



# Early Electrophysiological Markers of Navigational Affordances in Scenes

Assaf Harel<sup>1</sup>, Jeffery D. Nador<sup>1</sup>, Michael F. Bonner<sup>2</sup>, and Russell A. Epstein<sup>3</sup>

## Abstract

■ Scene perception and spatial navigation are interdependent cognitive functions, and there is increasing evidence that cortical areas that process perceptual scene properties also carry information about the potential for navigation in the environment (navigational affordances). However, the temporal stages by which visual information is transformed into navigationally relevant information are not yet known. We hypothesized that navigational affordances are encoded during perceptual processing and therefore should modulate early visually evoked ERPs, especially the scene-selective P2 component. To test this idea, we recorded ERPs from participants while they passively viewed computer-generated room scenes matched in visual complexity. By simply changing the number of doors (0 doors, 1 door, 2 doors, 3 doors), we were able to systematically vary the number of pathways that afford movement in the local

environment, while keeping the overall size and shape of the environment constant. We found that rooms with 0 doors evoked a higher P2 response than rooms with three doors, consistent with prior research reporting higher P2 amplitude to closed relative to open scenes. Moreover, we found P2 amplitude scaled linearly with the number of doors in the scenes. Navigability effects on the ERP waveform were also observed in a multivariate analysis, which showed significant decoding of the number of doors and their location at earlier time windows. Together, our results suggest that navigational affordances are represented in the early stages of scene perception. This complements research showing that the occipital place area automatically encodes the structure of navigable space and strengthens the link between scene perception and navigation. ■

## INTRODUCTION

How do we find our way in the environment? What allows us to successfully move about in the world without getting lost? Navigation, the act of finding one's way to a given destination, is a multisensory process requiring the integration of multiple types of sensory information (visual, vestibular, proprioceptive) about the environment, including information about direction, distance, and location (Ekstrom, Spiers, Bohbot, & Rosenbaum, 2018). In humans, the most prominent sensory modality guiding navigation in the environment is vision (Ekstrom, 2015). The particular advantage that vision confers to navigation is that it allows observers to recognize their surroundings remotely, even before they embark on any movement in their environment. In this regard, visual scene perception, that is, the recognition of one's surroundings, can be considered as an essential part of navigation, serving as a first vital stage in a cascade of processes, which ultimately culminate in successful navigation (Julian, Keinath, Marchette, & Epstein, 2018). Under this conceptualization, scene perception serves as the gate to navigation, demonstrating more broadly how visual processing (i.e., scene recognition) is critical for and cannot be

dissociated from action, planning, and memory systems (i.e., navigation).

Support for the link between scene recognition and spatial navigation<sup>1</sup> comes from neuroimaging. Several studies using fMRI reported that scene-selective regions sensitive to visual scene properties, such as category (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011; Walther, Caddigan, Fei-Fei, & Beck, 2009) and spatial expanse (Harel, Kravitz, & Baker, 2013; Kravitz, Peng, & Baker, 2011), also carry pertinent information for navigation (Persichetti & Dilks, 2018; Bonner & Epstein, 2017; Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2016). One region in particular that has been suggested to be involved in representing visually guided navigation information is the occipital place area (OPA: Dilks, Julian, Paunov, & Kanwisher, 2013; Hasson, Harel, Levy, & Malach, 2003; Nakamura et al., 2000). OPA activity is sensitive to various forms of navigation-relevant information, including egocentric position and heading (i.e., "viewpoint": Epstein, Higgins, Jablonski, & Feiler, 2007; Epstein, Higgins, & Thompson-Schill, 2005) "sense" (left-right: Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011), egocentric distance (proximal-distal: Persichetti & Dilks, 2016), first-person perspective motion through scenes (Kamps, Lall, & Dilks, 2016), and number of local elements in a scene, which can be used for obstacle avoidance (Kamps, Julian, et al., 2016). Importantly, OPA response patterns have also been reported to specifically

<sup>1</sup>Wright State University, Dayton, OH, <sup>2</sup>Johns Hopkins University, Baltimore, MD, <sup>3</sup>University of Pennsylvania, Philadelphia

index the spatial structure of navigational affordances (i.e., potential paths for movement in a scene), as operationalized by the number and locations of paths in the environment (Bonner & Epstein, 2017).

Although there is clear evidence that scene-selective cortical regions carry navigation-related information, it is still not clear how incoming visual information is transformed into navigation-relevant information that could potentially be used to guide movement in space. A recent TMS study demonstrated that OPA plays a causal role in transforming perceptual inputs into spatial memories linked to environmental boundaries (Julian, Ryan, Hamilton, & Epstein, 2016), but this study used a continuous theta-burst protocol that lacked the temporal resolution to probe the temporal sequence of these processes. Information about timing is essential for determining the extent to which the extraction of navigation-related information from a scene is indeed a visually driven perceptual process. One alternative, for instance, is that OPA sensitivity to navigational affordances might reflect recurrent feedback from posterior parietal cortex (Kravitz, Saleem, Baker, & Mishkin, 2011) rather than early sensory processing.<sup>2</sup> Determining the nature of scene affordances processing thus requires a method with high temporal resolution, such as magnetoencephalography (MEG) and EEG, which can establish the temporal dynamics of scene perception for navigation. Yet another advantage of EEG is that it can establish the processes involved in the extraction of navigationally relevant information: Various ERPs have been shown to index multiple cognitive processes (e.g., attentional allocation, semantic categorization, memory encoding; for a review, see the work of Luck, 2014) and thus be used to determine the mechanisms underlying a certain task or experimental manipulation.

M/EEG studies have recently started to uncover the time course of scene perception, by examining how different scene properties across a variety of complexity levels get processed over time. At a categorical level, the processing of scenes can be distinguished from the processing of other complex visual categories by 220 msec poststimulus onset (Harel, Groen, Kravitz, Deouell, & Baker, 2016; Sato et al., 1999). Specifically, the amplitude of the posterior P2 ERP component (peaking around 220 msec after stimulus onset) is higher in response to scenes than to faces and common objects. The P2 component has thus been suggested to index scene-selective processing (analogous to the face-selective N170 ERP component), particularly the processing of high-level global scene information (Harel et al., 2016, 2020; Hansen, Noesen, Nador, & Harel, 2018). Indeed, P2 amplitude not only distinguishes between scenes and other visual categories but is also sensitive to various global scene properties (GSPs), such as spatial expanse and naturalness: It is higher to closed than open scenes, can distinguish natural from man-made scenes (Harel et al., 2016, 2020; Hansen et al., 2018), and notably, it is not modulated by local texture information, in contrast to earlier visually evoked components

(Harel et al., 2020). P2 amplitude is also diagnostic of scene naturalness and spatial expanse at the level of individual scene images, with its response variance to individual scenes significantly explained by both summary image statistics (approximating naturalness and spatial expanse) and subjective behavioral ratings (Harel et al., 2016; see also the work of Cichy, Khosla, Pantazis, & Oliva, 2017). Studies using single-trial decoding approaches revealed that scene naturalness and spatial expanse (as well as basic-level scene category) can also be decoded from the neural signals even earlier, within the first 100 msec of processing (Henriksson, Mur, & Kriegeskorte, 2019; Lowe, Rajsic, Ferber, & Walther, 2018; see also the works of Groen, Ghebreab, Lamme, & Scholte, 2016; Groen, Ghebreab, Prins, Lamme, & Scholte, 2013).<sup>3</sup> The latency of both time windows suggests that low- as well as high-level diagnostic scene information is extracted at the early perceptual stages of processing, supporting rapid pre-attentive scene categorization (Hansen et al., 2018; Groen et al., 2016; Greene & Oliva, 2009; Rousselet, Joubert, & Fabre-Thorpe, 2005).

Although the above M/EEG studies cannot directly establish OPA as the generator of the spatial expanse signals indexed by the P2 component, they are nevertheless invaluable in providing a framework for thinking about the visual system's time course for extracting information about scene navigability. If the processing of ecological, navigability information occurs at the perceptual stages of processing, this should be reflected in a modulation of visually evoked activity during the first 250 msec (primarily the scene-selective P2 component) by navigationally relevant information, suggesting that scene perception and navigation are indeed intrinsically linked. Alternatively, the extraction of navigability information from the scene might reflect postperceptual processes (e.g., stimulus evaluation, decision-making, action planning), which would result in a late effect of navigability, and hence not impact the P2. One indication that the more probable alternative of the two is former rather than the latter is the ubiquity of spatial expanse—the extent to which a scene depicts an enclosed or an open space—in modulating early scene-evoked neural responses. The effect of spatial expanse can be observed during early stages of visual processing (Henriksson et al., 2019; Hansen et al., 2018), across a variety of stimulus sets and physical image properties (e.g., with both line drawings and photographs; grayscale as well as color images; with both artificial and naturalistic scene images: Harel et al., 2016, 2020; Hansen et al., 2018; Lowe et al., 2018; Cichy et al., 2017) and across various task contexts (Hansen et al., 2018; Lowe et al., 2018), indicating its centrality as a source of information for scene perception. One reason why spatial expanse may prove to be important for scene perception is because of its potential link with navigability. The spatial expanse of a scene conveys information not only about the structure and geometry of the scene but also about its function, namely, the conceivable

possibilities for movement in space. And because the spatial expanse of a scene, or rather its openness, is perceived by humans as a continuous dimension (Zhang, Houpt, & Harel, 2019), closed and open scenes can thus be thought of as two ends on an ease-of-navigation (navigability) continuum, with closed environments posing more constraints on navigation than open ones. Thus, it may be argued that the degree of openness or enclosure also confer constraints on navigability. It therefore stands to reason that early neural responses to spatial expanse index information about scene structure not only with regard to its openness, but also as it defines navigable space. Indeed, ERP amplitudes produced by spatial expanse variation are very slightly, if at all, modulated by task demands (Hansen et al., 2018), in line with the notion that navigational affordances are extracted automatically (Bonner & Epstein, 2017).

