regions that responded more strongly to scenes than objects ($p < 10^{-4}$, uncorrected), as described previously (Epstein & Kanwisher, 1998), and further constrained using a published atlas of “parcels” that identify the anatomical regions within which most subjects show activation for the contrast of scenes minus objects (Julian, Fedorenko, Webster, & Kanwisher, 2012). We also defined two control regions. First, we functionally defined foveal cortex (FC) using the contrast of scrambled objects > objects, as previously described (Kamps et al., 2016), using data from the localizer scans. Second, using an independent dataset from another experiment that included the same four experimental conditions used here (Dynamic Scenes, Static Scenes, Dynamic Faces, Static Faces), but that tested different hypotheses, we were able to functionally define middle temporal area (MT) (Tootell et al., 1995) as the region responding more to both Dynamic Scenes and Dynamic Faces than to

both Static Scenes and Static Faces in 8 of our 16 participants. The number of participants exhibiting each region in each hemisphere was as follows: rOPA: 16/16; rRSC: 16/16; rPPA: 16/16; rFC: 14/16; rMT: 8/8; lOPA: 16/16; lRSC: 16/16; lPPA: 16/16; lFC: 15/16; lMT: 8/8. Within each ROI, we then calculated the magnitude of response (percent signal change) to the Dynamic Scenes and Static Scenes, using the data from the Experimental runs. A 2 (hemisphere: Left, Right) \times 2 (condition: Dynamic Scenes, Static Scenes) repeated-measures ANOVA was conducted for each scene ROI. We found no significant hemisphere by condition interaction in OPA [$F_{(1,15)} = .04$, $p = .85$], RSC [$F_{(1,15)} = .38$, $p = .55$], PPA [$F_{(1,15)} = 2.28$, $p = .15$], FC [$F_{(1,13)} = .44$, $p = .52$], or MT [$F_{(1,7)} = 1.72$, $p = .23$]. Thus, both hemispheres for each ROI were collapsed for further analyses. After collapsing across hemispheres, the number of participants exhibiting each ROI in at least one hemisphere was as follows: OPA: 16/16; RSC: 16/16; PPA: 16/16; FC: 15/16; MT: 8/8 ([Supplemental Fig. 1](#)).

Finally, in addition to the ROI analysis described above, we also performed a group-level analysis to explore responses to the experimental conditions across the entire slice prescription. This analysis was conducted using the same parameters as were used in the ROI analysis, with the exception that the experimental data were registered to standard stereotaxic (MNI) space. We performed two contrasts: Dynamic Scenes versus Static Scenes and Dynamic Faces versus Static Faces. For each contrast, we conducted a nonparametric one-sample t-test using the FSL randomize program (Winkler, Ridgway, Webster, Smith, & Nichols, 2014) with default variance smoothing of 5 mm, which tests the t value at each voxel against a null distribution generated from 5000 random permutations of group membership. The resultant statistical maps were then corrected for multiple comparisons ($p < .01$, FWE) using threshold-free cluster enhancement, a method that retains the power of cluster-wise inference without the dependence on an arbitrary cluster-forming threshold (Smith & Nichols, 2009).

3. Results

3.1. OPA represents first-person perspective motion information through scenes, while RSC and PPA do not

Here we predicted that OPA represents first-person perspective motion information through scenes, consistent with the hypothesis that OPA plays a role in visually-guided navigation. Confirming our prediction, a 3 (ROI: OPA, RSC, PPA) \times 2 (condition: Dynamic Scenes, Static Scenes) repeated-measures ANOVA revealed a significant interaction [$F_{(2,30)} = 8.09$, $p = .002$, $\eta_p^2 = .35$], with OPA responding significantly more to the Dynamic Scene condition than the Static Scene condition, relative to both RSC and PPA (interaction contrasts, both p 's < .05). By contrast, RSC and PPA responded similarly to the Dynamic Scene and Static Scene conditions (interaction contrast, $p = .44$) ([Fig. 2](#)). Importantly, the finding of a significant interaction across these regions rules out the possibility that differences in attention between the conditions drove these effects, since such a difference would cause a main effect of condition, not an interaction of region by

