



Research Report

Two scene navigation systems dissociated by deliberate versus automatic processing



Shosuke Suzuki ^{a,1}, Frederik S. Kamps ^{a,c,1}, Daniel D. Dilks ^a and Michael T. Treadway ^{a,b,*}

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

^b Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, United States

^c Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:

Received 27 July 2020

Reviewed 19 September 2020

Revised 25 November 2020

Accepted 20 March 2021

Action editor Giuseppe Iaria

Published online 22 April 2021

Keywords:

fMRI

Navigation

Occipital place area

Retrosplenial complex

Scene perception

ABSTRACT

Successfully navigating the world requires avoiding boundaries and obstacles in one's immediately-visible environment, as well as finding one's way to distant places in the broader environment. Recent neuroimaging studies suggest that these two navigational processes involve distinct cortical scene processing systems, with the occipital place area (OPA) supporting navigation through the local visual environment, and the retrosplenial complex (RSC) supporting navigation through the broader spatial environment. Here we hypothesized that these systems are distinguished not only by the scene information they represent (i.e., the local visual versus broader spatial environment), but also based on the automaticity of the process they involve, with navigation through the broader environment (including RSC) operating *deliberately*, and navigation through the local visual environment (including OPA) operating *automatically*. We tested this hypothesis using fMRI and a maze-navigation paradigm, where participants navigated two maze structures (*complex* or *simple*, testing representation of the broader spatial environment) under two conditions (*active* or *passive*, testing deliberate versus automatic processing). Consistent with the hypothesis that RSC supports deliberate navigation through the broader environment, RSC responded significantly more to complex than simple mazes during active, but not passive navigation. By contrast, consistent with the hypothesis that OPA supports automatic navigation through the local visual environment, OPA responded strongly even during passive navigation, and did not differentiate between active versus passive conditions. Taken together, these findings suggest the novel hypothesis that navigation through the broader spatial environment is deliberate, whereas navigation through the local visual environment is automatic, shedding new light on the dissociable functions of these systems.

© 2021 Elsevier Ltd. All rights reserved.

* Corresponding author. Department of Psychology, Emory University, Atlanta, GA 30306, United States.

E-mail address: mtreadway@emory.edu (M.T. Treadway).

¹ These authors contributed equally to the present work.

<https://doi.org/10.1016/j.cortex.2021.03.027>

0010-9452/© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

To successfully navigate our spatial world, we must both move about the immediately visible environment, avoiding boundaries and obstacles (e.g., walking around one's bedroom, not bumping into the walls or furniture), and situate the immediately visible environment within the broader spatial environment (e.g., knowing that the bathroom is down the hall and to the right of one's bedroom), ultimately enabling us to find our way from the current place to another distant place. These navigational processes are logically independent: understanding that the kitchen is through the door to my right does not help me perceive how to locomote past the coffee table obstructing my path; likewise, correctly perceiving a passable route around the coffee table does not tell me whether I should take the door to the left or to the right to reach the kitchen. Consistent with this logic, recent fMRI evidence suggests that these two navigational abilities may depend on two different cortical scene processing systems. In particular, the scene-selective occipital place area (OPA) is hypothesized to support navigation through the local, immediately visible environment (often referred to as “visually-guided navigation”) (Dillon et al., 2018; Julian et al., 2016; Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2016; Kamps, Lall, & Dilks, 2016; Persichetti and Dilks, 2016, 2018), while another scene-selective cortical region – the retrosplenial complex (RSC) – is hypothesized to support navigation through the broader spatial environment beyond the currently visible scene (often referred to as “map-based” or “memory-guided navigation”) (Iaria et al., 2007; Maguire, 2001; Marchette et al., 2014; Persichetti & Dilks, 2019; Sherrill et al., 2013; Wolbers & Büchel, 2005). Importantly, the precise function of each navigation system, and the nature of the dissociation between them, remains poorly understood.

Here we hypothesize that these two scene navigation systems differ not only in the scene information they represent (e.g., about the immediately visible scene or the broader spatial environment), but also the kind of process they involve. In particular, we hypothesize that navigation through the broader environment involves *deliberate* processing (i.e., depending strongly on conscious, intentional control), while navigation through the local visual environment involves *automatic* processing (i.e., operating regardless of conscious, intentional control). To our knowledge this hypothesis has never been tested directly, despite being supported by several lines of evidence. First, a recent cellular calcium imaging study in mice found that RSC neurons show robust representation of spatial information during active locomotion through VR environments, but that this information is greatly diminished when rodents are presented with the same visual information under passive conditions (i.e., without locomoting) (Mao et al., 2020). By contrast, human fMRI studies of OPA show that even passive viewing is sufficient to evoke OPA responses to navigationally relevant motion or navigational affordance information (Kamps, Lall, et al., 2016; Bonner and Epstein, 2017). Further, most RSC studies to our knowledge have employed relatively “active” tasks, such as judgment of relative direction tasks, consistent with the idea that deliberate coding is critical for driving RSC responses even in

humans. Second, navigation through the broader spatial environment is challenging, with large individual differences in ability (Ishikawa & Montello, 2006; Weisberg et al., 2014; Wen et al., 2011), suggestive of a computationally intensive, and therefore relatively deliberate process. By contrast, navigation through the local visual environment operates seemingly effortlessly and unconsciously, suggestive of an automatic process. Third, navigation through the local visual environment must operate “online” on a rapid timescale in order to provide dynamic information about the changing distance and direction of boundaries and obstacles as one moves through space, which would be facilitated by automatic processing. By contrast, navigation through the broader spatial environment occurs over longer timescales, given that it takes time to move from the current scene to another scene out of view, and therefore requires only intermittent, deliberate processing. Consistent with this notion, a bottom-up perspective of an efficient navigation system in mobile robotics has long recognized the importance of having both a reactive system characterized by real-time responses to the environment as well as a system for acquiring knowledge about the world that can be accessed when difficulties are encountered by the reactive system (Arkin, 1990). Fourth, navigation through the broader environment requires integration of, and possibly reasoning about, numerous sources of visual, mnemonic, and somatosensory information, suggestive of a deliberate process, while navigation through the local environment may predominantly depend on inputs from earlier visual systems, therefore allowing a relatively automatic process. RSC is anatomically well-suited to integrate information from multiple modalities, with dense connectivity to regions of visual cortex, the medial temporal lobe, and the parietal cortex (Epstein, 2008; Kravitz et al., 2011), whereas OPA is situated relatively early in the dorsal visual pathway (adjacent to areas V3A/B) (Grill-Spector, 2003; Nasr et al., 2011; Silson et al., 2015), and therefore well-suited to process visual information automatically via the initial feedforward processing sweep from earlier visual cortex. Indeed, RSC is proposed to be part of a “memory” network, whereas OPA is proposed to be part of a “perception” network (Baldassano et al., 2016; Silson et al., 2016). Importantly, however, no study has directly compared RSC and OPA responses while participants actively engaged in navigation versus passively viewed navigation. Directly comparing active and passive navigation in these regions within the same study is essential, since an obvious alternative hypothesis is that any scene region will respond more during active than passive conditions, given that the active condition inherently involves greater attentional processing.