The present work sought to establish the time course of the extraction of navigational affordances and, specifically, to explicitly test the idea that early scene-evoked activity represents the potential for navigation in a scene. Although previous studies have uncovered the temporal dynamics of spatial layout processing, they did not explicitly establish the relation between these neural signatures and the extraction of navigational affordances. We argued that if scene perception involves the extraction of functional information for navigation, then the structure of navigable space (i.e., navigational affordances) should be encoded relatively early in processing and thus manifest in early visually evoked ERPs. We hypothesized that the navigational affordances of the visual environment would be resolved at the P2 time window and, specifically, that the amplitude of the P2 component would capture the ease of navigation in the environment given that the P2 distinguishes between open and closed scenes, (which can be thought of as offering more or less navigability, respectively). To test this hypothesis, we recorded ERPs from participants while they passively viewed computer-generated room scenes matched in visual complexity (used in a previous fMRI study of navigability, see the work of Bonner & Epstein, 2017). The rooms varied in the number of pathways that afford movement in the local environment: By simply changing the number of doors (0 doors, 1 door, 2 doors, 3 doors) in the room, we were able to systematically control the number of movement paths in the scene, while keeping the overall size and shape of the environment constant. If encoding navigational affordances engages stimulus-driven perceptual processes, then sensitivity to number of movement paths in the environment should emerge within the first 250 msec of processing. Specifically, neural activity during the P2 time window should be modulated by the number of doors in the scenes, with increasing P2 amplitude as a function of constraints on navigability (i.e., decreasing number of doors). Alternatively, if manipulating of number of doors in the scene does not result in early perturbations of neural activity and is observed later, then that

would lend support to the idea that encoding navigation information reflects postperceptual processing rather than stimulus-driven visual processing. Complementing the hypothesis-driven univariate ERP analysis of P2 amplitude, we also conducted a more data-driven, multivariate analysis to assess the decoding of navigability information over time, how early it emerges, and to what extent it represents the extraction of information about the local position of navigability cues, or whether that information is extracted in a position-invariant fashion.

## METHODS

### Participants

Thirty-six Wright State University students (21 women; mean age: 20 years) participated in the study. Participants signed an informed written consent form according to the guidelines of the institutional review board of Wright State University and were compensated monetarily or with course credit. All participants had normal or corrected-to-normal visual acuity and no history of neurological disease, and all but one were right-handed. Six participants were excluded from final analyses because of excessive EEG artifacts.

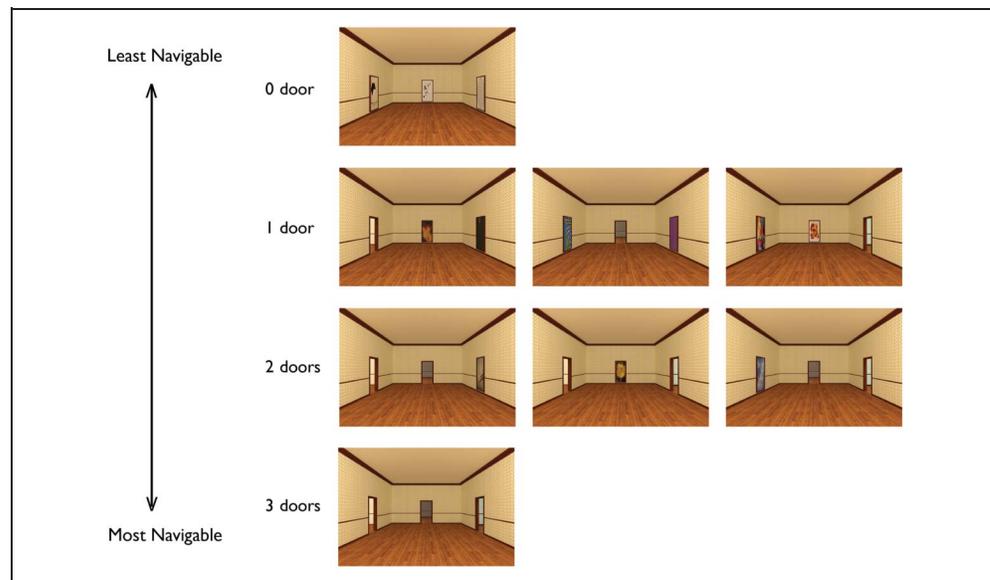
### Stimuli

The stimuli comprised 144 computer-generated images of simple rectangular rooms, used in a previous neuroimaging study (Bonner & Epstein, 2017). Every room scene contained either a door or a painting on each of its three visible walls, yielding eight navigability conditions according to the number of door elements present combined with the walls on which they appeared: 0 doors; 1 door left, right, or center; 2 doors left–center, left–right, or center–right; and 3 doors (see Figure 1 for examples of stimuli). For each navigability condition, 18 unique rooms were created by applying different textures to the walls, making a total of 144 unique exemplars. The stimuli were presented using Presentation software (Neurobehavioral Systems, Inc., [www.neurobs.com](http://www.neurobs.com)). Images were displayed in 8-bit color at the center of a Dell LCD monitor (1920 × 1080 pixels) at a viewing distance of about 150 cm, subtending 12 × 14 degrees of visual angle.

### Experimental Design and Procedure

Participants viewed the 144 individual scene stimuli 8 times (eight blocks), each block containing all 144 stimuli (a total of 1152 trials). Scene stimuli were pseudorandomized within individual blocks and across the eight blocks to prevent direct repetition of any stimulus, navigability condition, or texture. Stimuli were presented for 500 msec with a randomly jittered interstimulus interval ranging from 1000 to 2000 msec. Participants performed a fixation cross task, in which they were required to

**Figure 1.** Examples of room images used in the experiment. The rooms varied in the ease of navigability they afford, operationalized by the number of doors in the scene. By changing the number of doors in the room (0 door, 1 door, 2 doors, 3 doors), we were able to systematically control the number of movement paths in the scene, while keeping the overall size and shape of the environment constant. Depicted here are eight scenes (out of a total 144 scenes) spanning all four navigability conditions. Note that the 1-door and 2-doors conditions included three variations in door location. The rooms also varied in their spatial layout and texture, yielding 18 distinct room variants.



report whether the horizontal or vertical bar of the central fixation cross (1-degree visual angle) lengthened on each trial by a factor of 25%. Changes in the fixation cross were randomized across trials and, hence, were independent from the actual content of the underlying image, essentially requiring the participants to pay very little, if any, attention to the background images while completing this task. This same task has been employed in previous EEG studies of scene processing using naturalistic real-world stimuli (Hansen et al., 2018; Harel et al., 2016), as well as computer-generated room-like stimuli (Harel et al., 2020). We verified that the task conditions were indeed independent from the stimulus conditions by conducting a Task  $\times$  Condition ANOVA on participants' accuracy scores (see Results section).

### EEG Recording

Analog EEG signals were recorded using 64 Ag-AgCl pin-type active electrodes (Biosemi ActiveTwo) mounted on an elastic cap (Electro-Cap International, Inc.) according to the extended 10–20 system. EOGs were recorded from two additional pairs of pupil-aligned electrodes: One pair was placed on the skin over the right and left temporal zygomatic bones; the other was placed over the nasal zygomatic and frontal bones. Analog EEG data from all electrodes were referenced to the common mode signal electrode placed between electrodes PO3 and PO4. Impedance of all channels was measured before the start of each recording session, to ensure that all fell below 50 KOhms, and data for each electrode were inspected to ensure no “bridging” artifacts were present.

In postprocessing, data were rereferenced to an electrode placed on the tip of the nose. Both EEG and EOG were sampled at 512 Hz with a resolution of 24 bits with

an active input range of  $-262 \mu\text{V}$  to  $+262 \mu\text{V}$  per bit. The digitized EEG was saved and processed off-line.

### Data Processing

The data were preprocessed using Brain Vision Analyzer 2 (Brain Products GmbH). The raw data were first band-pass filtered from 0.3 to 80 Hz (24 dB), with a second-order Butterworth filter, and referenced to the tip of the nose. Eye movements were corrected in the scalp electrode data using an automated, restricted infomax ocular correction independent component analysis (for details see the work of Jung et al., 1998), effectively removing components heavily correlated with HEOG and VEOG artifacts following a meaned slope algorithm. Of the 64 independent components (one per scalp electrode) calculated for each participant,  $14 \pm 4$  were removed on average. Remaining artifacts exceeding  $\pm 100 \mu\text{V}$  in amplitude or containing a change of over  $100 \mu\text{V}$  in a period of 50 msec were rejected. The preprocessed data were then segmented into epochs ranging from  $-200$  msec before to 800 msec after stimulus onset for all conditions. Participants' data were entirely excluded from analyses if fewer than 80% of epochs could be retained following artifact removal. Six participants' data were thus excluded, and for the remaining 30, an average of  $1056 \pm 63$  (equivalent to  $91\% \pm 6\%$ ) epochs were retained for all analyses.

