Neuropsychologia 141 (2020) 107434

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://www.elsevier.com/locate/neuropsychologia

Artificially-generated scenes demonstrate the importance of global scene properties for scene perception

Assaf Harel [a,*], Mavuso W. Mzozoyana [b], Hamada Al Zoubi [b], Jeffrey D. Nador [a], Birken T. Noesen [a], Matthew X. Lowe [c], Jonathan S. Cant [d]

[a] Department of Psychology, Wright State University, Dayton, OH, USA
[b] Department of Neuroscience, Cell Biology and Physiology, Wright State University, Dayton, OH, USA
[c] Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA
[d] Department of Psychology, University of Toronto Scarborough, Toronto, ON, Canada

ARTICLE INFO

Keywords:
Scene perception
Scene recognition
ERP
EEG
Vision
P2

ABSTRACT

Recent electrophysiological research highlights the significance of global scene properties (GSPs) for scene perception. However, since real-world scenes span a range of low-level stimulus properties and high-level contextual semantics, GSP effects may also reflect additional processing of such non-global factors. We examined this question by asking whether Event-Related Potentials (ERPs) to GSPs will still be observed when specific low- and high-level scene properties are absent from the scene. We presented participants with computer-based artificially-manipulated scenes varying in two GSPs (spatial expanse and naturalness) which minimized other sources of scene information (color and semantic object detail). We found that the peak amplitude of the P2 component was sensitive to the spatial expanse and naturalness of the artificially-generated scenes: P2 amplitude was higher to closed than open scenes, and in response to manmade than natural scenes. A control experiment showed that the effect of Naturalness on the P2 is not driven by local texture information, while earlier effects of naturalness, expressed as a modulation of the P1 and N1 amplitudes, are sensitive to texture information. Our results demonstrate that GSPs are processed robustly around 220 ms and that P2 can be used as an index of global scene perception.

1. Introduction

Humans show a remarkable ability to perceive and recognize visual scenes. People can describe the gist of a scene (i.e. provide a basic-level account of what the scene is – a kitchen, an office, a forest, etc.) very rapidly, under very brief presentation rates (Intraub, 1981; Joubert et al., 2007; Potter, 1976; Potter and Levy, 1969), and with very little attentional resources (Greene and Fei-Fei, 2014, 2017; Li et al., 2002, but see Gronau and Izoutcheev, 2017). Similarly, human memory for scenes is highly efficient and accurate, both for short-term (Hollingworth, 2004; Velisavljević and Elder, 2008) and long-term episodic memory (Hollingworth, 2004, 2005; Konkle et al., 2010). These scene recognition abilities are quite impressive if one considers the computational complexity involved in scene understanding, particularly given the enormous variety of properties that scenes vary on, ranging from low-level features and summary image statistics to high-level semantics and action affordances (Groen et al., 2017; Malcolm et al., 2016).

In light of the richness of scene stimulus information, accounts of scene recognition vary dramatically on the question of what constitutes the critical information enabling rapid scene categorization. On the one hand, 'high-level' accounts posit that rapid scene categorization is achieved primarily through the extraction of basic-level category information (Tversky and Hemenway, 1983; Walther et al., 2009), one example of which is by assigning semantic labels to objects in scenes (Çukur et al., 2016; Stansbury et al., 2013). On the other hand, 'low-level' accounts suggest that physical properties of stimuli (often captured by image-based statistics) are sufficient for explaining the speed and accuracy with which people categorize scenes. Examples for such diagnostic image-based properties are contrast energy, spatial frequency, texture, and color, which have all been shown to contribute significantly to successful scene categorization (Andrews et al., 2015; Hansen et al., 2012, 2011; Lowe et al., 2018; Oliva and Schyns, 2000). Further complicating the question of which type of information plays a greater role in scene recognition is the fact that the low- and high-level

* Corresponding author. 335 Fawcett Hall 3640 Col. Glenn Highway, Dayton, Ohio, 45435, USA.
E-mail address: assaf.harel@wright.edu (A. Harel).

https://doi.org/10.1016/j.neuropsychologia.2020.107434
Received 10 June 2019; Received in revised form 4 March 2020; Accepted 9 March 2020
Available online 13 March 2020
0028-3932/

features often co-vary and cannot be easily teased apart (Groen et al., 2018, 2017; Harel et al., 2016; Lescroart et al., 2015). For example, a forest scene may be dominated by high spatial frequencies, near-vertical orientations, and particular colors, while at the same time it also can be described by the co-occurrence of specific objects (trees), or certain navigability affordances (e.g., how dense is the forest, does it contain navigable paths, etc.).

One approach for bridging the gap between accounts focusing on low-level physical stimulus properties and higher-level scene attributes has been the global view of scene recognition (Greene and Oliva, 2010). The global view of scene processing suggests that the gist of a scene is extracted by forming an initial representation of the scene's coarse layout without specifying its local elements (Brady et al., 2017; Kauffmann et al., 2015; Musel et al., 2014; Schyns and Oliva, 1994). Computationally, this might be achieved by calculating the spatial envelope of the scene, which preserves information about spatial frequency and orientation distribution (Aude Oliva and Torralba, 2001, 2006). Critically, particular distributions of spatial frequencies and orientations can be mapped onto psychologically-real, ecological scene dimensions known as global scene properties (GSPs) (Ross and Oliva, 2010). GSPs can thus be described as ecological scene primitives that capture the structure and function of real-world scenes (Greene and Oliva, 2009a). Examples of GSPs include the scene's mean depth, spatial layout, naturalness, openness and navigability (Greene and Oliva, 2009a; Joubert et al., 2007; Lowe et al., 2016; Ross and Oliva, 2010; Schyns and Oliva, 1994). GSPs are encoded at the early stages of scene processing, are processed swiftly with very little effects of attention (Harel et al., 2016; Hansen et al., 2018), and thus enable rapid scene categorization (Brady et al., 2017; Greene and Oliva, 2009a, 2009b; Ross and Oliva, 2010).

Recent neurophysiological research has started to uncover the temporal dynamics of GSP processing. Information about the spatial expanse of a scene was found to be processed in the brain by 250 ms post-stimulus onset (Cichy et al., 2017; Harel et al., 2016; Hansen et al., 2018), and the naturalness of the scene is processed at an even earlier time window, around 120 ms post-stimulus onset at medial occipital electrode sites (Groen et al., 2013) and around 170 ms at lateral occipito-temporal sites (Harel et al., 2016; Hansen et al., 2018). The early latency of these electrophysiological signatures lends support to the notion that global scene information is extracted at the early stages of visual processing, putatively supporting rapid pre-attentive scene categorization (Greene and Oliva, 2009b; Groen et al., 2016; Rousselet et al., 2005). Further support for the significance of global scene information for scene recognition comes from neuroimaging research showing that several key GSPs are processed in scene-selective cortex. For example, response patterns in the Parahippocampal Place Area (PPA) and Retrosplenial Complex (RSC) convey information about how enclosed a scene is (spatial expanse) (Harel et al., 2013; Kravitz et al., 2011; Lowe et al., 2016; Park et al., 2011), its spatial boundary (Ferrara and Park, 2016), mean depth (Kravitz et al., 2011), as well as how cluttered its contents are (Park et al., 2014). Notably, responses in the Occipital Place Area (OPA: Dilks et al., 2013), and to a lesser extent PPA, also contain specific information about the potential ease of navigation in a given scene (navigability affordances) (Bonner and Epstein, 2017).