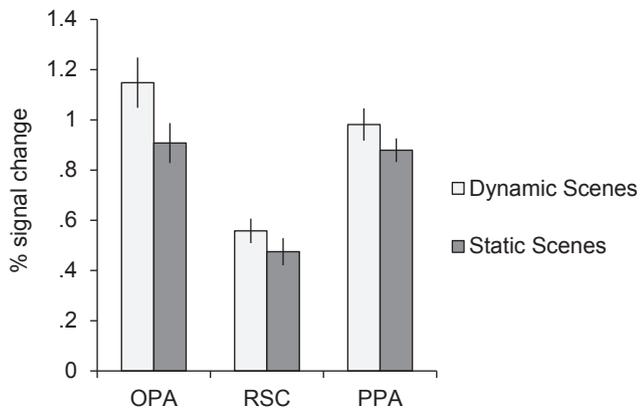


Fig. 2 – Average percent signal change in OPA, RSC, and PPA to the Dynamic Scenes condition, depicting first-person perspective motion information through scenes, and the Static Scenes condition, in which first-person perspective motion was disrupted. OPA responded more to the Dynamic Scenes than Static Scenes, relative to both RSC and PPA [$F_{(2,30)} = 8.091, p = .002$], demonstrating that OPA selectively represents first-person perspective motion information, a critical cue for visually-guided navigation.

condition. Thus, taken together, these findings demonstrate that OPA represents first-person perspective motion information through scenes—a critical cue for visually-guided navigation—while PPA and RSC do not, supporting the hypothesized role of OPA in visually-guided navigation.

But might OPA be responding to motion information more generally, rather than motion information in scenes, in particular? To test this possibility, we compared the

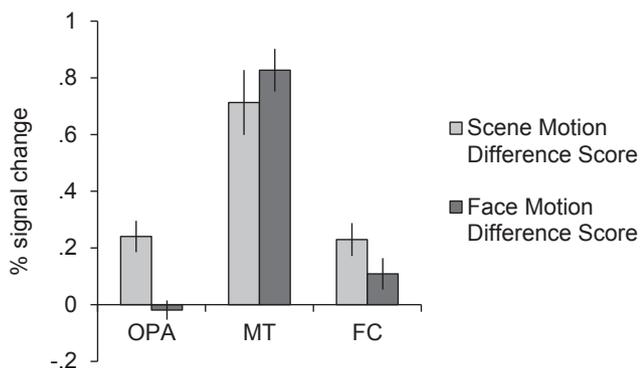


Fig. 3 – Difference scores (percent signal change) for Dynamic Scenes minus Static Scenes (“Scene Motion Difference Score”) and Dynamic Faces minus Static Faces (“Face Motion Difference Score”) in OPA, MT (a motion-selective region), and FC (a low-level visual region). OPA responded significantly more to scene motion than face motion, relative to both MT [$F_{(1,7)} = 95.41, p < .001$] and FC [$F_{(1,14)} = 9.96, p < .01$], indicating that the response to scene-selective motion in OPA does not reflect differences in the amount of motion information in the scene stimuli compared to the face stimuli, or information inherited from low-level visual regions.

difference in response in OPA to the Dynamic Scenes (with motion) and Static Scenes (without motion) (“Scene difference score”), with the difference in response to Dynamic Faces (with motion) and Static Faces (without motion) (“Face difference score”) (Fig. 4). A paired samples t-test revealed that the Scene difference score was significantly greater than the Face difference score in OPA [$t_{(15)} = 6.96, p < .001, d = 2.07$], indicating that OPA does not represent motion information in general, but rather selectively responds to first-person perspective motion in scenes. Of course, it is possible that OPA may represent other kinds of motion information in scenes beyond the first-person perspective motion information tested here, a question we explore in detail in Discussion.