### ERP Univariate Analysis

The peaks of the P1, N1, and P2 were determined for each individual participant, by automatically detecting the maximal amplitude in predetermined time windows in each experimental condition (most positive peak between 80 and 130 msec, most negative peak between 130 and

200 msec, and most positive peak between 200 and 320 msec, respectively). The mean latencies of the three components across conditions were consistent with previous research (P1: 121 msec [ $SEM = 3$ ]; N1: 167 msec [ $SEM = 3$ ]; P2: 230 msec [ $SEM = 3$ ]). Analyses were restricted to posterior lateral sites (averaged across P7, P5, P9, and PO7 for the left hemisphere, and across P8, P6, P10, and PO8 for the right hemisphere), where maximal scene effects were previously observed (Harel et al., 2016, 2020; Hansen et al., 2018). Mean peak amplitudes (across participants) were analyzed using a two-way within-subject ANOVA, with Hemisphere (left, right) and Navigability (number of doors) as independent factors. Because numbers of doors varied parametrically, we have also conducted a linear trend analysis to assess whether the amplitude of the ERP components scales with decreasing number of doors, implemented as a two-way ANOVA framework (for further details on this analysis, see the work of Pinhas, Tzelgov, & Ganor-Stern, 2012).

### ERP Multivariate Analysis

Representational similarity analyses (RSA) were conducted on the segmented EEG data, across all 64 channels (but not external electrodes), for all artifact-free segments. Data were exported from Brain Vision Analyzer and processed in MATLAB (Version 2016a, MathWorks) as four-dimensional matrices (Channel  $\times$  Condition  $\times$  Segment  $\times$  Time Point). Subject-level ERP data were submitted to RSAs against two-model representational dissimilarity matrices (RDMs), pertaining to the two-alternative hypotheses: location-specific processing, in which the spatial layout of the doors influenced processing (Figure 4A, left column); and location-invariant processing, in which only the “global” differences (i.e., between the total number of doors) influenced processing (Figure 4A, right column). Experimental conditions were coded as binary triplets, with ones and zeros corresponding to doors and paintings, respectively. Each digit within a triplet corresponded to one of the three possible locations (left, right, or center). So, for example, condition 1–0–0 would correspond to a single door on the left, whereas 0–0–1 would correspond to a single door on the right. Thus, the identity and location of each door or painting was retained in these condition codes. Codes were then entered in ascending order as the rows and columns of a confusion matrix to construct both model and neural RDMs. As such, each cell corresponded to one pairing of stimulus conditions. For the location-specific model RDM, values in each cell were derived from the number and location of doors (and paintings) in its component row and column stimulus conditions (Bonner & Epstein, 2017). Mathematically, these values were obtained by subtracting the similarity between stimulus conditions from the maximum possible similarity. Similarity was computed as the sum of the dot product of the condition codes corresponding to the row and column of the cell, yielding a value between 0

(maximally dissimilar) and 3 (maximally similar). Differences between conditions in a given cell could therefore be denoted by the difference between maximal similarity and the calculated similarity. Meanwhile, cell values in the location-invariant model RDM were derived only from the total number of doors (and paintings) present in each pair of conditions and not their locations. Mathematically, they were coded as the difference between sums of the row and column stimulus conditions. Because the neural RDMs were constructed using Pearson correlation coefficients of determination ranging from 0 to 1 (see below), we divided the cell values in the model RDMs (ranging from 0 to 3) by 3 to achieve common metrics for both. The resulting model RDMs are presented in Figure 4.

Neural RDMs were constructed for each subject by first calculating the Pearson correlation coefficient of determination between the ERP amplitudes of all possible pairs of conditions across electrodes, separately at each time point. As these correlation coefficients denote pairwise similarities between conditions, differences between conditions were represented by taking the complement of the coefficient of determination ( $1 - r^2$ ). In order to determine the level of agreement between the neural and model RDMs for each subject, we took the Spearman correlation between the two matrices (effectively comparing the obtained pattern of activation across electrodes in the neural RDMs to those predicted by each model RDM). Finally, the Spearman correlation coefficients were submitted to cluster analyses (with a cluster induction parameter corresponding to a Type I error rate of .01) across subjects to correct for multiple comparisons (Benjamini & Hochberg, 2000).

## RESULTS

### Behavioral Performance: Verification of Task Independence

We conducted a Task (horizontal change in fixation cross, vertical change in fixation cross)  $\times$  Stimulus (8 different door conditions, see above) ANOVA on participants’ accuracy scores to verify that the orthogonal fixation task was indeed independent of stimulus conditions. We found no significant main effects of either factor (Task:  $F(1, 29) = 2.19, p = .15, \eta_p^2 = .07$ ; Stimulus:  $F(7, 203) = 1.85, p = .08, \eta_p^2 = .06$ ), and notably, there was no interaction between them,  $F(7, 203) = 1.02, p = .42, \eta_p^2 = .03$ . Overall, this supports the orthogonality of the fixation cross task to the stimulus (i.e., door) conditions.

### Univariate Analysis

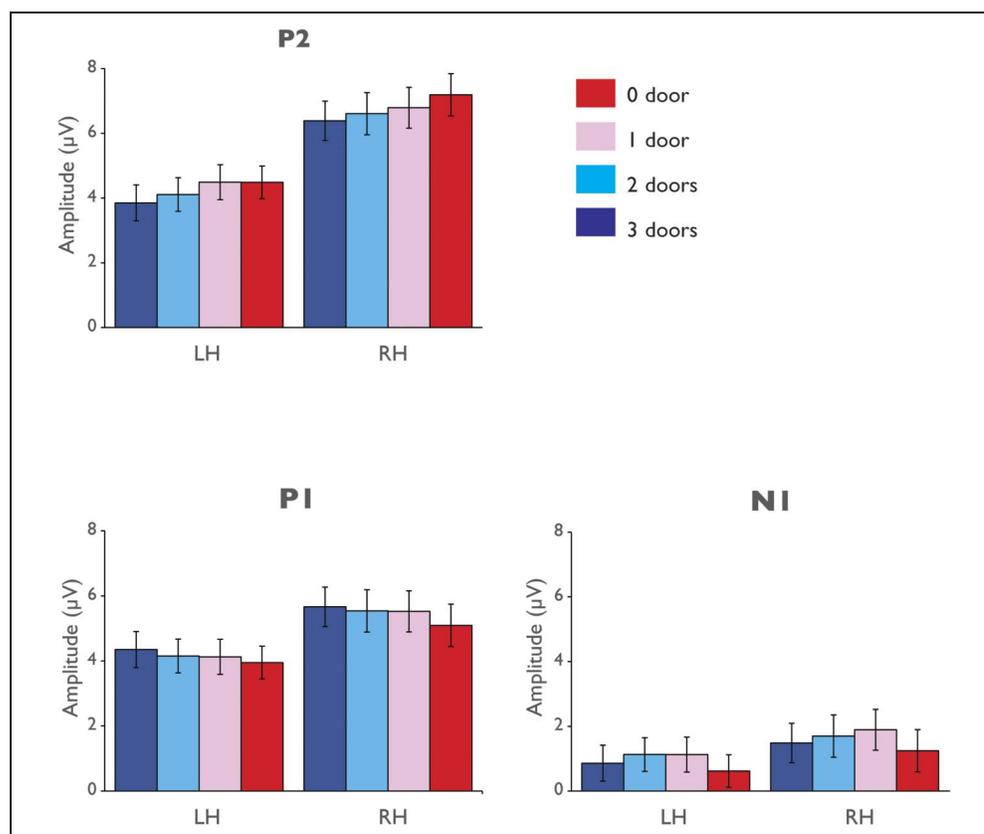
To examine the extent to which navigational affordances, operationalized as the number of doors in a room (i.e., the number of pathways that afford movement in the local environment), are encoded by early neuromarkers of scene perception, we compared the amplitude of the early

visually evoked ERP components (P1, N1, and P2) in response to the different room conditions (0 doors, 1 door, 2 doors, 3 doors). We conducted a two-way repeated-measures ANOVA on the amplitude of the individually defined peaks of each of the ERP components, with Hemisphere (left, right), and Number of Doors (0 doors, 1 door, 2 doors, 3 doors) as independent factors. The significant results of these analyses are reported in Figure 2, and the grand-average waveforms are depicted in Figure 3.

### P2 Component

We found that the P2 amplitude is sensitive to the number of doors contained in the room scenes, expressed by a significant main effect of Number of Doors,  $F(3, 87) = 4.29$ ,  $p = .007$ ,  $\eta_p^2 = .13$ . To assess the source of the main effect of Number of Doors, we have conducted two follow-up analyses: linear trend analysis and pairwise post hoc comparisons. The former analysis showed that P2 amplitude scales with the number of doors present in the scene, manifesting in a significant linear trend,  $F(1, 29) = 7.58$ ,  $p = .01$ ,  $\eta_p^2 = .21$ . Post hoc comparisons revealed a significantly higher P2 amplitude for the 0-doors condition ( $M = 5.83$  mV,  $SEM = 0.51$ ) compared to the 3-doors condition ( $M = 5.12$  mV,  $SEM = 0.42$ ),  $t(29) = 2.45$ ,  $p = .009$ , and a higher amplitude to the 1-door ( $M = 5.64$  mV,  $SEM = 0.47$ ) relative to the 2-door condition ( $M = 5.36$  mV,  $SEM = 0.45$ ),  $t(29) = 2.09$ ,  $p = .02$ .