As suggested above, GSPs provide an intermediate level of representation that could serve as a link between low-level features and high-level semantic scene representations. It is still an open question, however, as to what extent the early electrophysiological responses to GSPs are driven exclusively by mid-level global scene information, or whether they also incorporate low-as well as high-level scene information. For instance, much of the early brain responses discriminating manmade and natural scenes could potentially be explained by low-level image statistics, such as spatial coherence and contrast (Groen et al., 2012; Groen et al., 2013). Image statistics can be used to distinguish GSPs (including spatial expanse and naturalness) as well as basic-level scene categories from neural signals within the first 100 ms post-stimulus onset (Lowe et al., 2018). Alternatively, high-level semantic information can also have an impact on early responses to scene naturalness (Joubert et al., 2007). The problem is that irrespective of the level of information that may be associated with the GSP, the use of real-world naturalistic scene images, which is essential for maintaining ecological validity, poses at the same time a real challenge for determining the unique role that GSPs play in scene processing. How then can the processing of GSPs be disentangled from the processing of both types of scene information? To what extent do the early brain signatures of scene processing convey uniquely global scene information? In the present study we addressed these questions by using a set of artificially-generated scene stimuli that were specifically designed to systematically vary along two GSPs (spatial expanse and naturalness) while minimizing one source of low-level information (color) and high-level semantic details (object identities). The scene stimuli, originally used by Lowe et al. (2016), are grayscale scene images based on real-world photographs which have been digitally manipulated to remove salient semantic objects and features (see Fig. 1 for examples).[1] These 'impoverished' scenes allowed us to investigate global scene information while minimizing (but not abolishing; see discussion below) the involvement of non-global low- or high-level sources of information.

Color is one important source of diagnostic scene information (Oliva and Schyns, 2000; for a review, see Tanaka et al., 2001) that may be used in conjunction with other low-level features to form the scene's gist (Oliva and Schyns, 2000). Thus, one may argue that the neural responses to GSPs to a large degree reflect color processing rather than the extraction of global scene structure. For example, open scenes on average might contain more blue image patches than closed scenes; natural scenes may contain more greens and brown earth tones than manmade scenes. At the same time, on the other end of the (representational) spectrum, real-world scenes also convey high-level semantic features and contextual associations, carried, for instance, by local object information (Võ and Henderson, 2009; Võ and Wolfe, 2013) and semantic category information (Walther et al., 2009; Walther et al., 2011). Basic-level scene category might be detected or inferred based on salient objects, which may lead to subsequent categorization of the scene as manmade or natural, or even as open or closed. In that case, removing most of the prominent objects from the scene would make scene category information more ambiguous and thus should hamper the processing of these respective GSPs. Therefore, our rationale in the present study was that if the early extraction of global scene information is based predominantly on either one of these low- or high-level sources of scene information, one should expect the processing of GSPs to be significantly attenuated when presented with the "impoverished" artificially-generated scene stimuli used in this study.

To test this hypothesis, we measured event-related potentials (ERPs) in two experiments. In Experiment 1 we tested whether the removal of color and semantic object information would hamper the processing of diagnostic GSPs, such as spatial expanse and naturalness. Research from our lab demonstrates that the posterior P2 ERP component is scene-selective, and is sensitive to the GSPs of spatial expanse and naturalness: P2 amplitude is higher to closed than open scenes and is also higher to natural than manmade scenes (Harel et al., 2016; Hansen et al., 2018). According to the above logic, early evoked responses to GSPs, particularly the P2 component, should be expected to be reduced or altogether absent using our artificially-generated scenes, compared with the ERP responses to rich naturalistic scene images used in our previous

[1] Note that we use the term "artificially-generated" to refer to the stimulus format we used (i.e., scenes images that were manipulated in Photoshop, as compared with 'naturalistic' photographs of scenes that were not altered using image-processing software in previous studies), whereas the use of "manmade" refers to the content within the scene, irrespective of the stimulus format (i.e., materials made by humans for 'manmade' versus naturally occurring materials unaltered by humans for 'natural').

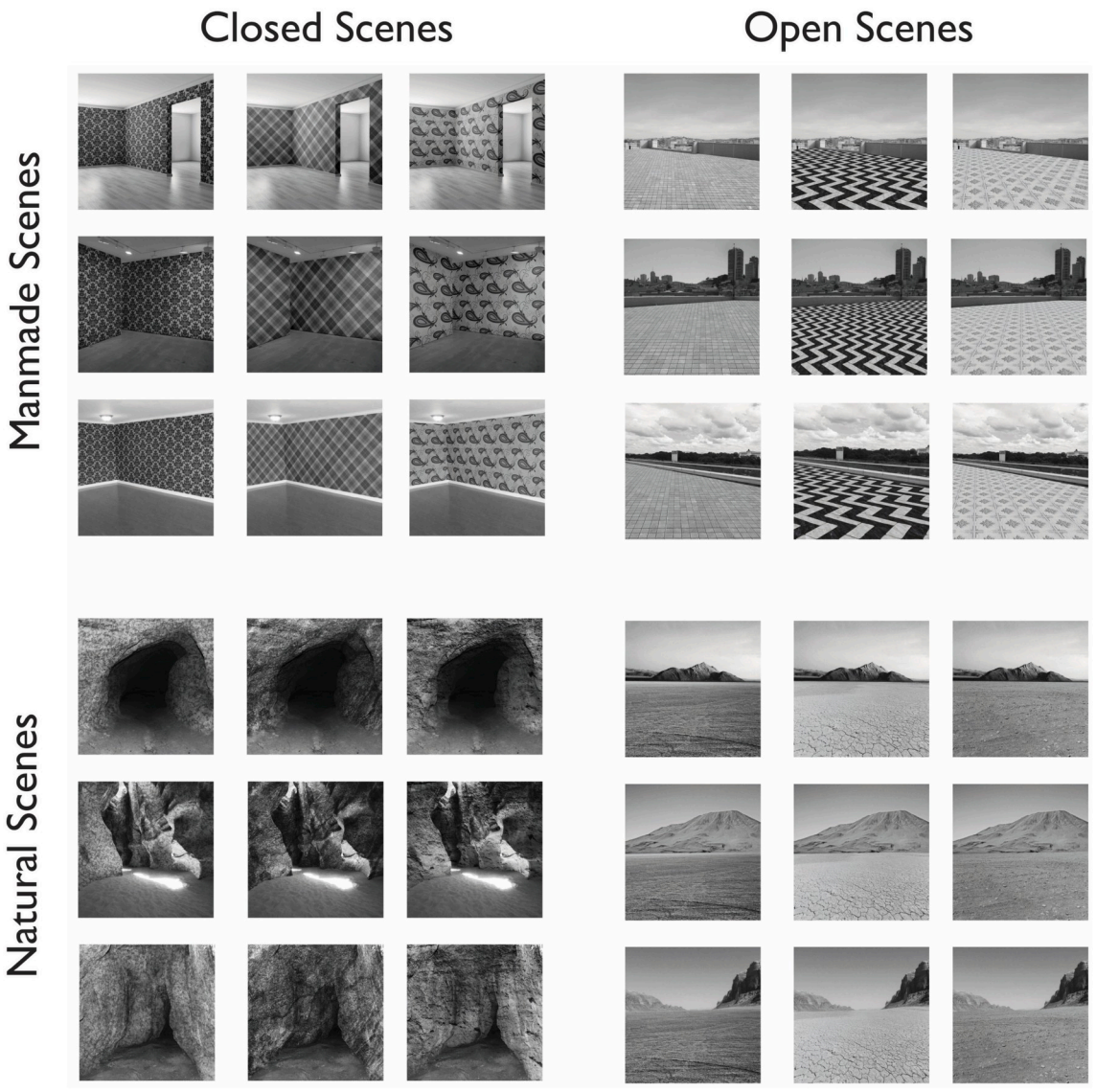

Fig. 1. Examples of scene images used in Experiment 1. Stimuli spanned both Naturalness and Spatial Expanse dimensions, resulting in four categories: Manmade Closed scenes (rooms), Manmade Open scenes (rooftops), Natural Closed scenes (caves), and Natural Open scenes (deserts). Depicted here are nine scene examples (out of 144 scenes) per category. Exemplars within each category varied in the layout (presented here across rows) and the texture (presented here across columns) of the scenes.