However, given that we did not precisely match the amount of motion information between scene and face stimuli, might OPA be responding more to Dynamic Scenes than Dynamic Faces because Dynamic Scenes have more motion information than Dynamic Faces, rather than responding specifically to scene-selective motion? To test this possibility, we compared the Scene difference score and the Face difference score in OPA with those in MT—a domain-general motion-selective region. A 2 (region: OPA, MT) \times 2 (condition: Scene difference score, Face difference score) repeated-measures ANOVA revealed a significant interaction [$F_{(1,7)} = 95.41, p < .001, \eta_p^2 = .93$], with OPA responding significantly more to motion information in scenes than faces, and MT showing the opposite pattern (Bonferroni corrected post-hoc comparisons, both p 's $< .05$). The greater response to face motion information than scene motion information in MT suggests that in fact there was more motion information present in the Dynamic Faces than Dynamic Scenes, ruling out the possibility that differences in the amount of motion information in the scene stimuli compared to the face stimuli can explain the selective response in OPA to motion information in scenes.

3.2. Responses in OPA do not reflect information inherited from low-level visual regions

While the above findings suggest that OPA represents first-person perspective motion information in scenes—unlike RSC and PPA—and further that OPA does not represent motion information in general, might it still be the case that the response of OPA simply reflects visual information inherited from low-level visual regions? To rule out this possibility, we compared the Scene difference score (Dynamic Scenes minus Static Scenes) and Face difference score (Dynamic Faces minus Static Faces) in OPA with those in FC (i.e., a low-level visual region) (Fig. 3). A 2 (ROI: OPA, FC) \times 2 (condition: Scene difference score, Face difference score) repeated-measures ANOVA revealed a significant interaction [$F_{(1,14)} = 9.96, p < .01, \eta_p^2 = .42$], with the Scene difference score significantly greater than the Face difference score in OPA, relative to FC. This finding reveals that OPA is not simply inheriting information from a low-level visual region, but rather is responding selectively to first-person perspective motion information through scenes.