**Figure 2.** Grand average ERP univariate analysis results. Top row: mean P2 peak amplitudes in response to the four navigability conditions (from least navigable to most navigable: 0 door, 1 door, 2 doors, 3 doors) presented separately for the left and right hemisphere. Bottom row: mean P1 and N1 peak amplitudes for the navigability conditions presented separately for the left and right hemispheres. Error bars indicate between-subjects  $SE$ ; all data are plotted for the posterior lateral sites. P2 amplitude showed a significant linear increase in magnitude of response as a function of number of doors, with less doors resulting in higher amplitude. Notably, no such increase was observed on the amplitude of the P1 or N1 components (see text for details).

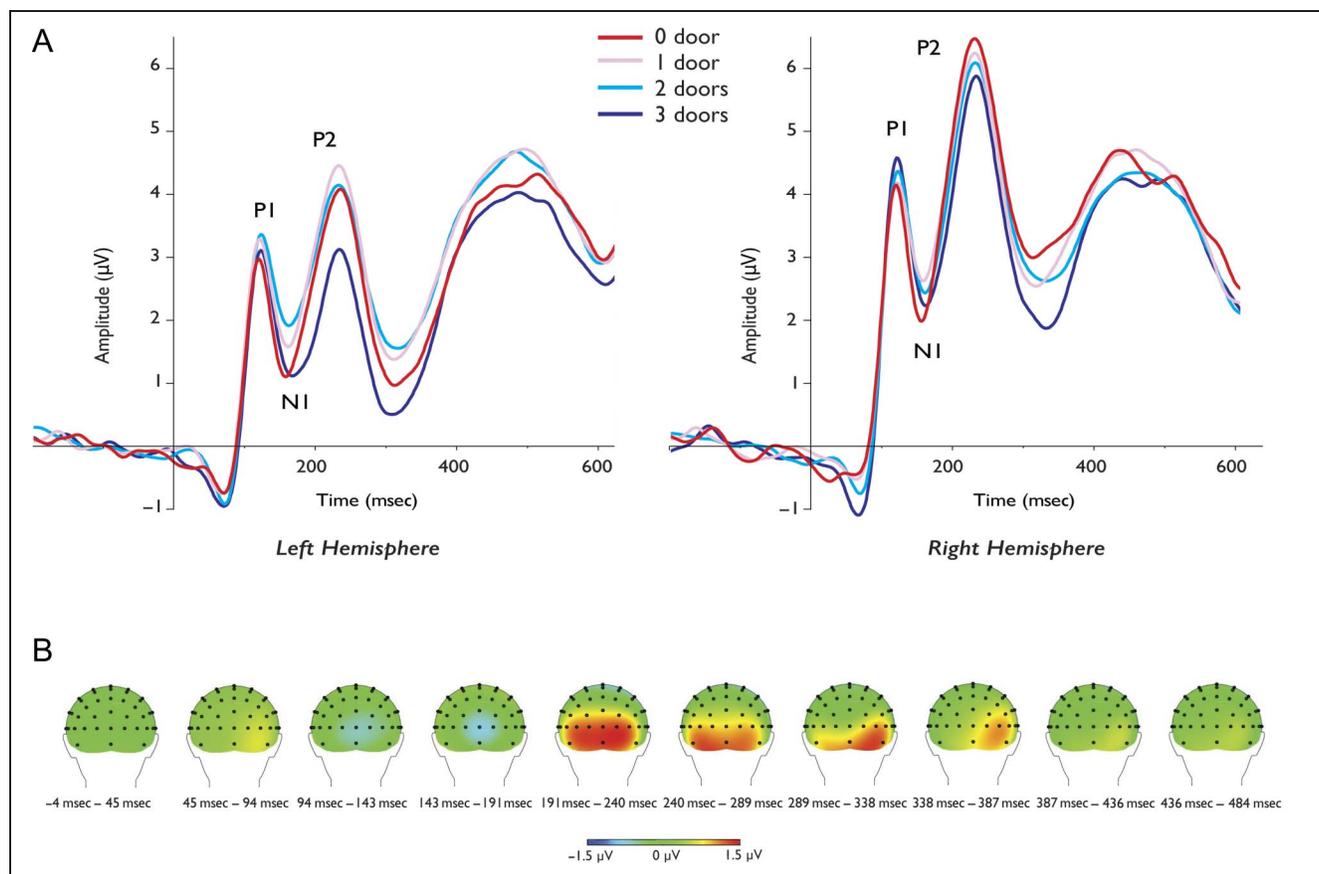


A main effect of Hemisphere was also observed,  $F(1, 29) = 18.90$ ,  $p = .001$ ,  $\eta_p^2 = .40$ , with higher amplitude in the right hemisphere ( $M = 6.74$  mV,  $SEM = 0.62$ ) compared with the left hemisphere ( $M = 4.23$ ,  $SEM = 0.42$ ). The effect of Number of Doors, however, was not found to significantly differ across hemispheres (interaction of Hemisphere with Number of Doors,  $F(3, 87) = 1.36$ ,  $p = .26$ ,  $\eta_p^2 = .05$ ; as standard ERP practice, we depict both hemispheres in our figures).

Notably, the navigational affordances effect persisted beyond the P2 time window. Figure 3B depicts the difference ERP waveform contrasting the 3-doors and the 0-doors conditions across the whole scalp using current source density (CSD) topographical maps. As can be seen, the signal depicting difference in perceiving minimally navigable scenes to maximally navigable scenes was present from around 200 msec to 350 msec post-stimulus onset.

### N1 Component

An analysis of the N1 component revealed the effect of Number of Doors did not reach significance,  $F(3, 87) = 3.02$ ,  $p = .06$ ,  $\eta_p^2 = .09$ . No significant linear trend was noted as a function of number of doors,  $F(1, 29) < 1.00$ ,  $\eta_p^2 = .02$ , and the difference between the no-doors condition and the three doors condition was not found to be significant (planned  $t$  test:  $t(29) = 0.85$ ,  $p = .20$ ). No significant effects of hemisphere,  $F(1, 29) = 1.30$ ,  $p =$



**Figure 3.** (A) Group-averaged waveforms for the four navigability conditions (from least navigable to most navigable: 0 door, 1 door, 2 doors, 3 doors) presented separately for the left and right hemisphere at posterior lateral sites. (B) Scalp CSD topographical maps depicting the difference between the 3-doors and the 0-door conditions show a persistent effect of navigability.

.26,  $\eta_p^2 = .04$ , or a Hemisphere  $\times$  Number of Doors interaction,  $F(3, 87) < 1.00$ ,  $\eta_p^2 = .01$ , were observed.

### P1 Component

Analysis of the amplitude of the P1 component showed the effects of Number of Doors, and the Number of Doors  $\times$  Hemisphere did not reach significance,  $F(3, 87) = 2.56$ ,  $p = .08$ ,  $\eta_p^2 = .08$ ;  $F(3, 87) = 1.03$ ,  $p = .37$ ,  $\eta_p^2 = .03$ , respectively. No significant main effect of Hemisphere was observed,  $F(1, 29) = 3.94$ ,  $p = .06$ ,  $\eta_p^2 = .12$ .

We also performed a secondary analysis investigating the extent to which the specific location of the door, which might provide local information about potential movement path, could have an effect on the early visually evoked ERPs.<sup>4</sup> We conducted two separate two-way ANOVAs with Hemisphere and Door Location as independent variables: one ANOVA for the single-door condition and another one for the two-doors conditions. For the single-door condition, we were not able to find any significant effects of Door Location or Door Location  $\times$  Hemisphere on either the P1, N1, or P2 peak amplitudes (all  $ps > .30$ ). A significant effect of Hemisphere was found on the amplitude of the P2 and P1 components (P2:  $F(1,$

29) = 18.52,  $p = .001$ ,  $\eta_p^2 = .40$ ; P1:  $F(1, 29) = 4.58$ ,  $p = .04$ ,  $\eta_p^2 = .13$ ), with higher amplitude in the right than in the left hemispheres.

For the two-doors condition, a significant main effect of Door Location was found on the amplitude of the P2 component,  $F(2, 58) = 3.70$ ,  $p = .04$ ,  $\eta_p^2 = .11$ , and the interaction between Door Location and Hemisphere was not found to be significant,  $F(2, 58) < 1.00$ ,  $\eta_p^2 = .01$ . Post hoc comparisons showed significantly ( $p < .05$ , Bonferroni-corrected) lower amplitude to rooms in which the doors were located in the center and right positions compared to scenes with doors on the left and right, albeit the latter were not significantly different than scenes with left- and center-positioned doors.<sup>5</sup> No significant effects of Door Location or Door Location  $\times$  Hemisphere were observed on the N1 amplitude (all  $ps > .78$ ). For the P1 component, we found a significant main effect of Door Location,  $F(2, 58) = 6.00$ ,  $p = .006$ ,  $\eta_p^2 = .17$ , whereas the interaction between Door Location and Hemisphere was not found to be significant,  $F(2, 58) < 1.00$ ,  $\eta_p^2 = .00$ . Lastly, a main effect of Hemisphere was found for both the P1 and P2 components (P1:  $F(1, 29) = 4.21$ ,  $p = .05$ ,  $\eta_p^2 = .13$ ; P2:  $F(1, 29) = 17.48$ ,  $p = .001$ ,  $\eta_p^2 = .37$ ), reflecting a right hemisphere advantage (see main analysis reported above).