studies. Alternatively, if the extraction of information about the spatial expanse and naturalness of the scene is not mediated solely by color and semantic information, we should observe robust spatial expanse and naturalness effects on the P2 amplitude. In Experiment 2, we tested whether potential GSP effects observed with our set of 'impoverished' artificially-ge f GSPs per se. Specifically, scene naturalness, which has been shown to be extracted earlier than spatial expanse (Harel et al., 2016; Hansen et al., 2018), may be inferred based on local textural information (e.g. soil vs. tiles), and thus, a naturalness effect on P2 may be driven by the texture of the scene and not by global scene information. To address this concern, we devised a new set of scenes by manipulating the artificially-generated scene stimuli used in Experiment 1. Specifically, we replaced the original textures of the natural scenes with manmade textures and vice versa. We then measured the effects of naturalness on P2 amplitude using the original scenes containing textures that were consistent with their category (e.g. a rooftop paved with tiles), and the newly manipulated scenes containing textures that were inconsistent with their category (e.g. a rooftop with a ground soil texture). Our hypothesis was that if the naturalness effect was driven mainly by texture rather than scene category, we should expect a reversal (or at least a minimization) of the direction of the P2 effect when comparing the texture-inconsistent scenes to the texture consistent ones. However, if P2 indexes naturalness using information other than local texture, then we should expect an effect of naturalness independent of texture consistency.

2. Materials and Methods

2.1. Participants

Seventeen Wright State University students (three females, mean age: 22.5 years) participated in Experiment 1. Twenty-nine participated in Experiment 2 (18 females, mean age: 19.6 years). All participants signed an informed written consent form according to the guidelines of the Institutional Review Board (IRB) of Wright State University. All had normal or corrected-to-normal visual acuity with no history of neurological diseases. Three participants in Experiment 1 and six participants in Experiment 2 were excluded from the final analyses due to excessive

EEG artifacts. All participants were compensated monetarily or with course credit.

2.2. Stimuli

2.2.1. Experiment 1

Stimuli were scene images used in a previous neuroimaging study (Lowe et al., 2016) (See Fig. 1 for examples of scene images). The stimulus set comprised of 576 grayscale computer-generated scene images devoid of any foreground objects, to avoid any contextual or semantic effects. The scenes spanned two GSPs: Spatial Expanse (closed/open) and Naturalness (manmade/natural), which mapped onto four coarse scene categories: a room scene (closed manmade), a rooftop scene (open manmade), a cave scene (closed natural), and a 'desert plateau' scene (open natural). Each scene category contained twelve unique structural arrangements (i.e., layouts), and each layout had twelve appropriate textures applied to their dominant surface (mapped onto scene gradient and depth using Adobe Photoshop CS3), totaling 144 unique scene exemplars per category (12 layouts/category × 12 textures/layout × 4 scene categories = 576 total images).

2.2.2. Experiment 2

To test the extent to which local texture is intrinsically linked with naturalness, a new subset of scenes was created by manipulating the scenes used in Experiment 1. Specifically, we applied twelve new textures to each of the twelve layouts of the scenes used in Experiment 1. Textures were applied to the scene's dominant surface, and were mapped onto its gradient and depth using the same procedures as in Experiment 1. Critically, however, these textures were inappropriate (i.e. inconsistent) with the overall naturalness of the scene. For example, the walls in the room scenes (closed manmade) were covered with textures of soil (natural) rather than a standard manmade textured wallpaper. Similarly, the ground of the desert scenes (open natural) now contained geometric manmade patterns rather than a more naturalistic sand texture (see Fig. 2 for a sample of the stimuli). Scenes from the cave category were not used due to difficulties in fitting the manmade textures to their undulating surfaces. These naturalness-inconsistent scenes, together with their original, naturalness-consistent scene counterparts comprised of the stimulus set used in Experiment 2, totaling 144 unique scene exemplars per category (12 layouts per category × 12 textures per layout) and a total of 864 individual scenes (144 scenes/category x 3 scene categories x 2 consistency levels = 864 total images).

Images across both experiments were 500 × 500 pixels and were presented on a Dell LCD monitor at the center of the screen, viewed from a distance of about 110 cm corresponding to 8.86° of visual angle. The stimuli were presented using Presentation software (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com).

2.3. Experimental design and procedure

Participants in Experiment 1 viewed the 576 individual scene stimuli across 12 blocks. Each block comprised of 96 images spanning the four scene categories, with the scene stimuli pseudo-randomized within individual blocks and across the 12 blocks. Each scene stimulus was presented twice across the entire length of the experiment, totaling 1152 trials. In Experiment 2, participants viewed the 864 individual scene stimuli across six blocks, each block comprised of 144 images, and presentation of scene stimuli were pseudo-randomized both within and across blocks. Each scene stimulus was presented once across the entire length of the experiment.