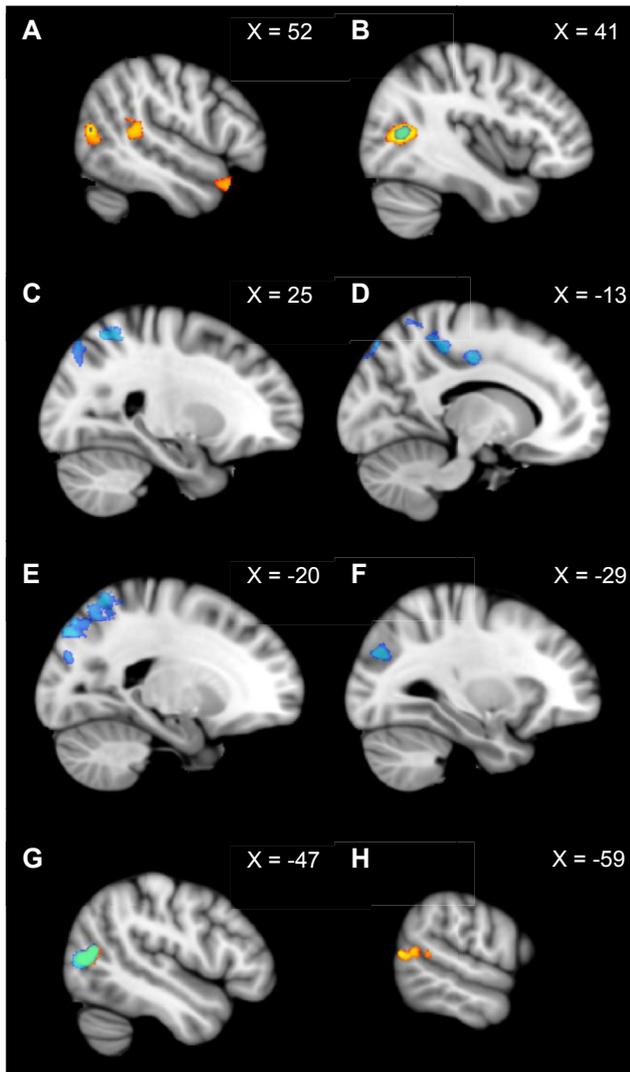


Fig. 4 – Group analysis exploring representation of first-person perspective motion information beyond OPA. The contrast of “Dynamic Scenes > Static Scenes” is shown in blue ($p < .01$, FWE corrected), while the contrast of “Dynamic Faces > Static Faces” is shown in yellow ($p < .01$, FWE corrected). The right hemisphere is depicted in panels A–C, while the left hemisphere is depicted in panels D–H. X coordinates in MNI space are provided for each slice. A network of regions including lateral superior occipital cortex (corresponding to OPA; see F), superior parietal lobe, and precentral gyrus (see C–E) responded significantly more to “Dynamic Scenes > Static Scenes” (blue), but similarly to “Dynamic Faces vs. Static Faces” (yellow). One bilateral region in lateral occipital cortex (corresponding to motion-selective MT) showed overlapping activation across both contrasts (see B, G). Finally, regions in bilateral posterior superior temporal sulcus and anterior temporal pole responded more to “Dynamic Faces > Static Faces,” but similarly to “Dynamic Scenes vs. Static Scenes” (see A and H).

3.3. Do regions beyond OPA represent first-person perspective motion through scenes?

To explore whether regions beyond OPA might also be involved in representing first-person perspective motion through scenes, we performed a group-level analysis examining responses across the entire slice prescription (Fig. 4, Table 1). If a region represents first-person perspective motion through scenes, then it should respond significantly more to the Dynamic Scene condition than the Static Scene condition ($p = .01$, FWE corrected). We found several regions showing this pattern of results: i) the left lateral superior occipital lobe (which overlapped with OPA as defined in a comparable group-level contrast of scenes vs objects using data from the Localizer scans), consistent with the above ROI analysis; ii) a contiguous swath of cortex in both hemispheres extending from the lateral superior occipital lobe into the parietal lobe, including the intraparietal sulcus and superior parietal lobule,

Table 1 – Summary of peak activations within clusters for the contrasts of “Dynamic Scenes > Static Scenes” and “Dynamic Faces > Static Faces.”

Region	MNI coordinate			t-Stat	Cluster size (voxels)
	X	Y	Z		
<i>Dynamic Scenes > Static Scenes</i>					
R lateral superior occipital cortex (extending into parietal lobe)	21	-77	48	4.88	1403
L lateral superior occipital cortex (contiguous with OPA, and extending into parietal lobe)	-23	-84	41	5.48	5810
R precentral gyrus	10	-14	44	5.66	4643
L precentral gyrus	-14	-36	47	5.08	867
L precentral gyrus	-12	-15	41	6.01	429
L precentral gyrus	-14	-27	42	3.86	8
R lateral middle occipital cortex (contiguous with MT) ^a	41	-64	8	6.16	494
R lateral middle occipital cortex (contiguous with MT) ^a	51	-68	11	4.86	24
L lateral middle occipital cortex (contiguous with MT) ^b	-52	-72	2	4.89	3630
<i>Dynamic Faces > Static Faces</i>					
R lateral middle occipital cortex (contiguous with MT, and extending into posterior superior temporal sulcus) ^a	42	-65	8	6.4	6116
R anterior temporal pole	52	18	-25	5.82	451
L lateral middle occipital cortex (contiguous with MT, and extending into posterior superior temporal sulcus) ^b	-52	-73	5	6.2	4134

^{a,b} Overlapping activation across the two contrasts in right and left hemispheres, respectively.