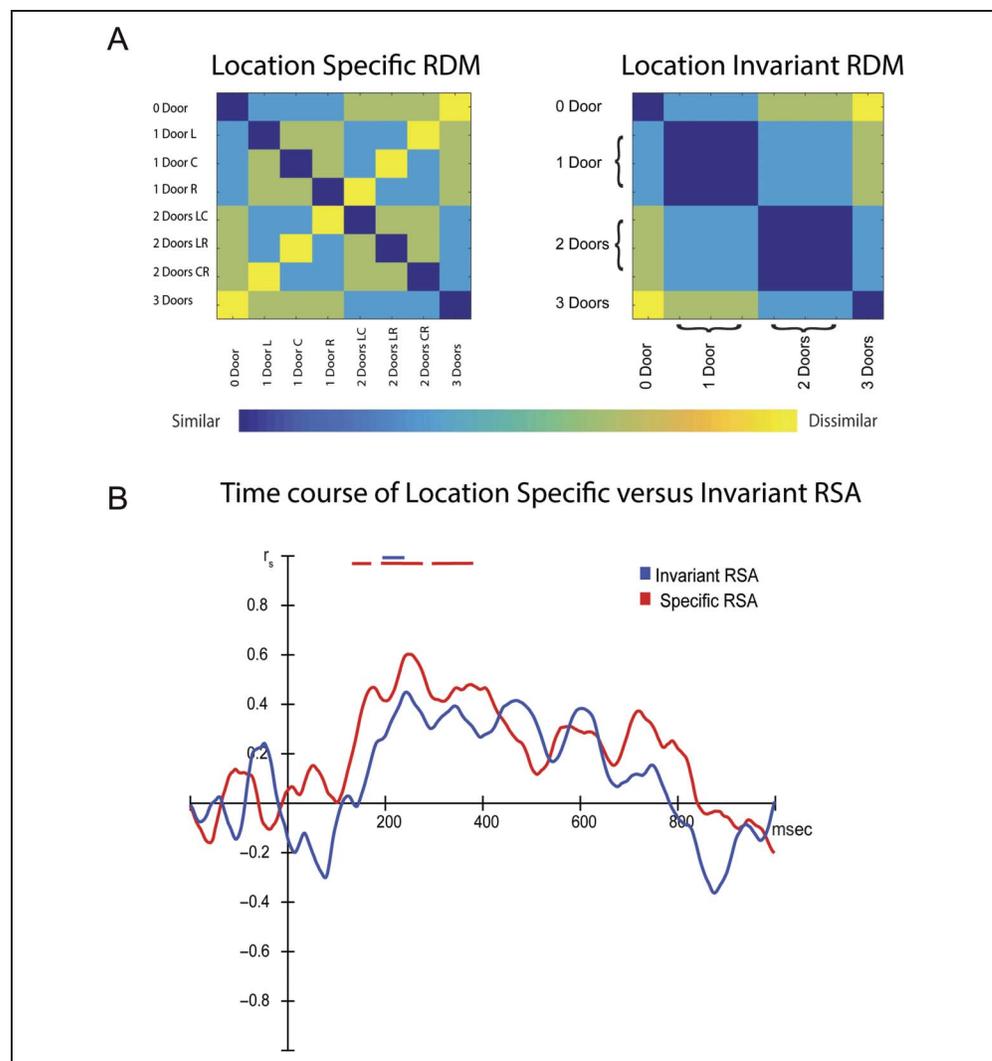
In summary, we did not observe consistent effects of the specific location of the door/s in the scene. This may stem from our univariate approach lacking the sensitivity to detect what might be subtle differences in the visual input (for similar results in fMRI, see the work of Bonner & Epstein, 2017). To address this possibility, we conducted a multivariate analysis, which allowed us to examine (a) the extent to which both position-dependent and position-invariant information can be extracted from the EEG signal, and (2) the different time windows during which the signals may be observed.

### Multivariate Analysis

To determine the time course of location-specific and location-invariant navigability-related processing, RSAs were conducted by comparing the obtained neural RDMs from each participant at each time point with two-model RDMs, one corresponding to each hypothesis (see Methods section and Figure 4A). The pattern of “activation” across electrodes within their averaged ERP

waveforms was correlated with differences between conditions predicted by either or both of the two-model RDM. This was followed by a cluster analysis across subjects, allowing us to determine the time intervals in which the data were best explained by one of the predicted models. We found significant Spearman correlations (Figure 4B) between the neural RDMs and the location-specific RDM at several time windows: from 134–170 msec (cluster-corrected values of  $p = .035$ , two-tailed), 193–275 msec (cluster-corrected values of  $p = .0024$ , two-tailed), and 295–380 msec (equivalent to all cluster-corrected values of  $p = .004$ , two-tailed). Notably, the cluster analyses also uncovered significant Spearman correlations between the neural RDMs and location-invariant RDM, from 196 to 237 msec (Figure 4B). Interestingly, this time window corresponds with the P2 time window. Together, the multivariate analyses suggest that local featural differences between stimuli (i.e., the number and location of doors) in each navigability condition are processed as early as 134 msec after stimulus onset, and more global featural differences (0 doors, 1 door, 2 doors, 3 doors)

**Figure 4.** Results of the multivariate analysis. (A) The two-model RDM used for the analyses, pertaining to the two-alternative hypotheses: location-specific processing, in which the local featural differences between conditions (i.e., between doors and paintings) influenced processing (left column); and location-invariant processing, in which only the “global” differences (i.e., between the total number of doors) influenced processing (right column). (B) Time course of decoding (measured using Spearman correlation coefficients), comparing the obtained pattern of activation to that predicted by the two models: location-specific (blue) and location-invariant (red). Solid lines indicate significant decoding ( $p < .01$ , corrected; see text for details).



are most likely processed after local featural differences, no earlier than 196 msec following stimulus onset.<sup>6</sup>

## DISCUSSION

The current study provides novel evidence that the brain codes information about the potential for navigation in the scene as early as 200 msec after stimulus onset, and that this coding involves both a global, position-invariant signal about the overall navigability of the space and local information regarding the positions of navigational pathways. A standard univariate ERP analysis revealed that the amplitude of the scene-selective P2 ERP component was higher in response to images of rooms with no doors compared to rooms with three doors, analogous to the higher P2 amplitude in response to closed relative to open scenes reported previously (Harel et al., 2016, 2020; Hansen et al., 2018). Furthermore, P2 peak amplitude scaled linearly with the constraints on navigation: The more constraints on navigation (i.e., less doors in a room), the higher was its amplitude. And although the effect of navigability was most pronounced on the P2 ERP component, the difference in amplitude between no-doors and three-doors condition continued beyond the P2 time window, lasting for additional 200 msec. Complementing these findings, a multivariate analysis revealed that the P2 time window contains significant information about navigational affordances: It is the first time period in which position-invariant navigability information is extracted, in addition to information regarding the location of the diagnostic feature (a specific door), which emerges earlier and persists at this time window as well. Together, these findings suggest that diagnostic information about the potential for navigation in a scene is present at the early stages of visual processing and extracted as early as 220 msec after stimulus onset.

Based on the current findings, we suggest that perceiving visual environments and navigating through them should not necessarily be considered as two separate processes, but rather as two points on a single continuum, scene perception being the first step in a sequence of stages that support navigation. At the neural level, this outlook has two spatiotemporal corollaries. First, scene-selective cortex should not only be engaged in the extraction of scene-diagnostic features but should also carry information about the potential for navigation in the scene. Second, navigability-related neural activity should be observed in the early stages of visual scene processing. The current ERP study joins the original fMRI study, which used the current scene stimuli (Bonner & Epstein, 2017) to establish these two points. Bonner and Epstein (2017) showed that information about the potential for movement in the scenes is, indeed, represented in scene-selective OPA.<sup>7</sup> Our study shows that this same information is represented as early as 200 msec poststimulus onset, the same latency during which global properties of the scene are extracted (Harel et al., 2016, 2020; Hansen

et al., 2018). Because the two studies use the exact scene stimuli, they form a crucial link in connecting the spatial and temporal aspects of perceptual processing of the potential for action in scenes. Furthermore, in a follow-up study, Bonner and Epstein (2018) reanalyzed their imaging data and showed using deep convolutional networks that the affordance properties of scenes could be represented through just a few stages of purely feedforward hierarchical computations, implying that computations of navigational affordances in OPA could be achieved rapidly, in line with the current findings.