Here we used fMRI during a maze-navigation paradigm that required participants to virtually navigate through a 3D maze environment. To study representation of the broader environment, participants navigated through two kinds of mazes differing in the complexity of the broader environment: a simple, one-turn environment or a more complex, two-turn environment. To study automatic versus deliberate processing, participants navigated through the mazes by either controlling their own movements (i.e., “active”) or by simply watching a prerecorded animation (i.e., “passive”). Finally, given that our prediction that an “automatic” navigation

system should not differentiate between active and passive conditions (i.e., a null effect), we also included an additional “no-navigation” control condition in which participants simply viewed the scene from a stationary perspective, before being teleported to the end of the maze. If RSC supports deliberate navigation through the broader spatial environment, then RSC will respond more to information about the broader environment (i.e., more to complex than simple mazes) during active navigation relative to passive navigation, and will not respond to the passive visual experience of navigation through scenes. By contrast, if OPA supports automatic navigation through the local visual environment, then OPA will respond strongly to even the passive visual experience of navigation through scenes (relative to the “no navigation” condition), and will not distinguish simple from complex mazes, nor differentiate between active and passive navigation conditions.

2. Materials and methods

We report how we determined our sample size, all data exclusions, all inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all manipulations, and all measures in the study.

2.1. Participants

We recruited 20 healthy adults from the Atlanta community through fliers and online advertisements. The sample size was predetermined based on previous publications (Kamps, Lall, et al., 2016; Marchette et al., 2015). Eligibility was determined using an online pre-screening survey. Eligible participants were right-handed, English-speaking individuals between the ages of 18–35. Individuals were excluded if they: (1) had any contraindications for magnetic resonance imaging (e.g., claustrophobia, metallic implants, central nervous system disorders, pregnancy in females); (2) were currently taking psychoactive medications, investigational drugs, or drugs that affect blood flow (e.g., for hypertension); or (3) reported current medical, neurological, or psychiatric illnesses. All participants gave informed consent and had normal or corrected-to-normal vision.

One participant was excluded because we could not localize any scene-selective cortical regions from the functional localizer. Thus, the final sample included 19 participants ($M_{\text{age}} = 25.36$; $SD_{\text{age}} = 5.55$; 13 females). Upon completion of study procedures, participants were compensated for their participation. All procedures were approved by the Emory Institutional Review Board.

2.2. Experimental design

We used a region of interest (ROI) approach, in which we localized scene-selective cortical regions via a functional localizer run. Then, we used an independent set of experimental runs to investigate the responses of these regions while participants completed the maze-navigation task. Post-hoc whole-brain analyses were conducted to examine the responses to the experimental conditions beyond scene-

selective cortex. Data are accessible through the Open Science Framework (<https://osf.io/8xvz7/>). No part of the study procedures or analyses were pre-registered prior to the research being conducted. Prior to scanning, all participants underwent a 15-min training procedure on the maze-navigation task (see *Training and Calibration*).

For the functional localizer run, ROIs were identified using a standard method (Epstein & Kanwisher, 1998). Specifically, in a blocked design, participants saw pictures of scenes and objects for a total of 336 s, consisting of 8 blocks per stimulus category. Each block was 16 s long and contained 20 pictures from the same category. Each picture was presented for 300 msec, followed by a 500 msec interstimulus interval (ISI). We also included five 16 s fixation blocks: 1 at the beginning; 3 in the middle interleaved after every 4 blocks; and 1 at the end of the run. Participants performed a one-back task, responding with a button press every time the same picture presented twice in a row.

For the experimental runs (i.e., the maze-navigation task), participants were required to virtually navigate through single-path mazes via a 4-button box with the right (i.e., dominant) hand. Specifically, one button moved the participant forward ($r = 2.2$ units/s), and two buttons each rotated the participant clockwise and counterclockwise ($\omega = .5\pi$ rad/sec). Holding down the buttons continuously applied the translational/rotational effects. Acceleration was applied to participant motion to mimic real-life motion. Only one type of motion was allowed at any given moment (i.e., pressing multiple buttons resulted in no motion).

Each trial was associated with one of four navigation conditions, including i) an *active navigation* condition, in which participants were required to advance through the maze using button presses to move and rotate, ii) a *passive navigation* condition, in which participants were required to simply view moving through the maze without making any action, and iii) a *no navigation* condition, in which participants waited for the approximate duration of the maze until they were teleported to the end of the maze. One additional condition was included for the purpose of another study, an *effortful navigation* condition in which participants were required to repeatedly press the button to advance through the maze. Critically, to test for representation of the broader environment, we varied the structure of the mazes. Specifically, there were “simple” mazes which required a single 90° turn (left or right) and “complex” mazes which required two 90° turns (left-then-right or right-then-left). The presence of an additional turn therefore specifically increased information about the broader environment, since an additional turn i) may be coded as an additional “landmark”, ii) further segments the space into a greater number of distinct locations (e.g., 3 corridors instead of 2), and iii) increases the number of changes in heading direction as one moves through the maze. To ensure any differences in response to the complex and simple mazes depended on representations derived from prior experience (as opposed to direct perception of the current environment), the complete structure of the maze was never fully in view (the participants never see the maze from above, for example), and there was no perceptual information that differentiated the complex and simple mazes: Both types of mazes were comprised of 1×1 unit² floors placed adjacently to form a

path, bounded by 1×1 unit² walls, and the participant view was set at .6 units above the floor, with a 10° nose-down pitch. All other aspects of the two maze types were matched. Consequently, because at any given moment the only perceptual information available was an egocentric view of the scene (Fig. 1), there was no moment-to-moment perceptual information that could distinguish (for example) the second left turn of a complex maze from the first left turn of a simple maze. This aspect of the design was crucial for our hypothesis, as it implies that if a brain area responds more to a complex maze than a simple maze, this difference cannot be attributed to perceptual information, and rather must be due to what the navigator “knows” about the larger path she/he has traveled through based on prior experience.