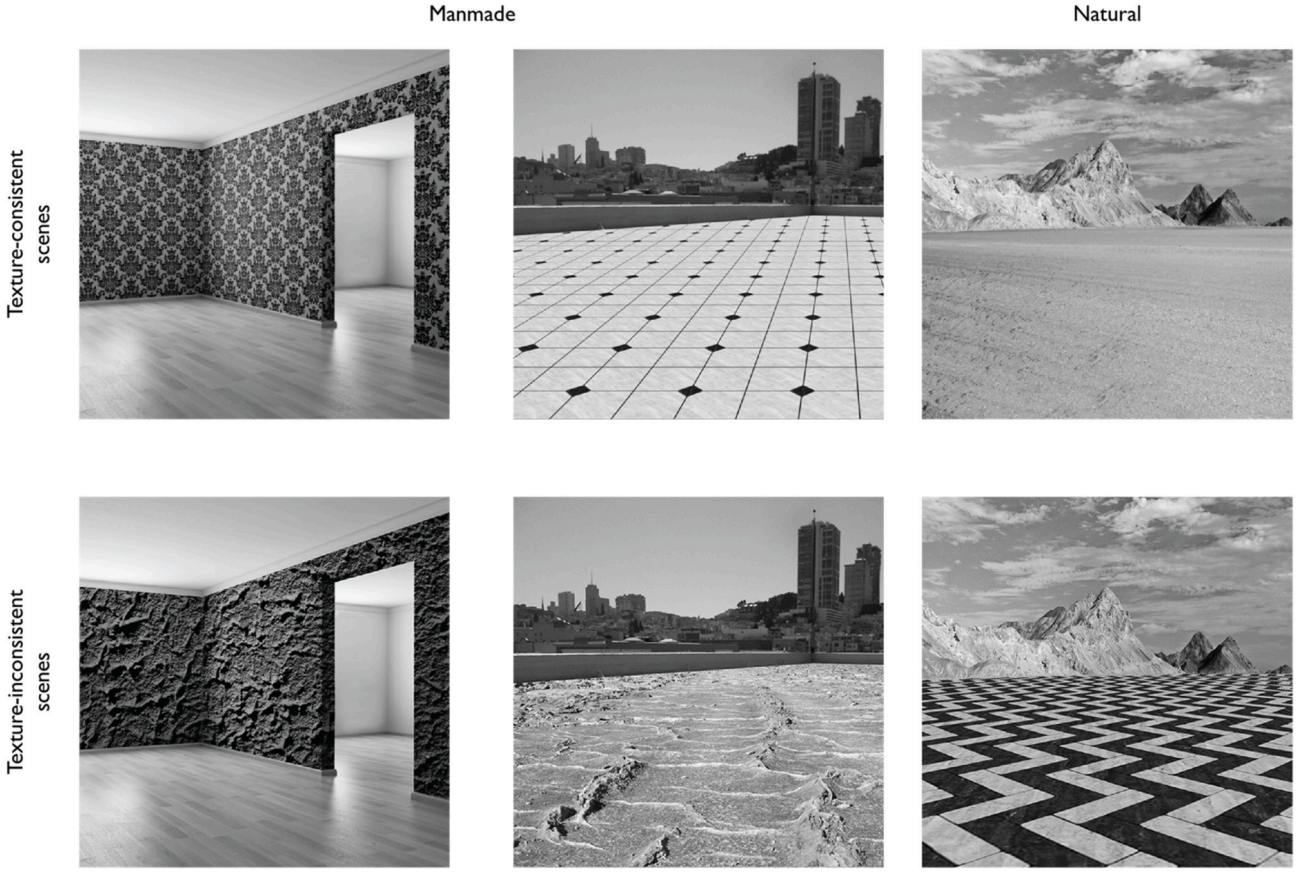

Fig. 2. Examples of scene images used in the Experiment 2. Top row: Texture-consistent scenes, namely, the same manmade and natural scenes used in Experiment 1 (caves were not included). Bottom row: Texture-inconsistent scenes. These were the same scenes with their main surface textures swapped for textures of the opposing naturalness category (e.g. manmade scenes with natural textures).

In both experiments, scene stimuli were presented for 500 ms with a jittered interstimulus interval ranging from 750 ms to 1250 ms. Participants performed a fixation cross task, in which they were required to report whether the horizontal or vertical bar of the central fixation cross lengthened on each trial. Changes in the fixation cross were pseudo-randomized across all trials, and hence were independent from the actual content of the underlying image, in essence requiring the participants to pay very little, if any, attention to the background images. This is the same task employed in prior EEG studies of scene processing using naturalistic real-world stimuli (Hansen et al., 2018; Harel et al., 2016).

2.4. EEG recording

The EEG analog signals were recorded using 64 Ag–AgCl pin-type active electrodes (Biosemi ActiveTwo, Amsterdam) mounted on an elastic cap (ECI) according to the extended 10–20 system, and from two additional electrodes placed at the right and left mastoids, and an electrode placed on the tip of the nose. All electrodes were referenced to the Common Mode Signal (CMS) electrode placed between electrodes PO3 and PO4. Eye movements, as well as blinks, were monitored using two pairs of EOG electrodes, one pair attached to the external canthi, and the other to the infraorbital and supraorbital regions of the right eye. Both EEG and EOG were sampled at 512 Hz with a resolution of 24 bits with an active input range of −262 mV to +262 mV per bit, with online low pass filter of 51 Hz to prevent aliasing. The digitized EEG was saved and processed off-line.

2.5. Data processing

The data were preprocessed using Brain Vision Analyzer 2 (Brain Products GmbH, Munich, Germany). The raw data were first 1.0 Hz high-pass filtered (24 dB) and referenced to the tip of the nose. Eye movements were corrected using an ocular correction ICA procedure (for details see Jung et al., 1998). Remaining artifacts exceeding ±100 mV in amplitude or containing a change of over 100 mV in a period of 50 ms were rejected. The preprocessed data was then segmented into epochs ranging from −200 ms before to 800 ms after stimulus onset for all conditions.

2.6. ERP analysis

For each participant, the peaks of the P1, N1 and P2 in each experimental condition were determined as the most positive peak between 80 and 130 ms, the most negative peak between 130 and 200 ms, and the most positive peak between 200 and 320 ms, respectively. 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). Mean peak amplitudes (across participants) were analyzed using a three-way within subject ANOVA. In Experiment 1, this included Hemisphere (left, right), Naturalness (manmade, natural), and Spatial Expanse (open, closed) as independent factors. In Experiment 2 this included Hemisphere (left, right), Naturalness (manmade, natural), and Texture Consistency (consistent, inconsistent) as independent variables. Only significant effects are reported.

3. Results

3.1. Experiment 1

Grand-average waveforms are depicted in Fig. 3. We conducted an omnibus 3-way repeated-measures ANOVA on the amplitude of the individually defined peaks of each of the ERP components, with Hemisphere (left, right), Naturalness (manmade, natural), and Spatial Expanse (closed, open) as independent factors. The significant results of these analyses are reported in Fig. 4.

3.1.1. P2 component

The posterior P2 component has previously been reported to be a key ERP component in scene perception, carrying scene information at multiple levels of representation, including between-category information (differentiating scenes from objects and faces), within-category features (GSPs), and summary image statistics (contrast and spatial frequency) (Harel et al., 2016). We found that the P2 amplitude is sensitive to the global information contained in the artificial scenes, with both spatial expanse and naturalness modulating the amplitude of the P2 component (Fig. 4a). We found a significant main effect of Spatial Expanse ($F(1,13) = 8.37$, $MSE = 1.45$, $p < 0.01$), with closed scenes evoking a higher positive response than open scenes ($M_{closed} = 2.00$ mV, $SE = 0.51$; $M_{open} = 1.34$, $SE = 0.55$). The effect of Spatial Expanse was

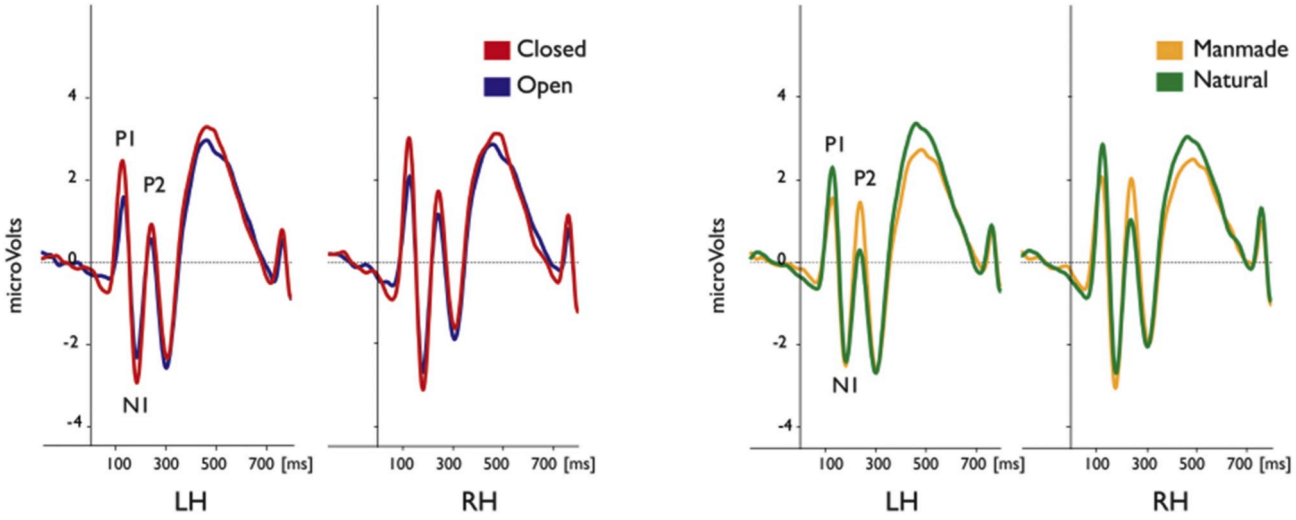

Fig. 3. Group-averaged waveforms for the two global scene properties plotted for the left and right hemisphere (LH, RH, respectively) at posterior lateral sites. Left: Spatial expanse (closed vs. open; red and blue, respectively). Right: Naturalness (man-made vs. natural; yellow and green, respectively). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