Given the limited temporal precision of fMRI, the functional nature of the observed OPA activation cannot be determined unequivocally; whereas OPA activity could indeed reflect stimulus-driven processing of navigability information available in the scene, it could also potentially reflect recurrent feedback from posterior parietal cortex as part of the occipito-parietal circuit (Kravitz, Saleem, et al., 2011). Our current results suggest the former, rather than the latter, alternative is more probable, as we show that by 200 msec, sufficient evidence has accumulated for determining the potential for navigation in the scene. Moreover, the fact that the navigability effect on the EEG manifests without any apparent need for encoding or planning movement (see below) strengthens the notion that navigational affordances are encoded mandatorily, as originally suggested by Bonner and Epstein (2017). It is important to note, however, that despite the similarities between the studies, we cannot unequivocally conclude that OPA is the neural generator of the observed effects of navigational affordances on the P2 component. We have not performed source localization in the current study, as the relationship between ERP generator locations and scalp electrodes is complex (Nunez & Srinivasan, 2006), and oftentimes the answers to which cortical area generates a certain ERP effect could vary as a function of the mathematical solution favored by the researcher (for a comprehensive discussion, see the work of Luck, 2014, Online Chapter 14; for an empirical demonstration of the limits of localization, see the work of Petrov, 2012). Future research combining ERP and fMRI (e.g., simultaneously recording ERPs in an MRI scanner; see the work of Sadeh, Podlipsky, Zhdanov, & Yovel, 2010) is needed to determine the relationship between OPA activity and the P2 sensitivity to navigationally relevant information. Although the current data suggest that navigational affordances are encoded rapidly, it is still an open question just how rapid “rapid” is. Some recent studies show earlier encoding of navigable space, specifically, around 100–120 msec poststimulus onset. A study combining MEG and fMRI showed significant scene boundary encoding (i.e., sensitivity to navigation-constraining large-scale geometrical boundaries) in OPA with corresponding MEG response patterns emerging as early as 65 msec and peaking at about 100 msec poststimulus onset (Henriksson et al., 2019). In a similar vein, differential processing of closed and open scenes (which

confer distinct navigational affordances, see below) has been reported to manifest not only at the P2 time window but also earlier, around 120 msec poststimulus onset (Lowe et al., 2018). Findings from our multivariate analyses support these studies demonstrating early coding of spatial geometry. Information about the direction of potential movement (i.e., specific door position) was found to be represented as early as 140 msec poststimulus onset, suggesting early encoding of navigational affordances. Notably, position-invariant information, which is perhaps more related to spatial expanse was found later, at the P2 time window, between 200 and 240 msec. Consequently, we suggest that spatial affordances are picked up at different points in times, each characterized by separate neural representations: Location-specific affordance information is encoded as early as 134 msec, persists until 170 msec, and re-emerges around 200 msec. This is followed by more global scene representations that are affected by the affordances of the space, namely, when there are more affordances (i.e., less constraints on movement overall), the global space becomes larger and more “open.” This second stage reflects global processing of the scene, entailing the integration of features across the entire scene (for discussion, see the work of Harel et al., 2020). As we note above, this later time window is coincident with the univariate results highlighting the sensitivity of the P2 component to navigation-related information (although the multivariate analyses detect the onset of this effect slightly earlier than the peak analyses, this is to be expected, given that the P2 only peaks around 230 msec).

Our focus on the P2 component, suggesting it is the electrophysiological marker of navigational affordances, follows our previous work highlighting the P2 as an index of the processing of high-level global scene information. The posterior P2 is the first ERP component to show evidence of scene-selectivity, with higher amplitude to scenes compared with faces and objects (Harel et al., 2016), and the only visually evoked component to be modulated by scene inversion, as would be expected if global scene information is indeed extracted during this period (Harel & Al Zoubi, 2019). Furthermore, P2 amplitude is sensitive to GSPs, such as spatial expanse (closed/open) and naturalness (man-made/natural), and these effects are automatic, evident across a variety of stimulus presentation conditions, and are largely unperturbed by manipulations of local texture (Harel et al., 2016, 2020; Hansen et al., 2018). Together, these studies suggest that the scene-selective P2 and the P2 time window in general (approx. 200–250 msec) is indicative of and essential for the processing of the global spatial structure of scenes (see also the work of Kaiser, Häberle, & Cichy, 2020; Kaiser, Turini, & Cichy, 2019; Cichy et al., 2017). Spatial structure is used here broadly as an umbrella term to capture related concepts, such as scene layout, expanse, and boundary. Notably, spatial structure is one of the key categories of GSPs proposed by Greene and Oliva (2009) to be central for rapid scene categorization. According to

Greene and Oliva, rapid scene categorization is not primarily mediated through objects and parts, but rather through global, ecological properties that describe spatial and functional aspects of scene space (such as navigability or mean depth). Based on Greene and Oliva’s theory, we hypothesized that closed and open scenes not only describe the large-scale geometry of the scenes (i.e., its spatial structure) but may also capture the functional property of navigability. As such, closed and open scenes represent two ends of a continuum wherein enclosed spaces posit more constraints on movement than open spaces. And because closed scenes consistently evoke a stronger neural response than open scenes (Harel et al., 2020; Hansen et al., 2018; Lowe et al., 2018), we predicted that scenes conferring more constraints on navigability should evoke a higher P2 amplitude than scenes that confer less such constraints. This prediction bore out, not only in its dichotomous form (0-doors vs. 3-doors) but also as a continuous, parametric effect of navigability, evident in a linear decrease of P2 amplitude as a function of the number of doors, as well as in a significant decoding of number of doors independent of their local location around the P2 time window, 200–240 msec poststimulus onset. It is still an open question to what extent the P2 exclusively indexes global scene information, or whether it also incorporates the processing of local scene information. Specifically, in spite of the studies described above, we still found some modulation of the P2 amplitude by door location (at least for the two-door condition, see secondary univariate analyses above), which may suggest it is not entirely independent of local diagnostic information. The link, however, between P2 amplitude and local information is not straightforward, as no modulation was found for the single door condition, and the modulation of the two-door condition was not systematic and, furthermore, was not related to hemisphere, as one might have expected given a retinotopic mapping of the visual field. Future research will be needed to elucidate the relationship between local and global scene processing and how these processes map onto the P2 amplitude.

Our study reveals another interesting characteristic of the P2. We found that the P2 navigability effect was lateralized to the right hemisphere; that is, the effect of the number of doors on P2 amplitude was more pronounced in the right than in the left hemisphere. This interaction resonates and converges with previous ERP studies of P2, which report laterality effects on the P2 amplitude, with overall higher amplitude in the right hemisphere, as well as specific GSPs being more discriminable in the right hemisphere. Specifically, both scene naturalness and spatial expanse were reported to have a greater effect in the right hemisphere (Harel et al., 2016, 2020; Hansen et al., 2018). The finding that the right hemisphere is more involved than the left in the processing of global scene information extends previous research on local/global processing and hemispheric asymmetries and is in line with the long-lived proposal that the right hemisphere

specializes in global processing (Brederoo, Nieuwenstein, Lorist, & Cornelissen, 2017; Flevaris & Robertson, 2016; but see the work of Wiesmann, Friederici, Singer, & Steinbeis, 2020). Thus, the right hemisphere advantage for GSP effects further supports our proposal that P2 indexes the processing of the spatial structure of the scene via the processing of global scene information.

Our findings point to the putative connection between spatial structure and navigability, which, arguably, may serve as the mechanism by which perceptual information is transformed into action-relevant information. Given the close link between spatial structure and navigability, our findings raise two intriguing questions. First, are navigability and spatial structure (or spatial expanse, in the case of closed vs. open scenes) one and the same thing? Can the two constructs be used synonymously, or are they two independent dimensions? Second, given that they are independent, does the neural encoding of navigational affordances reflect intermediate levels of representation (i.e., GSPs), low-level image statistics (Groen, Silson, & Baker, 2017) or rather the extraction of higher-level ecological scene properties (for discussion, see the work of Malcolm, Groen, & Baker, 2016)? Our current design cannot directly address these questions, as we only used closed indoor scenes (rooms). However, future research will be able to shed further light on this issue. To test the independence of the two dimensions, one would have to vary the amount of movement a scene affords, in both closed and open scenes. If spatial expanse and navigability are independent, then varying navigability should have a similar effect on the neural responses to open scenes (this could be manipulated, e.g., by adding an increasing number of obstacles like boulders, or by varying the number of paths in an open field). In addition, an alternative, more naturalistic approach could use a large set of real-world scene images (instead of highly controlled artificial scenes), in which people would rank these scenes on both spatial expanse (Zhang et al., 2019) and navigability (Bonner & Epstein, 2017, Experiment 2) and then in a separate electrophysiological study integrate these rankings with the neural response patterns to determine the relative contribution of each dimension and other image properties to the neural representations using computational model-based approaches (e.g., Lescroart & Gallant, 2019; Bonner & Epstein, 2018; Cichy, Kholsa, Pantazis, & Oliva, 2017; Lescroart, Stansbury, & Gallant, 2015).