Each trial proceeded as follows (Fig. 1): (1) Start phase: At the beginning of each trial, the participant's position was initialized to the beginning of the maze, immediately followed by a cue phase. (2) Cue phase: The floor tile immediate to the participant's view changed color, which represented a cue informing the participants to the navigation condition for that trial. The colored cue lasted for 2 s before returning to its original floor texture, and the participant was rendered immobile during the cue presentation. (3) Jittered interstimulus interval (ISI): The cue was immediately followed by a jittered fixation period, whereby a ‘+’ was rendered on top of the current scene for a Poisson-distributed duration with a mean of 2.5 s. The participant was still restricted from moving during this period. (4) Navigation phase: After the fixation plus disappeared, participants completed the respective navigation condition, as detailed above. Participants failed the trial if

they did not complete navigation within a lenient time limit (5.5 s) during the active and effortful navigation conditions, or if they pressed any buttons during the passive or no navigation conditions. (5) Jittered ISI: Once participants successfully reached the door at the end of the maze (the “goal”), the player was again rendered immobile for an ISI jittered around 2.5 s. (6) Reward phase (used for another study): Following the ISI, participants were presented with an animation of the door opening followed by a monetary reward, represented by a coin with the dollar amount rendered on its surface. Each trial was associated with one of 4 bins of reward magnitudes (\$0, \$1.68–2.78, \$2.79–\$3.89, \$3.90–5.00), from which an amount was randomly selected. (7) Rating phase (used for another study): Once every 4 trials, the participant was asked to make a mood rating on a Likert-scale between 1 (not happy at all) to 4 (very happy) using a button-press. (8) Jittered inter-trial interval (ITI): Finally, participants were presented with a ‘+’ rendered on a gray screen for a duration jittered around 3 s.

Each participant completed 3 runs with 32 trials each (~11 min/run). Order of trials was pseudorandomized to balance the number of task conditions across runs and presented in a fixed order (see Supplementary Table 1 for trial order). Participants successfully completed navigation on $93 \pm 1\%$ of the trials (active navigation: 95%, passive navigation: 99%, no navigation: 93%, effortful navigation: 85%). Mean completion times for the Navigation phase fell within 100 msec across conditions (active navigation: 4.6 s, passive navigation: 4.7 s, no navigation: 4.7 s, effortful navigation: 4.7 s). A video demo of the maze-navigation task is available on <https://osf.io/8xvz7/>.

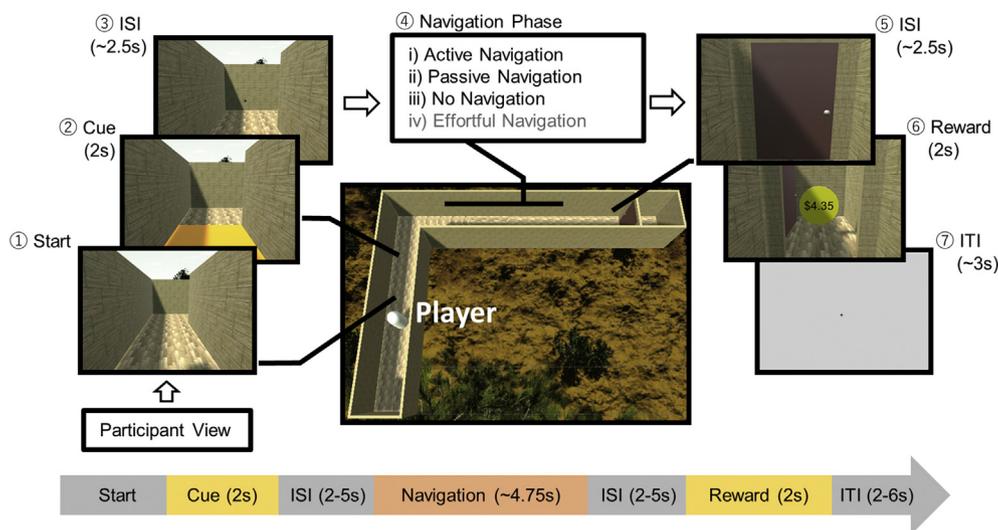


Fig. 1 – Schematic representation of a single trial of the maze-navigation task. On each trial, participants (labeled as Player) navigated through a maze structure in first-person viewpoint. First, participants made an approach to the cue location, which triggered the floor color to change for 2 s, indicating the navigation condition on each trial. Second, given a particular color, the participant then engaged in one of four navigation conditions: (i) the active navigation condition requiring individuals to move using button pressing; (ii) the passive-navigation condition requiring individuals to watch navigation without pressing any buttons; (iii) the no-navigation condition requiring individuals to wait for a specified duration without pressing any buttons before being ‘teleported’ to the goal, and (iv) the effortful-navigation condition (included for the purposes of another study) requiring individuals to rapidly press buttons in order to move. The center image shows the top view of one of the four maze structures (i.e., single right turn; the others were single left turn, right-then-left turn, left-then-right turn). Participants never saw the maze structure from this top view perspective.

2.3. Training and calibration

Participants completed a 15-min training procedure of the maze-navigation task prior to the scanning session, to ensure that they understood the instructions and could complete the task. First, participants were told that they will be navigating through virtual mazes to obtain monetary rewards. They were introduced to the player controls and all navigation conditions. To incentivize the participants, they were also told that a proportion of the reward they obtain on each trial will be added to their payment as a bonus. Once participants indicated that they understood the task and could follow the instructions, they completed 16 practice trials on a laptop computer. To maximize the effect of practice, participants used the same hand and fingers used to perform the task in the MRI scanner.

2.4. Data acquisition & analysis

The maze-navigation task was programmed using Unity 3D (Unity Technologies ApS). Stimuli were presented via back-projection mirror, and participants completed the maze-navigation task and functional localizer runs using an MR-compatible 4-button box (Current Designs Inc). Foam pads placed around participants' heads were used to minimize motion.

Participants were scanned in a 3-T Siemens TIM Trio scanner (Siemens AG) with a 32-channel head-coil. We used multiband functional and structural imaging (Feinberg et al., 2010; Feinberg & Setsompop, 2013; Xu et al., 2013), similarly used by the Human Connectome Project consortium (Van Essen et al., 2013). Each session began with a 3-plane localizer scan for slice alignment, and a single-shot, high-resolution structural MPRAGE sequence (TR/TE = 1900/2.27 msec; flip angle = 9°; FoV = 250 × 250 mm; 192 × 1.0 mm slices). BOLD functional images were acquired with T2*-weighted EPI sequences with a multiband acceleration factor of 4 (TR/TE = 1000/30.0 msec; flip angle = 65°; FoV = 220 × 220 mm; 52 × 3.00 mm slices).

Functional images were preprocessed using SPM12 scripts through NeuroElf v1.1, following best practice guidelines described by the Human Connectome Project for multiband data analysis (Glasser et al., 2013). Specifically, images were co-registered to the structural image, motion-corrected, warped to the Montreal Neurological Institute (MNI) template, and smoothed using a Gaussian filter (6 mm full width-half maximum). Raw and preprocessed data were subjected to multiple tests for quality assurance and inspected for spiking and motion. Volumes were discarded if the root mean square of motion parameters exceeded a single voxel dimension (3 mm), or if striping was identified through visual inspection. Subject-level modeling of trial events was conducted using robust regression to reduce the influence of strong outliers.