consistent with other studies implicating these regions in navigation (van Assche, Kebets, Vuilleumier, & Assal, 2016; Burgess, 2008; Kravitz, Saleem, Baker, & Mishkin, 2011; Marchette Vass, Ryan, & Epstein, 2014; Persichetti & Dilks, 2016; Spiers & Maguire, 2007); and iii) the right and left precentral gyrus, perhaps reflecting motor imagery related to the task (Malouin, Richards, Jackson, Dumas, & Doyon, 2003). Crucially, none of these regions showed overlapping activation in the contrast of Dynamic Faces versus Static Faces ($p = .01$, FWE corrected), suggesting that this activation is specific to motion information in scenes. Next, we observed two additional regions in right lateral middle occipital cortex, and one other region in left lateral middle occipital cortex, which responded more to Dynamic Scenes versus Static Scenes. Importantly, however, these same regions also responded more to Dynamic Faces versus Static Faces, revealing that they are sensitive to motion information in general. Indeed, consistent with the ROI analysis above, these regions corresponded to MT, as confirmed by overlaying functional parcels for MT that were created using a group-constrained method in an independent set of subjects (see Method and Julian et al., 2012). Finally, we observed several regions responding more to Dynamic Faces versus Static Faces, including bilateral posterior superior temporal sulcus, consistent with previous studies of dynamic face information processing (Pitcher et al., 2011) and thus validating our paradigm, as well as a region in the right anterior temporal pole—a known face selective region (Kriegeskorte, Formisano, Sorger, & Goebel, 2007; Sergent, Ohta, & MacDonald, 1992)—suggesting that this region may also be sensitive to dynamic face information. Crucially, these same regions did not show overlapping activation with the contrast of Dynamic Scenes versus Static Scenes, indicating that this activation is specific to stimuli depicting dynamic face information.

To further explore the data, we also examined activation to the contrast of Dynamic Scenes minus Static Scenes at lower thresholds ($p = .05$, uncorrected). Here we found the same network of regions responding more to Dynamic Scenes than Static Scenes, as well as additional regions in the right and left calcarine sulcus (consistent with the ROI analysis, insofar as FC also responded more to scene motion than face motion, albeit less so than OPA), right insula, right temporal pole, and right and left precentral gyrus.

4. Discussion

Here we explored how the three known scene-selective regions in the human brain represent first-person perspective motion information through scenes—information critical for visually-guided navigation. In particular, we compared responses in OPA, PPA, and RSC to i) video clips depicting first-person perspective motion through scenes, and ii) static images taken from these very same movies, rearranged such that first-person perspective motion could not be inferred. We found that OPA represents first-person perspective motion information, while RSC and PPA do not. Importantly, the pattern of responses in OPA was not driven by domain-general motion sensitivity or low-level visual information. These findings are consistent with a

recent hypothesis that the scene processing system may be composed of two distinct systems: one system supporting navigation (including OPA, RSC, or both), and a second system supporting other aspects of scene processing, such as scene categorization (e.g., recognizing a kitchen vs a beach) (including PPA) (Dilks et al., 2011; Persichetti & Dilks, 2016). This functional division of labor mirrors the well-established division of labor in object processing between the dorsal (“how”) stream, implicated in visually-guided action, and the ventral (“what”) stream, implicated in object recognition (Goodale & Milner, 1992). Further, these data suggest a novel division of labor even among regions involved in navigation, with OPA particularly involved in guiding navigation through the immediately visible environment, and RSC supporting other aspects of navigation, such as navigation through the broader environment.

The hypothesized role of OPA in guiding navigation through the immediately visible environment is consistent with a number of recent findings. First, OPA represents two kinds of information necessary for such visually-guided navigation: sense (left vs right) and egocentric distance (near vs far from me) information (Dilks et al., 2011; Persichetti & Dilks, 2016). Second, OPA represents local elements of scenes, such as boundaries (e.g., walls) and obstacles (e.g., furniture)—which critically constrain how one can move through the immediately visible environment (Kamps et al., 2016). Third, the anatomical position of OPA within the dorsal stream, which broadly supports visually-guided action (Goodale & Milner, 1992), suggests that OPA may support a visually-guided action in scene processing, namely visually-guided navigation. Thus, given the above findings, along with the present report that OPA represents the first-person perspective motion information through scenes, we hypothesize that OPA plays a role in visually-guided navigation, perhaps by tracking the changing sense and egocentric distance of boundaries and obstacles as one moves through a scene.