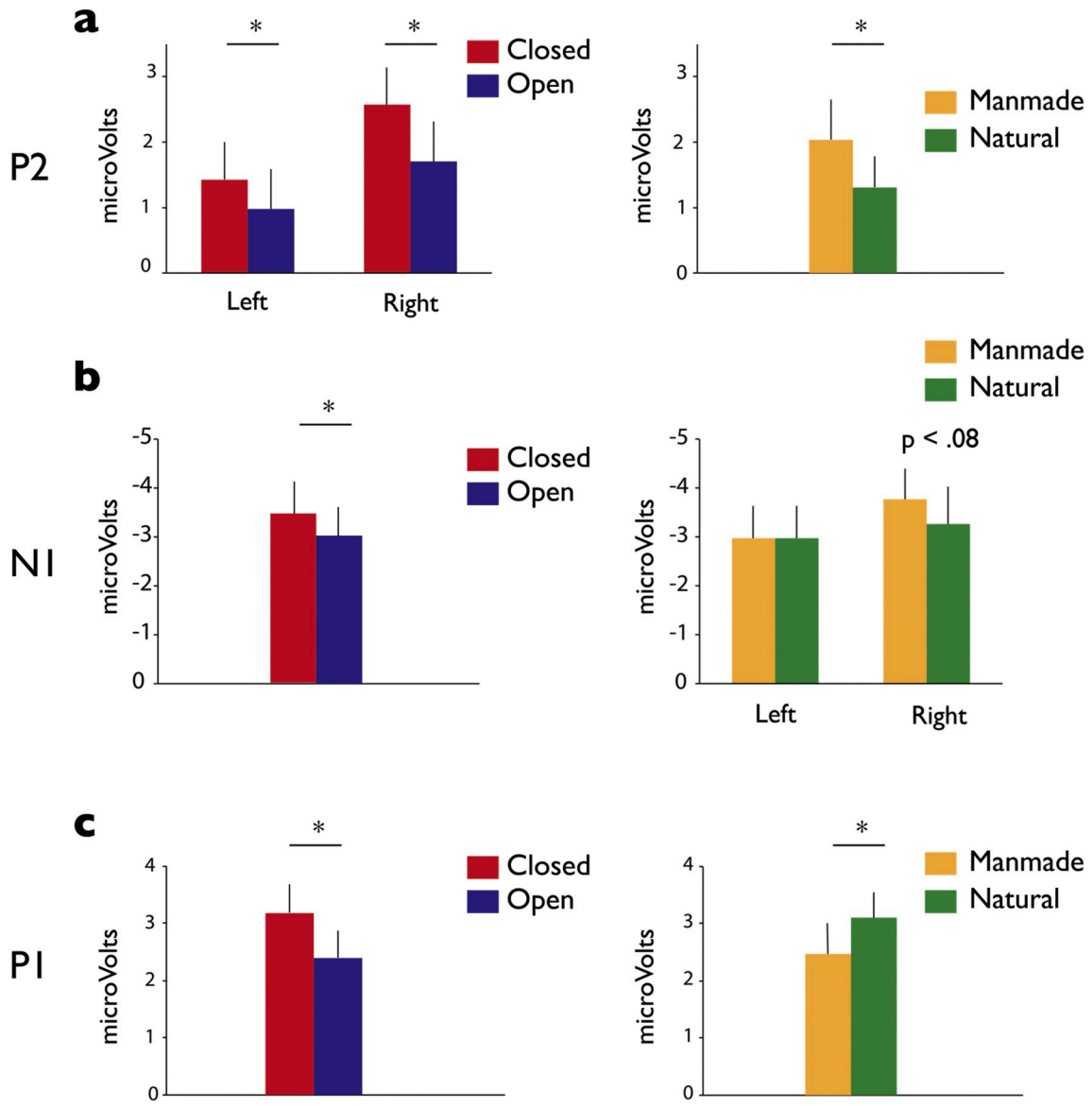

Fig. 4. Grand average ERP analysis results of Experiment 1. (a) Mean P2 peak amplitudes in response to closed and open scenes (red and blue, respectively) presented separately for the left and right hemispheres and in response to manmade and natural scenes (yellow and green, respectively). (b) Mean N1 peak amplitudes in response to closed and open scenes (red and blue, respectively) and in response to man-made (yellow) and natural scenes (green) presented separately for the left and right hemispheres. (c) Mean P1 peak amplitudes in response to closed and open scenes (red and blue, respectively) and in response to man-made and natural scenes (yellow and green, respectively). Data from the left and right hemispheres are plotted on the same graph only when there is a significant interaction with Hemisphere. Otherwise, data are collapsed across hemisphere. Significant differences ($p < 0.05$) between pairs of categories are denoted by an asterisk (error bars indicate between-subjects SE). All data are plotted for the posterior lateral sites. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

more pronounced in the right hemisphere (t (13) = 3.26, $p < 0.003$: Mean difference = 0.87 mV; M_{closed} = 2.57 mV, SE = 0.57; M_{open} = 1.70 mV, SE = 0.60) than in the left hemisphere (t (13) = 2.06, $p < 0.03$: Mean difference = 0.44 mV; M_{closed} = 1.42 mV, SE = 0.52; M_{open} = 0.98 mV, SE = 0.56) (post-hoc comparison following a significant interaction between Spatial Expanse and Hemisphere (F (1,13) = 6.28, MSE = 0.20, $p < 0.03$)) (Figs. 3 and 4). We also found a significant main effect of Naturalness (F (1,13) = 5.56, MSE = 2.67, $p < 0.03$), with manmade scenes evoking a greater positive response than natural scenes ($M_{manmade}$ = 2.04 mV, SE = 0.60; $M_{natural}$ = 1.31 mV, SE = 0.47). Interestingly, whereas the spatial expanse effect was in line with previous findings from our lab (Harel et al., 2016; Hansen et al., 2018), the naturalness effect was not (i.e., manmade greater than natural scenes, in contrast to the opposite direction found in Harel et al., 2016, and Hansen et al., 2018). Lastly, we also found a significant main effect of Hemisphere (F (1,13) = 6.86, MSE = 3.60, $p < 0.02$), with higher responses in the right hemisphere (2.14 mV, SE = 0.57) than in the left hemisphere (M = 1.0 mV, SE = 0.53).

3.1.2. N1 component

An analysis of the N1 component revealed a significant main effect of Spatial Expanse (F (1,13) = 5.79, MSE = 0.90, $p < 0.04$), with closed scenes evoking a more negative response than open scenes (M_{closed} = −3.46 mV, SE = 0.67, M_{open} = −3.03, SE = 0.60). In addition, we found a marginally significant Naturalness by Hemisphere interaction (F (1,13) = 4.26, MSE = 0.42, $p < 0.06$) with the Naturalness effect manifest predominantly in the right (t (13) = 1.55, $p < 0.08$), but not in the left hemisphere (t (13) = 0.006, $p < 1.00$) (Fig. 4b). Namely, manmade scenes evoked a more negative response than natural scenes in the right hemisphere ($M_{manmade}$ = −3.76 mV, SE = 0.63; $M_{natural}$ = −3.26

mV, SE = 0.75).

3.1.3. P1 component

Analysis of the amplitude of the P1 component revealed a significant main effect of Spatial Expanse (F (1,13) = 15.31, MSE = 1.12, p < 0.002), with closed scenes evoking a higher positive response than open scenes (M_{closed} = 3.18 mV, SE = 0.50; M_{open} = 2.39, SE = 0.48). We also observed a significant main effect of Naturalness (F (1,13) = 7.45, MSE = 1.51, p < 0.02), with natural scenes evoking a greater positive response than manmade scenes ($M_{natural}$ = 3.10 mV, SE = 0.44; $M_{manmade}$ = 2.47 mV, SE = 0.53) (Fig. 4c). No significant interactions were observed.