To identify scene-selective regions, the scene and object blocks in the functional localizer run were included in a subject-level GLM as regressors of interest. Motion parameters and their squares, as well as predictors forming a Discrete Cosine Transform basis set (cut-off: 128 s) for high-pass temporal filtering, were included as additional nuisance regressors. For each participant, fixed-effects contrasts

(Scenes > Objects) were generated to individually define ROIs. Specifically, we identified scene-selective ROIs (bilateral OPA, RSC, and a third scene-selective control region – the parahippocampal place area, PPA) at a voxelwise threshold of $p < 10^{-6}$, uncorrected (Fig. 2; Supplementary Fig. 1). Peak MNI coordinates and cluster sizes for each ROI in each hemisphere are provided in Supplementary Table 2.

For the maze-navigation task, the Cue, Navigation, and Reward phases were included in a subject-level GLM as regressors of interest. A separate regressor was included for each navigation condition at each phase, and additionally for each maze structure during the Navigation phase. Regressors for the Reward phase were further separated into rewarded and non-rewarded trials. We also included the Start and Rating phases as well as the ISIs in the model to omit their influences on the implicit baseline. In addition, motion parameters and their squares, as well as DCT basis functions for high-pass temporal filtering were again included as nuisance regressors. We then extracted beta parameters for the navigation conditions for each maze structure from each scene-selective ROI for each participant, and conducted repeated-measures ANOVAs on the neural response for each ROI. Specifically, we compared the complex and simple mazes to test for spatial representation of the broader environment, and their interaction with the active and passive navigation conditions to test deliberate versus automatic processing. Additionally, to examine responses to the visual experience of navigation, we compared the passive- and no-navigation conditions. Further, to evaluate the relative evidence for null effects within OPA, we used the R “BayesFactor” package (<https://cran.r-project.org/web/packages/BayesFactor/>) to conduct post-hoc Bayesian analyses. We specified a standard Cauchy prior for the effect size and a noninformative Jeffreys

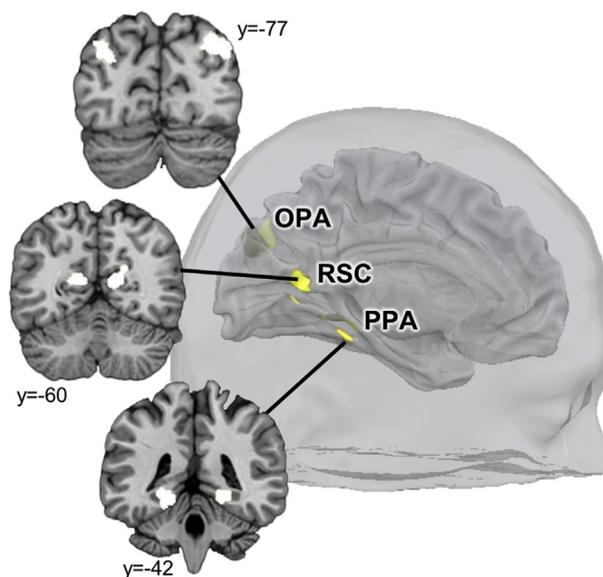


Fig. 2 – Localization of scene-selective regions (OPA, RSC, and PPA) in a sample participant. Each region was individually localized for each participant using the Scene > Object contrast in the functional localizer task, using a voxel-wise threshold of $p < 10^{-6}$, uncorrected.

prior for the variance (i.e., JZS prior) (Rouder et al., 2009) to compute Bayes factors (BF_{01}) for pairwise t-tests examining the effect of maze structure and complexity within OPA.

To examine whole-brain effects, we combined the subject-level GLMs to compute group-level contrasts. Specifically, two contrast maps that parallel the ROI analyses were generated: (i) Interaction between navigation condition (active, passive) and maze structure (complex, simple) and (ii) Comparison between the passive- and no-navigation conditions. For the purpose of our post-hoc analyses, regions were considered significant if they survived familywise-error correction at $p < .05$, achieved using a voxelwise threshold of $p < .005$ combined with a cluster-extent threshold of 192 voxels for the interaction map, and 128 voxels for the simple comparison. The latter analysis was also repeated using a more stringent voxelwise threshold of $p < .001$, combined with a cluster-extent threshold of 62 voxels.

For completeness, we also tested the effect of effort by comparing the effortful and active navigation conditions, and the effect of reward by comparing the rewarded and non-rewarded trials, with the prediction that neither OPA nor RSC will be sensitive to either of these effects. We confirm this prediction and report these results in the [Supplementary materials](#). We note that these data have been included in a separate publication focused on striatal encoding of effort and reward during navigation (Suzuki, Lawlor, Cooper, Arulpragasam, & Treadway, 2020). However, there is no overlap in any of the analyses performed, and the two manuscripts address distinct questions, anatomical regions, and neural systems.

3. Results

Given our hypothesis that RSC supports deliberate navigation through the broader spatial environment, while OPA supports automatic navigation through the local visual environment, we predicted that RSC will respond more to complex than simple mazes (reflecting information about the broader spatial environment) during active navigation to a greater extent than during passive navigation, while OPA responses will not depend on either the complexity of the broader environment or active navigation. To directly test this prediction, we conducted a 2 (region: RSC, OPA) \times 2 (maze structure: simple, complex) \times 2 (navigation condition: active, passive) repeated-measures ANOVA. As predicted, we found a significant three-way interaction between region, maze structure, and navigation condition ($F_{(1,18)} = 14.41$, $p = .001$, $\eta^2 = .45$). We then examined each region separately to parse out the nature of this interaction, and to test our predictions regarding the specific roles of RSC and OPA.

Given our hypothesis that RSC deliberately integrates the immediately visible environment with a broader spatial environment, we predicted that RSC will respond more to complex mazes than simple mazes to a greater extent during the active experience of navigating than to passive viewing. As predicted, a 2 (maze structure: simple, complex) \times 2 (navigation condition: active, passive) repeated-measures ANOVA revealed a significant interaction between maze structure and navigation condition ($F_{(1,18)} = 6.11$, $p = .02$, $\eta^2 = .25$), such that

RSC responded significantly more to complex than simple mazes during active navigation ($t_{(18)} = 4.90$, $p < .001$, $d = 1.12$), as compared to passive navigation ($t_{(18)} = 1.95$, $p = .07$, $d = .80$; [Fig. 3A](#); also see [Supplementary Fig. 2](#) for simple effects). By contrast, given our hypothesis that OPA is not involved in deliberate navigation through the broader environment, but instead supports automatic navigation through the local visual environment, we predicted that OPA will not represent information about the broader environment, nor respond more during active navigation than passive navigation. As predicted, we found only marginal effects of maze structure or navigation condition in OPA (maze structure: $F_{(1,18)} = 3.19$, $p = .09$, $\eta^2 = .15$; navigation condition: $F_{(1,18)} = 4.09$, $p = .06$, $\eta^2 = .19$). Crucially, unlike RSC, we found no maze structure by navigation condition interaction ($F_{(1,18)} = .01$, $p = .93$, $\eta^2 = .00$; [Fig. 3A](#)). Post-hoc Bayesian analysis revealed evidence supporting greater likelihood for the null hypothesis relative to the alternative hypothesis for the effects of maze structure and navigation condition within OPA (maze structure: $BF_{01} = 1.4$, navigation condition: $BF_{01} = 1.3$). Importantly, the significant difference in responses across RSC and OPA rules out the possibility that responses in RSC might simply be driven by general cognitive factors (e.g., increased attention, engagement, or effort, particularly during active navigation through complex mazes), since these general factors would be expected to drive responses across the cortex, not only in RSC specifically.