Critically, we found that OPA only responded to motion information in scenes, not faces. This finding rules out the possibility that OPA is sensitive to motion information in general, and suggests that OPA may selectively represent motion information in scenes. However, our study did not test other kinds of motion information within the domain of scene processing, and thus it may be the case that OPA represents other kinds of scene motion information in addition to the first-person perspective motion tested here. One candidate is horizontal linear motion (e.g., motion experienced when looking out the side of a car). However, one recent study (Hacialihafiz & Bartels, 2015) found that while OPA is sensitive to horizontal linear motion, OPA does not selectively represent such motion information in scenes, but also responds to horizontal linear motion in phase-scrambled non-scene images. This lack of specificity suggests that the horizontal linear motion sensitivity in OPA may simply be inherited from low-level visual regions (indeed, while many low-level features were matched between the stimuli, the study did not compare responses in OPA to those in a low-level visual region, such as FC), and thus may not be useful for scene processing in particular. Another candidate is motion parallax information, a 2D motion cue allowing inference of the surrounding 3D layout.

Interestingly, however, a second recent study (Schindler & Bartels, 2016) found that OPA was not sensitive to such motion parallax information, at least in the minimal line drawing scenes tested there. Yet another candidate is optic flow information, which was likely abundant in our Dynamic Scene stimuli. Optic flow information provides critical cues for understanding movement through space (Britten & van Wezel, 1998), and thus may be a primitive source of information for a visually-guided navigation system. Indeed, while optic flow information is typically studied outside the context of scenes (e.g., using moving dot arrays), OPA has been shown to be sensitive to other scene “primitives” such as high spatial frequencies (Kauffmann, Ramanoel, Guyader, Chauvin, & Peyrin, 2015) and rectilinear features (Nasr, Echavarría, & Tootell, 2014), supporting this possibility. Taken together, these findings on motion processing in OPA are thus far consistent with the hypothesis that OPA selectively represents motion information relevant to visually-guided navigation. However, future work will be required to address the precise types of motion information (e.g., optic flow information) in scenes that drive OPA activity.

As predicted, RSC did not respond selectively to first-person perspective motion through scenes, consistent with current hypotheses that RSC supports other aspects of navigation involving the integration of information about the current scene with representations of the broader environment (Burgess, Becker, King, & O’Keefe, 2001; Byrne, Becker, & Burgess, 2007; Epstein & Vass, 2015; Marchette et al., 2014). For example, RSC has been suggested to play a role in landmark-based navigation (Auger, Mullally, & Maguire, 2012; Epstein & Vass, 2015), location and heading retrieval (Epstein et al., 2007; Marchette et al., 2014; Vass & Epstein, 2013), and the formation of environmental survey knowledge (Auger et al., 2015; Wolbers & Buchel, 2005). Importantly, our stimuli depicted navigation through limited portions (each clip lasted only 3 sec) of unfamiliar scenes. As such, it was not possible for participants to develop survey knowledge of the broader environment related to each scene, or what is more, to integrate cues about self-motion through the scene with such survey knowledge. The present single dissociation, with OPA, but not RSC, responding selectively to dynamic scenes, therefore suggests a critical, and previously unreported division of labor amongst brain regions involved in scene processing and navigation more generally. In particular, we hypothesize that while RSC represents the broader environment associated with the current scene, in order to support navigation to destinations outside the current view (e.g., to get from the cafeteria to the psychology building), OPA rather represents the immediately visible environment, in order to support navigation to destinations within the current view (e.g., to get from one side of the cafeteria to the other). Of course, since here we did not test how these regions support navigation through the broader environment, it might still be the case that OPA supports both navigation through the immediately visible scene and the broader environment. Future work will be required to test this possibility.

Finally, our group analysis revealed a network of regions extending from lateral superior occipital cortex (corresponding to OPA) to superior parietal lobe that were sensitive to first-

person perspective motion information through scenes. This activation is consistent with a number of studies showing parietal activation during navigation tasks (Burgess, 2008; Kravitz et al., 2011; Marchette et al., 2014; Persichetti & Dilks, 2016; Spiers & Maguire, 2007; van Assche, Kebets, Vuilleumier, & Assal, 2016). Interestingly, this activation is also consistent with neuropsychological data from patients with damage to posterior parietal cortex who show a profound inability to localize objects with respect to the self (a condition known as egocentric disorientation) (Aguirre & D’Esposito, 1999; Ciaramelli, Rosenbaum, Solcz, Levine, & Moscovitch, 2010; Stark, Coslett, & Saffran, 1996; Wilson et al., 2005).