In sum, we found that the early visually-evoked ERP components are sensitive to the GSPs of artificially-generated scene stimuli containing no color or semantic local object details, with differential responses of the P2 as well as the P1 and N1 components to open and closed scenes, as well as to manmade and natural scenes. Interestingly, while both GSPs modulated the P2 amplitude as previously observed with naturalistic scenes, the naturalness GSP effect observed with the current artificial scenes was in the opposite direction to that observed when using naturalistic scenes (cf. Harel et al., 2016). One might interpret this seeming inconsistency as stemming from differences in image properties, raising the possibility that P2 is not sensitive to naturalness per se, but rather to low-level image properties that co-vary with scene naturalness. Texture, in particular, may be one of these properties. Indeed, our scenes contain prominent manmade and natural textural elements, which can serve as a robust diagnostic feature that is picked up by the visual system during the early stages of scene processing. To address this possibility, we conducted Experiment 2 in which we directly tested the impact that texture has on the early visually-evoked components.

3.2. Experiment 2

In Experiment 2, we directly tested the extent to which the naturalness effect on P2 amplitude reflects lower-level texture processing, by substituting natural textures for manmade textures, while holding scene layout constant. Participants saw the exact scene images as in Experiment 1 (except for the cave stimuli, see Materials and Methods above), but in addition they saw a texture-inconsistent version of each scene. If extraction of naturalness information depends on local textures, then P2 activity should distinguish manmade-textured scenes from natural ones, irrespective of the global scene naturalness, resulting in a reversal of the P2 naturalness effect (i.e., natural > manmade). Alternatively, if P2 reflects global processing of naturalness information, its effect will be observed independent of texture manipulations. To test these hypotheses, we conducted an omnibus 3-way repeated-measures ANOVA on the amplitude of the individually defined peaks of each of the ERP components, with Hemisphere (left, right), Naturalness (manmade, natural), and texture Consistency (consistent, inconsistency) as independent factors. Fig. 5 depicts the grand-average waveforms with the impact of naturalness and scene consistency on the mean peak amplitudes presented in Fig. 6.

3.2.1. P2 component

The posterior P2 component was highly sensitive to the naturalness of the scene, irrespective of its texture as evident in a significant main effect of Naturalness (F (1,22) = 19.55, MSE = 3.63, p < 0.0001). Manmade scenes evoked a higher response than natural scenes ($M_{manmade}$ = 4.80 mV, SE = 0.86; $M_{natural}$ = 3.56 mV, SE = 0.79), similar to the findings of Experiment 1. Critically, there was no significant interaction between Naturalness and Texture Consistency (F (1,22) = 1.57, MSE = 1.68, p < 0.22), that is, the effect of Naturalness was observed both in the texture-consistent scenes (t (23) = 2.83, p < 0.005) and the texture inconsistent scenes (t (23) = 4.55, p < 0.001) (Fig. 6a). No additional significant interaction effects were found (all F's < 1.00). In fact, Naturalness was the only significant factor to modulate the P2 amplitude, with the exception of a marginal effect of Hemisphere (F (1,22) = 3.97, MSE = 12.73, p < 0.06). Thus, foreground texture does not modulate the P2 sensitivity to whether the scene is manmade or natural, suggesting that P2 amplitude can be considered an index of Naturalness that is by and large independent of the local texture information.

3.2.2. N1 component

An analysis of the N1 component revealed that in contrast to the P2, which was not influenced by our manipulation of scene texture, N1 amplitude was substantially impacted by it. First, Texture Consistency had a significant main effect (F (1,22) = 19.74, MSE = 0.87, p < 0.0001), with more negative amplitude to the texture-inconsistent scenes (M = 0.13 mV, SE = 0.74) than to the texture-consistent scenes (M = 0.74 mV, SE = 0.75). More importantly, however, Texture

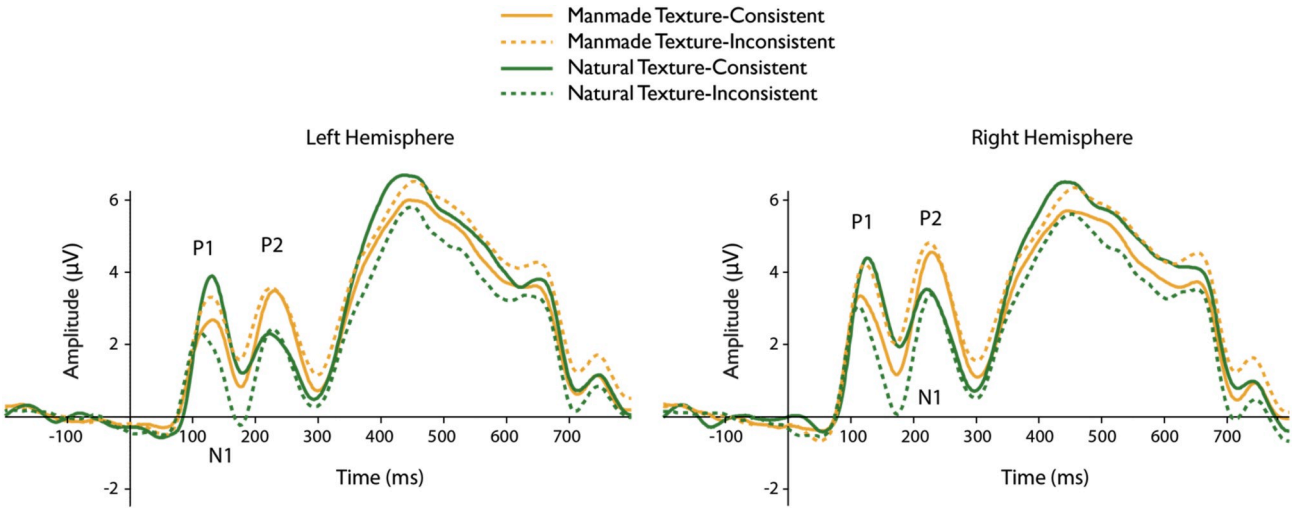

Fig. 5. Group-averaged waveforms for the manmade and natural scenes across the two levels of GSP naturalness consistency plotted for the left and right hemisphere at posterior lateral sites. Yellow and green solid lines represent activity for manmade and natural scenes containing consistent textures, respectively, while yellow and green stippled lines represent activity for manmade and natural scenes containing inconsistent textures, respectively. Note that initially, around the first 200 ms, the lines separate by local texture (natural textures have higher values than manmade textures), while after that, at the P2 level the lines separate by the global dimension of naturalness (manmade scenes have higher values than natural scenes), independent of local texture. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