Next, we compared the effects of navigation condition and maze structure in RSC to those in PPA, another scene-selective region thought to be involved in the categorization of scenes (e.g., recognizing a kitchen versus a beach), but not navigation (Walther et al., 2009; Persichetti and Dilks, 2018, 2019). A 2 (region: RSC and PPA) \times 2 (maze structure: simple, complex) \times 2 (navigation condition: active, passive) repeated-measures ANOVA revealed a significant three-way interaction between region, maze structure, and navigation condition ($F_{(1,18)} = 6.97$, $p = .02$, $\eta^2 = .28$). Whereas PPA exhibited significant main effects of maze structure ($F_{(1,18)} = 6.25$, $p = .02$, $\eta^2 = .26$) and navigation condition ($F_{(1,18)} = 16.82$, $p = .001$, $\eta^2 = .48$), it did not, like RSC, show a significant interaction between maze structure and navigation condition ($F_{(1,18)} = 1.08$, $p = .31$, $\eta^2 = .06$; [Fig. 3A](#)). These findings support the hypothesis that RSC and PPA play dissociable roles in scene processing, and provide further evidence that RSC responses in the present study cannot be explained by general cognitive factors.

To explore responses beyond the functionally defined scene regions (including in the immediate vicinity of each region, as well as across the entire cortex), we next performed a group-level analysis testing for an interaction between maze structure and navigation condition across the whole brain. Consistent with the ROI analyses, this analysis revealed a significant interaction effect in bilateral RSC (rRSC peak: [18, -51, 9], $k = 59$; lRSC peak: [-12, 60, 9], $k = 48$), but not in any other region overlapping with or neighboring OPA or PPA ([Fig. 4](#); [Supplementary Table 3](#)). Additional activations were observed in postcentral gyrus (extending anteriorly into precentral gyrus and supplemental motor cortex), posterior cingulate/cerebellum, and the midbrain/thalamus ([Supplementary Table 3](#)). While not the focus of the current

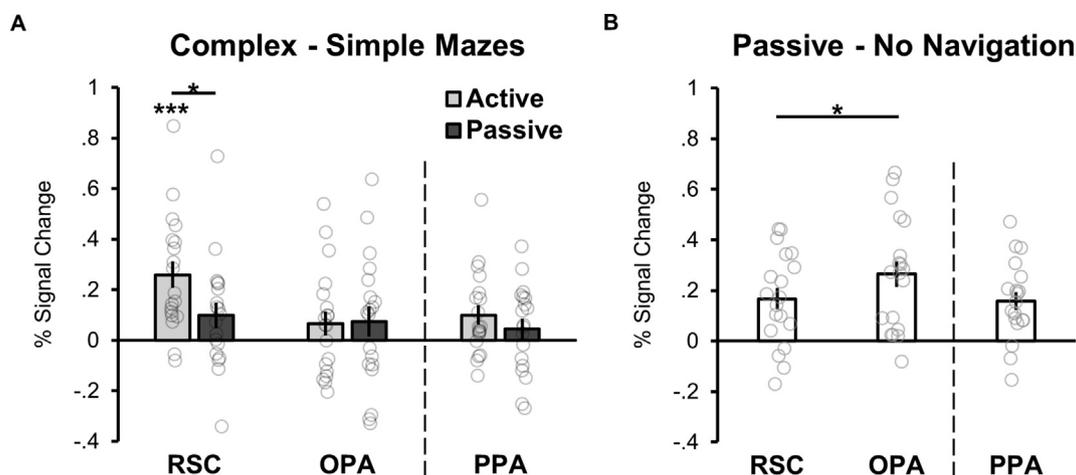


Fig. 3 – Effects of the broader spatial environment, active navigation, and the passive visual experience of navigation, in RSC and OPA. (A) Difference scores indexing representation of the broader spatial environment were calculated by subtracting the response to simple maze navigation from that to complex maze navigation, separately for the active and passive navigation conditions. RSC showed robust responses to information about the broader environment during active, but not passive navigation, while OPA did not respond differently to complex versus simple mazes, nor to active versus passive navigation. (B) Difference scores indexing representation of the passive experience of navigation through scenes were calculated by subtracting the response to the no navigation condition (i.e., simply looking at the maze, without moving through it) from that to the passive navigation condition. OPA responded significantly more than RSC to the passive visual experience of navigation. * $p < .05$, * $p < .001$. Error bars indicate standard error of the mean. Distributions of individual data are overlaid on each bar plot.**

investigation, we speculate that these activations may reflect motor planning or button pressing during the task. Taken together, these results confirm that RSC is involved in deliberate navigation through the broader spatial environment, while OPA and PPA are not.

The results above suggest unlike RSC, OPA responses did not differ depending on active versus passive navigation, consistent with the idea that OPA supports automatic navigation through the local visual environment. However, given that this inference relies on a null result, it is possible that we simply failed to measure navigational responses in OPA. Thus, to ensure that navigational responses could be detected in OPA in the present study, we next sought to replicate previous findings that OPA responds more to scene motion information (e.g., the first-person perspective motion experienced while moving through a scene) than does RSC (Hacihalhafiz & Bartels, 2015; Kamps, Lall, et al., 2016, Kamps et al., 2020; Pitcher et al., 2019). To do so, we compared responses to passive navigation – which should be sufficient to drive responses in a region that automatically processes information relevant to guiding navigation through space – with those to the “no navigation” control condition, which showed the same maze scene from a static perspective, and therefore did not depict the visual experience of navigation. A 2 (region: OPA, RSC) \times 2 (condition: passive navigation, no navigation) repeated-measures ANOVA revealed a significant region \times condition interaction ($F_{(1,18)} = 10.06$, $p = .005$, $\eta^2 = .36$), with OPA responding significantly more to the passive visual experience of navigation than RSC (Fig. 3B; also see Supplementary Fig. 2 for simple effects). The finding of a significant interaction also 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 condition. This finding replicates previous studies that OPA shows stronger responses to even the passive visual experience of navigation, relative to RSC, and confirms that navigational responses could be detected in the present study, consistent with the hypothesis that OPA represents the visual experience of navigation automatically. Note, however, that this comparison is incomplete as a test of the role of OPA in navigation through the local visual environment, since the passive and no navigation conditions are not matched in terms of lower-level, dynamic visual information. A complete test would therefore also measure responses to other dynamic and static conditions that do not involve navigationally relevant visual information, predicting selective responses in OPA to navigationally relevant dynamic information only. While not tested here, several previous studies have found precisely this pattern in OPA (Kamps, Lall, et al., 2016, Kamps et al., 2020; Pitcher et al., 2019).