In sum, here we found that OPA, PPA, and RSC differentially represent the first-person perspective motion information experienced while moving through a scene, with OPA responding more selectively to such motion information than RSC and PPA. This enhanced response in OPA to first-person perspective motion information, a critical cue for navigating the immediately visible scene, suggests the novel hypothesis that OPA is distinctly involved in visually-guided navigation, while RSC and PPA support other aspects of navigation and scene recognition.

Acknowledgments

We would like to thank the Facility for Education and Research in Neuroscience (FERN) Imaging Center in the Department of Psychology, Emory University, Atlanta, GA. We would also like to thank Alex Liu and Ben Deen for technical support. The work was supported by Emory College, Emory University (DD) and National Institute of Child Health and Human Development grant T32HD071845 (FK). The authors declare no competing financial interests.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.cortex.2016.06.022>.

REFERENCES

- Aguirre, G. K., & D’Esposito, M. (1999). Topographical disorientation: A synthesis and taxonomy. *Brain*, 122(Pt 9), 1613–1628.
- van Assche, M., Kebets, V., Vuilleumier, P., & Assal, F. (2016). Functional dissociations within posterior parietal cortex during scene integration and viewpoint changes. *Cerebral Cortex*, 26(2), 586–598. <http://dx.doi.org/10.1093/cercor/bhu215>.
- Auger, S. D., Mullally, S. L., & Maguire, E. A. (2012). Retrosplenial cortex codes for permanent landmarks. *Plos One*, 7(8), e43620. <http://dx.doi.org/10.1371/journal.pone.0043620>.
- Auger, S. D., Zeidman, P., & Maguire, E. A. (2015). A central role for the retrosplenial cortex in de novo environmental learning. *Elife*, 4. <http://dx.doi.org/10.7554/eLife.09031>.
- Britten, K. H., & van Wezel, R. J. (1998). Electrical microstimulation of cortical area MST biases heading perception in monkeys. *Nature Neuroscience*, 1(1), 59–63. <http://dx.doi.org/10.1038/259>.