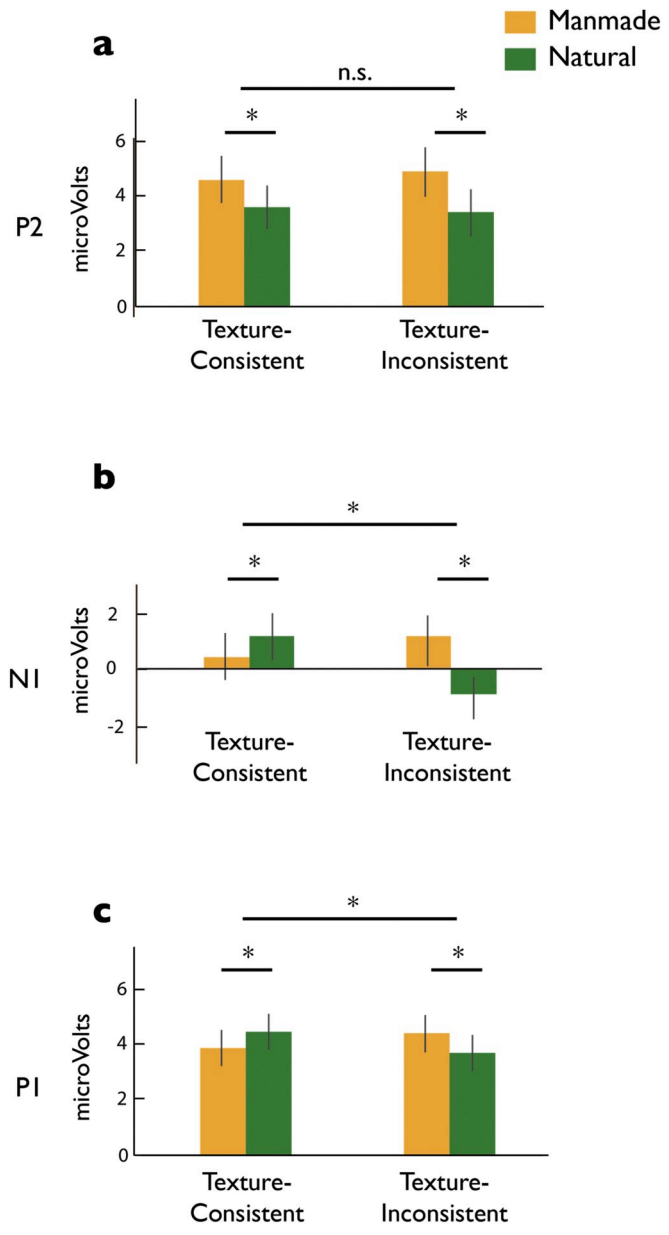

Fig. 6. Grand average ERP analysis results for Experiment 2. Mean peak amplitudes of the P2, N1 and P1components in response to manmade and natural scenes (yellow and green, respectively) presented separately for the Naturalness-consistent (left column) and Naturalness-inconsistent (right column) conditions. To facilitate visualization of all three ERP components, data from the consistent and inconsistent conditions are plotted on the same graph even when the interaction between Naturalness and Naturalness-Consistency was not significant (i.e. P2 component, top row: the long bar refers to the Naturalness-by-Consistency interaction and not to the main effect of consistency). Also note that while the N1 peak amplitude is negative-going compared to the peaks of the P2 and P1 components (see Fig. 5), some conditions have positive N1 values and others have negative values. This was not the case in Experiment 1 and that is why negative values are plotted downward for the N1 results in this graph. Regardless, when interpreting the N1 results from both experiments, the focus should be on which condition had a more negative peak amplitude. Data are collapsed across hemispheres. Significant differences ($p < 0.05$) between pairs of categories are denoted by an asterisk (error bars indicate between-subjects SE). All data are plotted for the posterior lateral sites. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Consistency significantly interacted with the Naturalness of the scene ($F(1,22) = 39.45$, $MSE = 1.98$, $p < 0.0001$). As a follow-up of this interaction, we assessed the effect of Naturalness separately in each Texture Consistency condition (consistent, inconsistent). We expected the texture-consistent scenes to elicit a comparable effect to that observed in Experiment 1 (as they are essentially the same scenes used in Experiment 1, the only difference being that the caves were not used in Experiment 2), namely, a more negative N1 amplitude to the manmade scenes compared with the natural scenes. The critical question was how would this effect play out in the texture-inconsistent scenes. If the Naturalness effect of this earlier component reflects the processing of local texture information than we should expect a reversal – a more negative response to the natural scenes, since they contain prominent manmade textures, than the manmade scenes, which now contain natural textures. This is exactly what we observed. For the texture-consistent scenes, N1 amplitude was significantly more negative to the manmade scenes ($M = 0.40$ mV, $SE = 3.68$) than to the natural scenes ($M = 1.09$ mV, $SE = 3.68$) ($t(22) = 1.95$, $p < 0.04$), replicating the results of Experiment 1. In contrast, in the texture-inconsistent scenes we found the opposite pattern, with N1 amplitude significantly more negative to the natural scenes ($M = -0.83$ mV, $SE = 3.31$) than to the manmade scenes ($M = 1.09$ mV, $SE = 3.91$) ($t(22) = 6.20$, $p < 0.0001$) (Fig. 6b). We also found a significant main effect of Naturalness ($F(1,22) = 5.70$, $MSE = 3.08$, $p < 0.03$), but this effect is of less consequence, given the disordinal interaction between Texture Consistency and Naturalness: the direction of the Naturalness effect is qualified by the extent to which the texture is commensurate with the overall scene category.

3.2.3. *P1 component*

Similar to the N1 component, a significant Naturalness by Texture Consistency effect ($F(1,22) = 8.73$, $MSE = 2.13$, $p < 0.007$) was observed on the P1 amplitude. This was the only significant effect on P1 amplitude. Post-hoc analyses revealed that, similar to the N1 results, the effect of Naturalness reversed when the texture consistency of the scene was manipulated. Specifically, contrasting between manmade and natural scenes in the texture-consistent condition manifested in a significant Naturalness effect ($t(22) = 2.35$, $p < 0.02$), with natural scenes evoking a higher P1 amplitude ($M = 4.50$, $SE = 0.48$) than the manmade scenes ($M = 3.93$, $SE = 0.58$), replicating the P1 findings of Experiment 1. Notably, when the texture was inconsistent with the semantic category, the effect flipped ($t(22) = 1.82$, $p < 0.04$), that is, the natural scenes now evoked a lower P1 response ($M = 3.73$, $SE = 0.55$) compared to the manmade scenes ($M = 4.44$, $SE = 0.58$) (Fig. 6c). Put differently, natural textures evoke a higher P1 amplitude relative to manmade textures, irrespective of the global scene background in which they are embedded. When the natural texture is part of natural scenes, this results in a higher P1 response to natural scenes, but when embedded in manmade scenes this results in higher response to the manmade scenes. Thus, unlike P2 which shows a naturalness effect that is independent of texture, the P1 shows a texture effect that is independent of scene naturalness.

Together, the results of Experiment 2 show that neural sensitivity to the GSP of naturalness cannot be simply explained by or reduced to the processing of lower-level scene texture in the P2 component. In contrast, at very early stages of visual processing (at the level of P1 and N1 components), texture consistency modulates the neural response to manmade and natural scenes, to the extent that the scene texture consistency determines the direction of the effect of naturalness. However, by 220 ms post stimulus onset, texture consistency ceases to have an impact on the neural response to scene naturalness. In other words, the P2 indices global processing of scene naturalness, irrespective of local texture information.

4. Discussion

The present study examined the extent to which early ERP responses to global scene properties (GSPs) are sensitive to low- and high-level scene features. To address this question, we perturbed low- and high-level sources of scene information by presenting our participants with computer-generated artificial scenes, which were devoid of color and rich semantic detail, respectively. We found that early, visually-evoked GSP responses could still be obtained in spite of the removal of color and detailed semantic scene content – arguably, two very salient types of scene information. We found that Spatial Expanse and Naturalness had a robust impact on the amplitude of the P1, N1 and P2 components in Experiment 1: P1 and P2 amplitudes were higher, and N1 amplitude was more negative in response to closed than open scenes. Scene naturalness also modulated the early components, with higher P2 amplitude in response to manmade than natural scenes, higher P1 amplitude to natural than manmade scenes, and N1 showed a weaker modulation by scene naturalness in the right hemisphere. As revealed in Experiment 2, local texture information was to a large extent the main driver of the naturalness effect at the earlier processing stages: P1 amplitude was higher in response to natural textures than manmade textures, and N1 amplitude was more negative to manmade textures, irrespective of the type of scene they were embedded in (manmade or natural). In contrast, however, P2 sensitivity to the naturalness of the scenes was not perturbed by manipulations of local texture, further strengthening the notion that the P2 level (i.e. 220 ms post-stimulus onset) represents the time window in which GSPs are extracted from the scene.