We next compared OPA responses to the passive experience of navigation with those in PPA. Once again, a 2 (region: OPA, PPA) \times 2 (condition: passive navigation, no navigation) repeated-measures ANOVA revealed a significant region \times condition interaction ($F_{(1,18)} = 9.72$, $p = .006$, $\eta^2 = .35$), with OPA responding significantly more to the passive visual experience of navigation than PPA (Fig. 3B), consistent with previous findings (Kamps, Lall, et al., 2016, Kamps et al., 2020; Pitcher et al., 2019).

Finally, to explore responses beyond the functionally defined scene regions, we performed a group-level analysis testing for the contrast of passive navigation > no navigation across the whole brain. Consistent with the ROI analyses, this

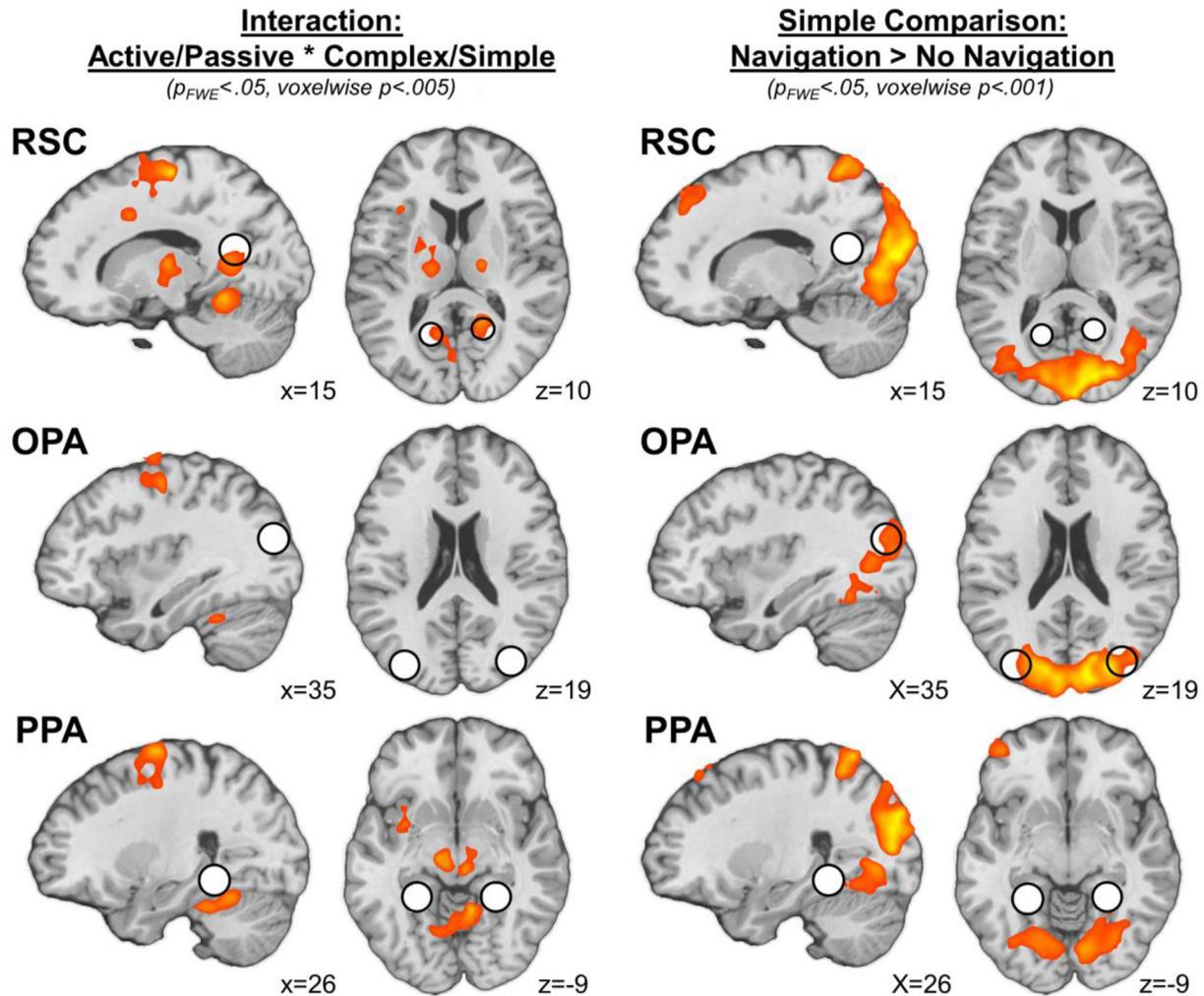


Fig. 4 – Results of whole-brain analyses. Consistent with ROI analyses, an interaction between navigation condition (active, passive) and maze structure (complex, simple) was observed in a region overlapping with RSC ($p_{FWE} < .05$ cluster corrected, voxelwise threshold $p < .005$). A region overlapping with OPA exhibited greater response to passive navigation compared to the no-navigation condition ($p_{FWE} < .05$ cluster corrected, voxelwise threshold $p < .001$). For visualization of ROI location, spheres ($r = 8$ mm) were generated around the average peak position of each scene-selective region. RSC: retrosplenial complex, OPA: occipital place area, PPA: parahippocampal place area.

analysis revealed that although all three scene-selective regions overlapped with the neural response to passive navigation > no navigation at a voxelwise threshold of $p < .005$, only bilateral OPA (local peak in rOPA: [42, -78, 18], $k = 60$) survived familywise-error correction using a more stringent voxelwise threshold of $p < .001$ (Fig. 4; Supplementary Table 3). Additional activations were observed in early visual cortex (including portions of early visual cortex) – presumably reflecting low-level visual differences between these conditions – as well as in a network of regions including the superior parietal lobe, superior and middle frontal gyrus, and the cerebellum (Supplementary Table 3). Intriguingly, a similar network of regions has been observed in previous studies of responses to (even the passive visual experience of) navigation (Kamps, Lall, et al., 2016; Marchette et al., 2014; Persichetti & Dilks, 2018; Spiers & Maguire, 2007). Taken together then, these results confirm that OPA shows the

strongest responses to the visual experience of navigation, relative to RSC or PPA.