- Burgess, N. (2008). Spatial cognition and the brain. *Annals of the New York Academy of Sciences*, 1124, 77–97. <http://dx.doi.org/10.1196/annals.1440.002>.
- Burgess, N., Becker, S., King, J. A., & O'Keefe, J. (2001). Memory for events and their spatial context: Models and experiments. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 356(1413), 1493–1503. <http://dx.doi.org/10.1098/Rstb.2001.0948>.
- 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), 340–375. <http://dx.doi.org/10.1037/0033-295X.114.2.340>.
- Ciaramelli, E., Rosenbaum, R. S., Solcz, S., Levine, B., & Moscovitch, M. (2010). Mental space travel: Damage to posterior parietal cortex prevents egocentric navigation and reexperiencing of remote spatial memories. *Journal of Experimental Psychology-Learning Memory and Cognition*, 36(3), 619–634. <http://dx.doi.org/10.1037/a0019181>.
- Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine*, 42(6), 1014–1018.
- Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *The Journal of Neuroscience*, 31(31), 11305–11312. <http://dx.doi.org/10.1523/JNEUROSCI.1935-11.2011>.
- Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of Neuroscience*, 33(4), 1331–1336a. <http://dx.doi.org/10.1523/JNEUROSCI.4081-12.2013>.
- Epstein, R. A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12(10), 388–396.
- Epstein, R. A., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598–601. <http://dx.doi.org/10.1038/33402>.
- 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*, 27(23), 6141–6149. <http://dx.doi.org/10.1523/JNEUROSCI.0799-07.2007>.
- Epstein, R. A., & Vass, L. K. (2015). Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 370(1663), 20150001. <http://dx.doi.org/10.1098/Rstb.2015.0001>.
- Gibson, J. J. (1950). *Perception of the visual world*. Boston: Houghton-Mifflin.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience*, 15, 20–25.
- Hacialihafiz, D. K., & Bartels, A. (2015). Motion responses in scene-selective regions. *NeuroImage*, 118, 438–444. <http://dx.doi.org/10.1016/j.neuroimage.2015.06.031>.
- Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, 60(4), 2357–2364.
- Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *NeuroImage*, 132, 417–424.
- Kanwisher, N., & Dilks, D. (2013). The functional organization of the ventral visual pathway in humans. In L. Chalupa, & J. Werner (Eds.), *The new visual neurosciences*. Cambridge, MA: The MIT Press.
- Kauffmann, L., Ramanoel, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, 112, 86–95. <http://dx.doi.org/10.1016/j.neuroimage.2015.02.058>.
- 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://dx.doi.org/10.1038/nrn3008>.
- Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104(51), 20600–20605. <http://dx.doi.org/10.1073/pnas.0705654104>.
- Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology*, 42(3), 225–238.
- Malouin, F., Richards, C. L., Jackson, P. L., Dumas, F., & Doyon, J. (2003). Brain activations during motor imagery of locomotor-related tasks: A PET study. *Human Brain Mapping*, 19(1), 47–62. <http://dx.doi.org/10.1002/hbm.10103>.
- 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://dx.doi.org/10.1038/nn.3834>.
- Nasr, S., Echavarria, C. E., & Tootell, R. B. H. (2014). Thinking outside the box: Rectilinear shapes selectively activate scene-selective cortex. *Journal of Neuroscience*, 34(20), 6721–6735. <http://dx.doi.org/10.1523/JNEUROSCI.4802-13.2014>.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Park, S., Brady, T. F., Greene, M. R., & 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. *The Journal of Neuroscience*, 31(4), 1333–1340. <http://dx.doi.org/10.1523/JNEUROSCI.3885-10.2011>.
- Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex*, 77, 155–163.
- Pitcher, D., Dilks, D. D., Saxe, R. R., Triantafyllou, C., & Kanwisher, N. (2011). Differential selectivity for dynamic versus static information in face-selective cortical regions. *NeuroImage*, 56(4), 2356–2363.
- Schindler, A., & Bartels, A. (2016). Motion parallax links visual motion areas and scene regions. *NeuroImage*, 125, 803–812. <http://dx.doi.org/10.1016/j.neuroimage.2015.10.066>.
- Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain*, 115(Pt 1), 15–36.
- Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., et al. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, S208–S219. <http://dx.doi.org/10.1016/J.Neuroimage.2004.07.051>.
- Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44(1), 83–98. <http://dx.doi.org/10.1016/j.neuroimage.2008.03.061>.
- Spiers, H. J., & Maguire, E. A. (2007). A navigational guidance system in the human brain. *Hippocampus*, 17(8), 618–626. <http://dx.doi.org/10.1002/hipo.20298>.
- Stark, M., Coslett, H. B., & Saffran, E. M. (1996). Impairment of an egocentric map of locations: Implications for perception and action. *Cognitive Neuropsychology*, 13(4), 481–523. <http://dx.doi.org/10.1080/0264322996381908>.
- Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., et al. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *The Journal of Neuroscience*, 15(4), 3215–3230.

- Vass, L. K., & Epstein, R. A. (2013). Abstract representations of location and facing direction in the human brain. *Journal of Neuroscience*, 33(14), 6133–6142. <http://dx.doi.org/10.1523/jneurosci.3873-12.2013>.
- Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *The Journal of Neuroscience*, 29(34), 10573–10581. <http://dx.doi.org/10.1523/JNEUROSCI.0559-09.2009>.
- Wilson, B. A., Berry, E., Gracey, F., Harrison, C., Stow, I., Macniven, J., et al. (2005). Egocentric disorientation following bilateral parietal lobe damage. *Cortex*, 41(4), 547–554. [http://dx.doi.org/10.1016/S0010-9452\(08\)70194-1](http://dx.doi.org/10.1016/S0010-9452(08)70194-1).
- Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M., & Nichols, T. E. (2014). Permutation inference for the general linear model. *NeuroImage*, 92, 381–397. <http://dx.doi.org/10.1016/j.neuroimage.2014.01.060>.
- Wolbers, T., & Buchel, C. (2005). Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *Journal of Neuroscience*, 25(13), 3333–3340. <http://dx.doi.org/10.1523/jneurosci.4705-04.2005>.