At a broader theoretical level, our results support a tight link between perception and action, a hallmark of sensorimotor and embodied cognition theories (e.g., Jelić, Tieri, De Matteis, Babiloni, & Vecchiato, 2016; Wilson, 2002; Clark, 1999). One aspect of these theories in the context of navigation is the constant need for visual updating when one explores an environment to minimize prediction error (Kaplan & Friston, 2018; Hassabis & Maguire, 2009; Kurby & Zacks, 2008; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). The idea is that exploring the environment requires continuously modeling the potential outcomes of the

intended action, with affordances serving this function by constraining visual perception to reflect experience-dependent, observer-relevant information (Sestito, Flach, & Harel, 2018). At the neural level, this should translate to dynamic changes in continuous activity in early visual areas as one walks around and explores their surroundings, even before execution of action (e.g., movement; Jelić et al., 2016). In line with this idea, a recent study using mobile brain/body imaging technology in which participants actively walked in a highly engaging immersive virtual environment demonstrated that environmental affordances are extracted and encoded throughout the entirety of the act of exploring one's surroundings, starting from early perceptual stages (P1-N1 complex) all the way to motor planning and execution (Djebbara, Fich, Petrini, & Gramann, 2019). Notably, our study adds several key observations to these findings. First, ambulatory, active, and continuous exploration of the environment is not necessary for observing electrophysiological responses representing the extraction of navigational affordances. Participants in our study were stationary, sitting in front of a computer screen, and watched briefly presented, minimalistic room images. The fact that one can observe similar electrophysiological responses as a function of navigability even without any movement or the presence of an immersive environment implies that scene affordances are extracted across multiple contexts and task demands (for a similar finding with spatial expanse, see the work of Hansen et al., 2018). Arguably, a lifetime of experience navigating in the world results in automatic activation of sensorimotor scene representations when presented with visual environments, even if these environments are sparse, minimalistic scenes deprived of rich detail (for the role of experience in perceiving novel scene affordances, see the work of Sestito, Harel, Nador, & Flach, 2018). The idea that navigational affordances are processed mandatorily is also supported by the finding that navigability effects are evident even without any explicit task-context or relevant task demands, as participants in our study were not required to either move about the environment, imagine themselves moving about it, or make any explicit judgments regarding its potential for navigation. The fact that we found navigability-based modulations although no movement in space was performed, nor was movement in space directly relevant to the task, supports the conclusion that extracting navigational affordances is rapid, mandatory, and task-independent. Furthermore, the similar patterns of results between the two studies suggest that our current laboratory-based findings are likely to generalize to real-world, realistic settings and thus expand their validity and utility for future research.

In summary, this study demonstrates that navigational affordances are extracted at the early, perceptual stages of visual scene processing, suggesting a close link between scene perception and navigation. Information about the potential for navigation in the scene is extracted rapidly and automatically, without any explicit task or movement

requirements. Complementing prior neuroimaging studies showing OPA encodes the structure of navigable space, the current work establishes the temporal dynamics of processing navigational affordances. Significant navigability information is present in two time windows: an early time window sensitive to the specific position of navigability diagnostic stimulus features, and a later one, which incorporates both position-specific and position-invariant information. The later time window overlaps with the univariate P2 ERP component, reflecting global processing of scene structure. Finally, the current findings are in line with sensorimotor accounts of perception, suggesting that perceiving visual environments and navigating through them should not necessarily be considered as two separate processes, but rather as two integrated processes.

Reprint requests should be sent to Assaf Harel, Department of Psychology, Wright State University, 335 Fawcett Hall, 3640 Col. Glenn Highway, Dayton, Ohio 45435, or via e-mail: [assaf.harel@wright.edu](mailto:assaf.harel@wright.edu).

### Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience (JoCN)* during this period were  $M(\text{an})/M = .407$ ,  $W(\text{oman})/M = .32$ ,  $M/W = .115$ , and  $W/W = .159$ , the comparable proportions for the articles that these authorship teams cited were  $M/M = .549$ ,  $W/M = .257$ ,  $M/W = .109$ , and  $W/W = .085$  (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

### Notes

1. The link between scene recognition and navigation resonates the putative link between perception and action proposed by sensorimotor accounts of perception, which posit that the potential for action in the environment, also known as affordances, is conveyed by the visual stimulus itself (e.g., Gibson, 1979).
2. This is in fact an inherent limitation of fMRI because of its low temporal resolution. For a more general discussion on the inferential challenges of fMRI, see the work of Ghuman and Martin (2019).
3. The extent to which these early signatures reflect the extraction of local image statistics or more global scene properties is still an open question. For instance, scene inversion effects are only observed around 220–250 msec poststimulus onset (Kaiser, Häberle, & Cichy, 2020; Harel & Al Zoubi, 2019), consistent with the idea that global scene structure information is extracted later than 100 msec.

4. The reason this analysis was considered secondary is that we did not expect to find door location effects in the univariate analysis based on Bonner and Epstein's (2017) fMRI study. No robust effects of door location were observed in that study in the univariate response magnitude analysis, while multivariate analysis of response patterns did not show sufficient sensitivity to report such effects.

5. Note that whereas in the 1-door condition, the comparison between conditions is relatively straightforward with door location being the diagnostic cue for movement, in the 2-doors conditions, this is less obvious, as the "odd one out" is not the door, but rather its absence—a single painting. This difficulty in interpretation is further exacerbated in the case of no significant interactions with Hemisphere, meaning that painting location is not the consistent source of the effect.

6. It should be noted, however, that significant clusters do not contain contiguous intervals of significant Spearman correlation coefficients; rather, clusters denote the intervals within which there is a conditional probability of  $p$  that the interval contains at least one significant coefficient, given the Type 1 error rate (cluster induction parameter) for evaluating the coefficients independently.

7. Notably, Bonner and Epstein (2017) found OPA sensitivity to navigability information not only using current stimuli, which are computer-generated, but also using naturalistic images of indoor scenes.

### REFERENCES

- Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25, 60–83. <https://doi.org/10.3102/10769986025001060>
- Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences, U.S.A.*, 114, 4793–4798. <https://doi.org/10.1073/pnas.1618228114>, PubMed: 28416669
- Bonner, M. F., & Epstein, R. A. (2018). Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLoS Computational Biology*, 14, e1006111. <https://doi.org/10.1371/journal.pcbi.1006111>, PubMed: 29684011
- Brederoo, S. G., Nieuwenstein, M. R., Lorist, M. M., & Cornelissen, F. W. (2017). Hemispheric specialization for global and local processing: A direct comparison of linguistic and non-linguistic stimuli. *Brain and Cognition*, 119, 10–16. <https://doi.org/10.1016/j.bandc.2017.09.005>, PubMed: 28923763
- Cichy, R. M., Khosla, A., Pantazis, D., & Oliva, A. (2017). Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks. *Neuroimage*, 153, 346–358. <https://doi.org/10.1016/j.neuroimage.2016.03.063>, PubMed: 27039703
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3, 345–351. [https://doi.org/10.1016/S1364-6613\(99\)01361-3](https://doi.org/10.1016/S1364-6613(99)01361-3)
- 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. *Journal of Neuroscience*, 31, 11305–11312. <https://doi.org/10.1523/JNEUROSCI.1935-11.2011>, PubMed: 21813690
- Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, 33, 1331–1336. <https://doi.org/10.1523/JNEUROSCI.4081-12.2013>, PubMed: 23345209