4. Discussion

The current study examined functional responses in two scene navigation systems – RSC and OPA – during both the active experience and passive viewing of virtual navigation through varying maze structures. Our results reveal that RSC and OPA are involved in distinct navigational processes, with OPA supporting automatic navigation through the local spatial environment, and RSC showing greater responses to mazes that were more complex and required deliberate navigation. Two lines of evidence supported this hypothesis. First, RSC and OPA represented different information during navigation, with RSC responding more strongly than OPA to

information about the broader spatial environment (i.e., complex mazes versus simple mazes), and OPA responding more strongly than RSC to the visual experience of (even passively) moving through a maze. Second, RSC and OPA represented their respective information via different processes, with RSC responses to the broader environment depending on active navigation (i.e., diminishing during passive navigation), and with OPA responses to the visual experience of moving through the maze relatively unaffected by active versus passive navigation. These results therefore provide direct evidence for two dissociable scene navigation systems in RSC and OPA, and further support the novel hypothesis that these systems differ not only based on the information they represent, but also based on the kind of processes they support, with RSC operating relatively deliberately, and OPA operating relatively automatically.

Our study sought to test representation of the broader environment using a subtle manipulation of the spatial layout of the broader environment (i.e., using complex, 2 turn environments versus more simple, 1 turn environments), while leaving other perceptual features of the mazes matched. Given this design, one possible alternative explanation of our results is that the increase in RSC signal represented an accumulation of more navigation-related visual information during the complex mazes, rather than representation of the broader environment beyond the currently visible scene. Under this interpretation, however, it is unclear why this effect would be specific to active vs. passive navigation conditions, or why it would only be found in RSC, and not other areas that represent navigation-related visual information (e.g., OPA). A second interpretation for our findings then is that the increase in RSC signal reflects representation of the broader spatial environment beyond the currently visible view, particularly during active navigation. The idea that RSC is sensitive to the broader spatial structure of the maze during navigation is consistent with abundant evidence that RSC supports navigation through the broader environment, potentially by recognizing spatial locations (including landmarks), situating those locations relative to other places in the environment, and representing the current orientation or heading relative to both the current scene and the broader map (Auger et al., 2015; Epstein et al., 2017; Marchette et al., 2014; Persichetti & Dilks, 2019; Schinazi & Epstein, 2010; Sherrill et al., 2013; Vann et al., 2009; Wolbers & Büchel, 2005). Indeed, although the precise function of RSC is not fully established, the greater response that we observed to complex mazes versus simple mazes is potentially consistent with any of these more specific functions, since more information about different locations, relations between those locations, and heading orientations is available in complex mazes than simple mazes. Notably, almost all of the work on RSC above used relatively “active” navigational tasks, such as the judgment of relative direction task (Marchette et al., 2014; Persichetti & Dilks, 2019) or other environmental learning tasks (Auger et al., 2015), which require relatively deliberate reasoning. Here we show that the active nature of these tasks is in fact critical to spatial coding in RSC, and that such spatial coding is greatly diminished during the passive experience of navigation. Further, many of these tasks used considerably more complex and large-scale environments than the mazes employed here, in order to maximize the available information

about the broader spatial environment. By contrast, here we took a different approach, employing an extremely minimal manipulation of maze complexity (i.e., one turn versus two turns), and still found robust representation of this information in RSC, highlighting the remarkable sensitivity of this region to the broader spatial environment. Nevertheless, given that our study did not actively assess memory for the broader environment, future work will be needed to determine the precise representations encoded by RSC during deliberate navigation.

Our finding that OPA responded strongly to the passive visual experience of navigation, and did not significantly differentiate between varying navigation conditions or maze structures extends prior research implicating OPA in navigation through the local visual environment by showing that this process is relatively automatic. Indeed, most previous work on navigational processing in OPA has employed passive tasks. For example, studies showing that OPA represents first-person perspective motion (Kamps, Lall, et al., 2016, Kamps et al., 2020), navigational affordances (Bonner & Epstein, 2017), and navigational boundary information (Park & Park, 2020) all employed either no task or a relatively simple one-back task. Importantly, the hypothesis that OPA supports automatic navigation through the local visual environment does not require that OPA responses are completely encapsulated from other neural processes (i.e., such that they will never be modulated by any active task or attention manipulation); rather, our claim is that OPA responses are relatively automatic, as compared with the considerably more striking effects found in RSC.

Finally, there are two potential limitations of this work, each leaving open questions for future work. First, as discussed above, the study took a highly conservative approach toward testing representation of the broader spatial environment, manipulating complexity by comparing mazes that differed by a single turn, and manipulating active versus passive navigation while participants simply walked through a maze, making no navigational decisions and having no requirement to commit the broader spatial map to memory. These design choices have the advantage of reducing the likelihood that any differences across the conditions are related to general visual processing or overall task demands. However, one might reasonably question then how we can be sure that the task evokes representation of the broader environment at all. In the absence of obvious alternative explanations, and because our task manipulated the spatial layout of the maze environment only (i.e., comparing 1 versus 2 turns), with all other aspects of the two mazes well matched, we believe that representation of the broader environment is the most plausible and parsimonious explanation of the results observed here. Nevertheless, future work should test the hypothesis put forward here using tasks that require navigation through more complex spatial environments and explicit navigational decisions. Second, our study did not test the specificity of these effects to navigation. It is therefore possible that a similar set of results might be observed for a comparable task that does not involve navigation through the broader environment. This possibility is also unlikely, given previous work demonstrating the clear selectivity of both RSC and OPA for scenes versus other domains (e.g., faces, objects), and especially scene information relevant to navigation.

Nevertheless, future work should investigate responses to similar, non-navigational tasks to test this possibility directly.

In sum, using a maze-navigation task and fMRI, we found evidence that the adult human navigation is composed of two systems: one system (including RSC) for navigating the broader spatial environment related to the current scene, which processes information relatively deliberately, and a second system (including OPA) for navigating the local visual environment, which processes information relatively automatically. This finding helps refine our understanding of the brain's navigation system and offers further evidence to suggest functional specialization across scene-selective cortex.

Funding

This work was supported by funding from the National Institute of Mental Health (grant number R00 MH102355 to MTT) and the National Eye Institute (grant number R01 EY029724 to DDD and T32 EY7092 to FSK).

Author contributions

SS wrote the first draft of the paper; FSK, DDD, and MTT edited the paper; SS and MTT designed the research; SS performed the research; SS and FSK analyzed data.

Credit author statement

Shosuke Suzuki: Methodology; Software; Formal Analysis; Visualization; Investigation; Writing – Original Draft; Writing – Review & Editing. **Frederik S. Kamps:** Formal Analysis; Writing – Review & Editing. **Daniel D. Dilks:** Writing – Review & Editing. **Michael T. Treadway:** Conceptualization, Methodology, Writing – Review & Editing, Supervision.

Open practices

The study in this article earned Open Data and Open Materials badges for transparent practices. Materials and data from this study are available at <https://osf.io/8xvz7/>.

Declaration of competing interest

The authors report no conflicts of interest, financial or otherwise. In the past three years MTT has served as a paid consultant to Blackthorn Therapeutics and Avanir Pharmaceuticals. None of these entities supported the current work, and all views expressed herein are solely those of the authors.

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 also would like

to thank Brittany DeVries, Makiah Nuutinen, Emma Hahn, Danielle Harrison, Annabel Lu, Maryam Rehman, Jeffrey Yang, Kristi Kwok, Samuel Han, and Nimra Ahad for their assistance in data collection. This work was supported by the Simons Center for the Social Brain at MIT.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2021.03.027>.

REFERENCES

- Arkin, R. C. (1990). Integrating behavioral, perceptual, and world knowledge in reactive navigation. *Robotics and Autonomous Systems*, 6, 105–122.
- Auger, S. D., Zeidman, P., & Maguire, E. A. (2015). A central role for the retrosplenial cortex in de novo environmental learning. *Elife*, 4, Article e09031.
- Baldassano, C., Esteva, A., Fei-Fei, L., & Beck, D. M. (2016). Two distinct scene-processing networks connecting vision and memory. *Eneuro*, 3.
- Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences*, 114, 4793–4798.
- Dillon, M. R., Persichetti, A. S., Spelke, E. S., & Dilks, D. D. (2018). Places in the brain: Bridging layout and object geometry in scene-selective cortex. *Cerebral Cortex*, 28, 2365–2374.
- Epstein, R. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends in Cognitive Sciences*, 12, 388–396.
- Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598–601.
- Epstein, R., Patai, E. Z., Julian, J. B., & Spiers, H. J. (2017). The cognitive map in humans: Spatial navigation and beyond. *Nature Neuroscience*, 20, 1504.
- Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., Miller, K. L., Ugurbil, K., & Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain fMRI and fast diffusion imaging. *PLoS One*, 5, Article e15710.
- Feinberg, D. A., & Setsompop, K. (2013). Ultra-fast MRI of the human brain with simultaneous multi-slice imaging. *Journal of Magnetic Resonance*, 229, 90–100.
- Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., & Polimeni, J. R. (2013). The minimal preprocessing pipelines for the Human Connectome project. *NeuroImage*, 80, 105–124.
- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, 13, 159–166.
- Haciilihafiz, D. K., & Bartels, A. (2015). Motion responses in scene-selective regions. *NeuroImage*, 118, 438–444.
- Iaria, G., Chen, J. K., Guariglia, C., Pitto, A., & Petrides, M. (2007). Retrosplenial and hippocampal brain regions in human navigation: Complementary functional contributions to the formation and use of cognitive maps. *European Journal of Neuroscience*, 25, 890–899.
- Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, 52, 93–129.

- Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The occipital place area is causally involved in representing environmental boundaries during navigation. *Current Biology*, 26, 1104–1109.
- Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016b). The occipital place area represents the local elements of scenes. *NeuroImage*, 132, 417–424.
- Kamps, F. S., Lall, V., & Dilks, D. D. (2016a). The occipital place area represents first-person perspective motion information through scenes. *Cortex*, 83, 17–26.
- Kamps, F. S., Pincus, J. E., Radwan, S. F., Wahab, S., & Dilks, D. D. (2020). Late development of navigationally relevant motion processing in the occipital place area. *Current Biology*, 30, 544–550. e543.
- Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12, 217–230.
- Maguire, E. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology*, 42, 225–238.
- Mao, D., Molina, L. A., Bonin, V., & McNaughton, B. L. (2020). Vision and locomotion combine to drive path integration sequences in mouse retrosplenial cortex. *Current Biology*, 30, 1680–1688. e1684.
- 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, 1598.
- Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside looking in: Landmark generalization in the human navigational system. *Journal of Neuroscience*, 35, 14896–14908.
- Nasr, S., Liu, N., Devaney, K. J., Yue, X., Rajimehr, R., Ungerleider, L. G., & Tootell, R. B. (2011). Scene-selective cortical regions in human and nonhuman primates. *Journal of Neuroscience*, 31, 13771–13785.
- Park, J., & Park, S. (2020). Coding of navigational distance and functional constraint of boundaries in the human scene-selective cortex. *Journal of Neuroscience*, 40, 3621–3630.
- Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex*, 77, 155–163.
- Persichetti, A. S., & Dilks, D. D. (2018). Dissociable neural systems for recognizing places and navigating through them. *Journal of Neuroscience*, 38, 10295–10304.
- Persichetti, A. S., & Dilks, D. D. (2019). Distinct representations of spatial and categorical relationships across human scene-selective cortex. *Proceedings of the National Academy of Sciences*, 116, 21312–21317.
- Pitcher, D., Ianni, G., & Ungerleider, L. G. (2019). A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. *Scientific Reports*, 9, 1–9.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.
- Schinazi, V. R., & Epstein, R. A. (2010). Neural correlates of real-world route learning. *NeuroImage*, 53, 725–735.
- Sherrill, K. R., Erdem, U. M., Ross, R. S., Brown, T. I., Hasselmo, M. E., & Stern, C. E. (2013). Hippocampus and retrosplenial cortex combine path integration signals for successful navigation. *Journal of Neuroscience*, 33, 19304–19313.
- Silson, E. H., Chan, A. W.-Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience*, 35, 11921–11935.
- Silson, E. H., Steel, A. D., & Baker, C. I. (2016). Scene-selectivity and retinotopy in medial parietal cortex. *Frontiers in Human Neuroscience*, 10, 412.
- Spiers, H. J., & Maguire, E. A. (2007). A navigational guidance system in the human brain. *Hippocampus*, 17, 618–626.
- Suzuki, S., Lawlor, V. M., Cooper, J. A., Arulpragasam, A. R., & Treadway, M. T. (2020). Distinct regions of the striatum underlying effort, movement initiation and effort discounting. *Nature Human Behaviour*, 5, 378–388.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & Consortium, W.-M. H. (2013). The Wu-Minn human connectome project: An overview. *NeuroImage*, 80, 62–79.
- Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience*, 10, 792.
- 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. *Journal of Neuroscience*, 29, 10573–10581.
- Weisberg, S. M., Schinazi, V. R., Newcombe, N. S., Shipley, T. F., & Epstein, R. A. (2014). Variations in cognitive maps: Understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 669.
- Wen, W., Ishikawa, T., & Sato, T. (2011). Working memory in spatial knowledge acquisition: Differences in encoding processes and sense of direction. *Applied Cognitive Psychology*, 25, 654–662.
- Wolbers, T., & Büchel, C. (2005). Dissociable retrosplenial and hippocampal contributions to successful formation of survey representations. *Journal of Neuroscience*, 25, 3333–3340.
- Xu, J., Moeller, S., Auerbach, E. J., Strupp, J., Smith, S. M., Feinberg, D. A., Yacoub, E., & Ugurbil, K. (2013). Evaluation of slice accelerations using multiband echo planar imaging at 3 T. *NeuroImage*, 83, 991–1001.