- Djebbara, Z., Fich, L. B., Petrini, L., & Gramann, K. (2019). Sensorimotor brain dynamics reflect architectural affordances. *Proceedings of the National Academy of Sciences, U.S.A.*, *116*, 14769–14778. <https://doi.org/10.1073/pnas.1900648116>, PubMed: 31189596
- Ekstrom, A. D. (2015). Why vision is important to how we navigate. *Hippocampus*, *25*, 731–735. <https://doi.org/10.1002/hipo.22449>, PubMed: 25800632
- Ekstrom, A. D., Spiers, H. J., Bohbot, V. D., & Rosenbaum, R. S. (2018). *Human spatial navigation*. Princeton University Press. <https://doi.org/10.2307/j.ctvc773wg>
- Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology*, *97*, 3670–3683. <https://doi.org/10.1152/jn.00003.2007>, PubMed: 17376855
- Epstein, R. A., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning places from views: Variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience*, *17*, 73–83. <https://doi.org/10.1162/0898929052879987>, PubMed: 15701240
- Flevaris, A. V., & Robertson, L. C. (2016). Spatial frequency selection and integration of global and local information in visual processing: A selective review and tribute to Shlomo Bentin. *Neuropsychologia*, *83*, 192–200. <https://doi.org/10.1016/j.neuropsychologia.2015.10.024>, PubMed: 26485158
- Ghuman, A. S., & Martin, A. (2019). Dynamic neural representations: An inferential challenge for fMRI. *Trends in Cognitive Sciences*, *23*, 534–536. <https://doi.org/10.1016/j.tics.2019.04.004>, PubMed: 31103440
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton, Mifflin and Company.
- Greene, M. R., & Oliva, A. (2009). The briefest of glances the time course of natural scene understanding. *Psychological Science*, *20*, 464–472. <https://doi.org/10.1111/j.1467-9280.2009.02316.x>, PubMed: 19399976
- Groen, I. I. A., Ghebreab, S., Lamme, V. A. F., & Scholte, H. S. (2016). The time course of natural scene perception with reduced attention. *Journal of Neurophysiology*, *115*, 931–946. <https://doi.org/10.1152/jn.00896.2015>, PubMed: 26609116
- Groen, I. I. A., Ghebreab, S., Prins, H., Lamme, V. A., & Scholte, H. S. (2013). From image statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *Journal of Neuroscience*, *33*, 18814–18824. <https://doi.org/10.1523/JNEUROSCI.3128-13.2013>, PubMed: 24285888
- Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *372*, 20160102. <https://doi.org/10.1098/rstb.2016.0102>, PubMed: 28044013
- Hansen, N. E., Noesen, B. T., Nador, J. D., & Harel, A. (2018). The influence of behavioral relevance on the processing of global scene properties: An ERP study. *Neuropsychologia*, *114*, 168–180. <https://doi.org/10.1016/j.neuropsychologia.2018.04.040>, PubMed: 29729276
- Harel, A., & Al Zoubi, H. (2019). Early electrophysiological correlates of scene perception are sensitive to inversion. *Journal of Vision*, *19*, 190. <https://doi.org/10.1167/19.10.190>
- Harel, A., Groen, I. I. A., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted EEG investigation. *eNeuro*, *3*, ENEURO-0139-16.2016. <https://doi.org/10.1523/ENEURO.0139-16.2016>, PubMed: 27699208
- Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: Gradients of object and spatial layout information. *Cerebral Cortex*, *23*, 947–957. <https://doi.org/10.1093/cercor/bhs091>, PubMed: 22473894
- Harel, A., Mzozoyana, M. W., Al Zoubi, H., Nador, J. D., Noesen, B. T., Lowe, M. X., et al. (2020). Artificially-generated scenes demonstrate the importance of global scene properties for scene perception. *Neuropsychologia*, *141*, 107434. <https://doi.org/10.1016/j.neuropsychologia.2020.107434>, PubMed: 32179102
- Hassabis, D., & Maguire, E. A. (2009). The construction system of the brain. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *364*, 1263–1271. <https://doi.org/10.1098/rstb.2008.0296>, PubMed: 19528007
- Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron*, *37*, 1027–1041. [https://doi.org/10.1016/S0896-6273\(03\)00144-2](https://doi.org/10.1016/S0896-6273(03)00144-2), PubMed: 12670430
- Henriksson, L., Mur, M., & Kriegeskorte, N. (2019). Rapid invariant encoding of scene layout in human OPA. *Neuron*, *103*, 161–171. <https://doi.org/10.1016/j.neuron.2019.04.014>, PubMed: 31097360
- Jelić, A., Tieri, G., Me Matteis, F., Babiloni, F., & Vecchiato, G. (2016). The enactive approach to architectural experience: A neurophysiological perspective on embodiment, motivation, and affordances. *Frontiers in Psychology*, *7*, 481. <https://doi.org/10.3389/fpsyg.2016.00481>, PubMed: 27065937
- Julian, J. B., Keinath, A. T., Marchette, S. A., & Epstein, R. A. (2018). The neurocognitive basis of spatial reorientation. *Current Biology*, *28*, R1059–R1073. <https://doi.org/10.1016/j.cub.2018.04.057>, PubMed: 30205055
- 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. <https://doi.org/10.1016/j.cub.2016.02.066>, PubMed: 27020742
- Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M. J., Iragui, V., et al. (1998). Extended ICA removes artifacts from electroencephalographic recordings. *Advances in Neural Information Processing Systems*, *10*, 894–900.
- Kaiser, D., Häberle, G., & Cichy, R. M. (2020). Cortical sensitivity to natural scene structure. *Human Brain Mapping*, *41*, 1286–1295. <https://doi.org/10.1002/hbm.24875>, PubMed: 31758632
- Kaiser, D., Turini, J., & Cichy, R. M. (2019). A neural mechanism for contextualizing fragmented inputs during naturalistic vision. *eLife*, *8*, e48182. <https://doi.org/10.7554/eLife.48182>, PubMed: 31596234
- 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. <https://doi.org/10.1016/j.neuroimage.2016.02.062>, PubMed: 26931815
- Kamps, F. S., Lall, V., & Dilks, D. D. (2016). The occipital place area represents first-person perspective motion information through scenes. *Cortex*, *83*, 17–26. <https://doi.org/10.1016/j.cortex.2016.06.022>, PubMed: 27474914
- Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*, 323–343. <https://doi.org/10.1007/s00422-018-0753-2>, PubMed: 29572721
- Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual cortex: It's the spaces more than the places. *Journal of Neuroscience*, *31*, 7322–7333. <https://doi.org/10.1523/JNEUROSCI.4588-10.2011>, PubMed: 21593316
- 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. <https://doi.org/10.1038/nrn3008>, PubMed: 21415848
- Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*, 72–79. <https://doi.org/10.1016/j.tics.2007.11.004>, PubMed: 18178125

- Lescroart, M. D., & Gallant, J. L. (2019). Human scene-selective areas represent 3D configurations of surfaces. *Neuron*, *101*, 178–192. <https://doi.org/10.1016/j.neuron.2018.11.004>, PubMed: 30497771
- Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Frontiers in Computational Neuroscience*, *9*, 135. <https://doi.org/10.3389/fncom.2015.00135>, PubMed: 26594164
- Lowe, M. X., Rajsic, J., Ferber, S., & Walther, D. B. (2018). Discriminating scene categories from brain activity within 100 milliseconds. *Cortex*, *106*, 275–287. <https://doi.org/10.1016/j.cortex.2018.06.006>, PubMed: 30037637
- Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.
- Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences*, *20*, 843–856. <https://doi.org/10.1016/j.tics.2016.09.003>, PubMed: 27769727
- Nakamura, K., Kawashima, R., Sato, N., Nakamura, A., Sugiura, M., Kato, T., et al. (2000). Functional delineation of the human occipito-temporal areas related to face and scene processing: A PET study. *Brain*, *123*, 1903–1912. <https://doi.org/10.1093/brain/123.9.1903>, PubMed: 10960054
- Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195050387.001.0001>
- Persichetti, A. S., & Dilks, D. D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex*, *77*, 155–163. <https://doi.org/10.1016/j.cortex.2016.02.006>, PubMed: 26963085
- Persichetti, A. S., & Dilks, D. D. (2018). Dissociable neural systems for recognizing places and navigating through them. *Journal of Neuroscience*, *38*, 10295–10304. <https://doi.org/10.1523/JNEUROSCI.1200-18.2018>, PubMed: 30348675
- Petrov, Y. (2012). Harmony: EEG/MEG linear inverse source reconstruction in the anatomical basis of spherical harmonics. *PLoS One*, *7*, e44439. <https://doi.org/10.1371/journal.pone.0044439>, PubMed: 23071497
- Pinhas, M., Tzelgov, J., & Ganor-Stern, D. (2012). Estimating linear effects in ANOVA designs: The easy way. *Behavior Research Methods*, *44*, 788–794. <https://doi.org/10.3758/s13428-011-0172-y>, PubMed: 22101656
- Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, *12*, 852–877. <https://doi.org/10.1080/13506280444000553>
- Sadeh, B., Podlipsky, I., Zhdanov, A., & Yovel, G. (2010). Event-related potential and functional MRI measures of face-selectivity are highly correlated: A simultaneous ERP-fMRI investigation. *Human Brain Mapping*, *31*, 1490–1501. <https://doi.org/10.1002/hbm.20952>, PubMed: 20127870
- Sato, N., Nakamura, K., Nakamura, A., Sugiura, M., Ito, K., Fukuda, H., et al. (1999). Different time course between scene processing and face processing: A MEG study. *NeuroReport*, *10*, 3633–3637. <https://doi.org/10.1097/00001756-199911260-00031>, PubMed: 10619657
- Sestito, M., Flach, J., & Harel, A. (2018). Grasping the world from a cockpit: Perspectives on embodied neural mechanisms underlying human performance and ergonomics in aviation context. *Theoretical Issues in Ergonomics Science*, *19*, 692–711. <https://doi.org/10.1080/1463922X.2018.1474504>
- Sestito, M., Harel, A., Nador, J., & Flach, J. (2018). Investigating neural sensorimotor mechanisms underlying flight expertise in pilots: Preliminary data from an EEG study. *Frontiers in Human Neuroscience*, *12*, 489. <https://doi.org/10.3389/fnhum.2018.00489>, PubMed: 30618676
- 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. <https://doi.org/10.1523/JNEUROSCI.0559-09.2009>, PubMed: 19710310
- Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences, U.S.A.*, *108*, 9661–9666. <https://doi.org/10.1073/pnas.1015666108>, PubMed: 21593417
- Wiesmann, C. G., Friederici, A. D., Singer, T., & Steinbeis, N. (2020). Two systems for thinking about others’ thoughts in the developing brain. *Proceedings of the National Academy of Sciences, U.S.A.*, *117*, 6928–6935. <https://doi.org/10.1073/pnas.1916725117>, PubMed: 32152111
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636. <https://doi.org/10.3758/BF03196322>, PubMed: 12613670
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind–brain perspective. *Psychological Bulletin*, *133*, 273–293. <https://doi.org/10.1037/0033-2909.133.2.273>, PubMed: 17338600
- Zhang, H., Houpt, J. W., & Harel, A. (2019). Establishing reference scales for scene naturalness and openness. *Behavior Research Methods*, *51*, 1179–1186. <https://doi.org/10.3758/s13428-018-1053-4>, PubMed: 29845553