Together, our findings demonstrate that the neural signatures underlying the early extraction of global scene information can be observed even when prominent sources of non-global scene information are absent from the scene. This of course does not mean we have eliminated (and in fact, can ever eliminate if we want to maintain ecological validity) all non-global factors (physical image properties in particular), and it does not imply that lower-level visual information plays no role whatsoever in the extraction of GSPs. But it does support the idea that early ERP responses to GSPs are resilient to manipulation of at least two non-global factors (i.e., color and semantic object details). Furthermore, the earlier components (P1, N1) are more sensitive to lower-level visual information such as texture, whereas P2 is not, and is likely where GSPs are extracted without influence of low-level visual information.

Introduced by Oliva and co-authors (Greene and Oliva, 2009b, 2009a; Ross and Oliva, 2010), GSPs can be considered as scene primitives (i.e., essential image elements) that contain crucial, ecological information about scenes, particularly their spatial structure, constancy and function. Notably, Oliva and colleagues argued that these scene primitives are global in that they do not necessitate the extraction of information about specific objects or confined locations. Global information has been suggested to be processed rapidly and pre-attentively in a mandatory, stimulus-driven fashion to support a subsequent, more elaborate visual analysis (Bar, 2003; Bar et al., 2006; Fabre-Thorpe, 2011; Hochstein and Ahissar, 2002). Previous research using a variety of behavioral paradigms has shown that global information is essential for scene perception, and is processed either in parallel with or even prior to local scene information (Brady et al., 2017; Kauffmann et al., 2015; Peyrin et al., 2006; Schyns and Oliva, 1994).

The findings from the current study are in line with the results from two previous ERP studies using naturalistic scene images in showing that early, visually-evoked potentials, particularly the P2 component, capture the global structure of scenes. In the first study, Harel et al. (2016) reported that the posterior P2 component, peaking around 220 ms after stimulus onset, contains scene information at three levels of increasing granularity: First, the amplitude of the P2 component was more positive in response to scenes than to other complex visual categories, such as faces and common objects. Second, P2 amplitude not only distinguished between scenes and other visual categories, but was also sensitive to distinctions within scenes: it was higher to closed natural scenes than open natural scenes, and higher to natural than manmade scenes (in contrast to the opposite direction observed here; see below). Third, P2 amplitude was not only diagnostic of the naturalness and spatial expanse of scenes at the average level (i.e. across scene exemplars), but was also diagnostic at the level of individual scenes. Specifically, the variance in P2 response to individual scene images was significantly explained by both summary image statistics approximating the GSPs of naturalness and spatial expanse, and by subjective behavioral ratings of the scenes based on their GSPs. In a second study, Hansen and co-authors (2018) used the same scene images as in Harel et al. (2016), replicating the finding that P2 amplitude is modulated by naturalness and spatial expanse, and demonstrating that the extraction of global scene information occurs in a mandatory, stimulus-driven fashion. This was based on the finding that P2 response to the two GSPs was not overly modulated by the behavioral relevance of the particular GSP to the experimental task. Our current findings are mostly consistent with the results of these studies, supporting the idea that P2 amplitude can be used as a marker for the scene's spatial expanse and naturalness (the different directionality of P2 results for naturalness will be discussed in further detail below).

Notably, the current set of minimalistic, artificially-generated scene stimuli is very different than the set of scene stimuli used in the Harel et al. (2016) and Hansen et al. (2018) studies. Theirs was a very rich stimulus set of 96 naturalistic colorful images, spanning sixteen different basic-level scene categories. In contrast, the current study employed a relatively homogenous set of grayscale scene stimuli, comprising only four generic scene categories (a room, a rooftop, a cave, and a barren 'desert' landscape). The scenes were lacking prominent objects and other various details which often characterize naturalistic scene images (the main source of variation across the individual scenes being the specific combinations of layout and texture). However, in spite of these differences in appearance, robust modulations of the P2 amplitude were found across all three studies for both spatial expanse and naturalness, demonstrating that GSP-related neural responses can be observed across substantial variations in physical and semantic scene properties. In particular, Spatial Expanse manifested similarly for the full-color real-world scenes and grayscale artificial scenes, with closed scenes evoking a higher P2 amplitude than open scenes.

As for naturalness, it also modulated the P2 amplitude for both real-world and artificial scenes, but interestingly, whereas in the former case natural scenes evoked a higher amplitude than manmade scenes, in the latter case manmade scenes evoked a higher response than natural scenes. These opposite directions of the naturalness effect might reflect differential contributions of local image properties to the two sets of scene stimuli, which nonetheless separate the manmade from the natural scenes. This possibility is in line with an alternative, low-level account of GSPs, which considers GSPs as diagnostic for scene recognition only to the extent that they are consistent with differences in local, low-level visual features (e.g. Loschky and Larson, 2010). Addressing this possibility was our main motivation for conducting Experiment 2, in which we tested the role that local texture information plays in driving the P2 naturalness effect. We found that the naturalness effect on P2 amplitude (manmade greater than natural scenes) was the same when using texture consistent scenes (i.e., the scenes from Experiment 1) and texture inconsistent scenes (i.e., using prominent manmade textures in natural scenes and vice versa), while in contrast, the earlier visually-evoked components (P1, N1) reversed the direction of their naturalness effect for texture inconsistent scenes, demonstrating sensitivity to local textures. This finding further strengthens the idea that P2 amplitude captures the extraction of intermediate-level, global scene information whereas the P1 and N1 effects are more sensitive to low-level visual information. The reason why the P2 naturalness effect reversed direction compared with previous studies still needs to be determined. Indeed, we haven't explored all possible low-level image properties that might have contributed to the difference in the direction of the naturalness effect. Recent research, for instance, has focused on

non-accidental scene properties as critical for scene categorization, primarily contour junctions (Choo and Walther, 2016; Walther and Shen, 2014), the spatial location of contour junctions (Wilder et al., 2018), and their symmetry (Wilder et al., 2019). It might be possible that the natural and manmade scenes in our current study differed along any of these nonaccidental properties, and this difference might have manifested in an opposite fashion compared with the scenes used in previous studies. It should be noted, though, that the task of elucidating the exact differences in stimulus properties that would counteract each other is not a straightforward one. For example, the manmade scenes of the current study differed in their contrast energy from the natural scenes, but not in their spatial coherence (Mzozoyana et al., 2017). In contrast, the real-world scenes differed in their spatial coherence, but not in their contrast energy (Harel et al., 2016). Moreover, the manmade- and natural-artificial scenes also differed in their Fourier spectrum properties, while no such difference existed between the manmade and natural real-world scenes (Harel et al., 2016; Mzozoyana et al., 2017). In addition, differences in image format (i.e., natural photographs vs. artificially-generated images), or experimental design (i.e., naturally occurring variations in scene texture and layout vs. experimentally controlled variations in these scene features) could be responsible for the reversal of the P2 naturalness effect in the present study. Future research will be required to elucidate the underlying nature of this reversal, particularly the extent to which the explanation for this reversal is weighted more towards low-level image properties or intermediate-level GSPs. Irrespective of the specific direction of the P2 naturalness effect, it is still notable in the current context that a significant difference between manmade and natural scenes was manifest not only in response to rich real-world scenes, but also in response to artificially-generated scenes deprived of color and object content.[2]

We suggest that the tolerance to changes in scene appearance evident in the P2 responses to the GSPs of spatial expanse and naturalness reflects the extraction of diagnostic scene information, which generalizes across variations in scene images and emphasizes the significance of intermediate-level representations to scene recognition. Intermediate-level representations have the computational advantage of being more informative than simple local features for describing large-scale scenes, which vary in the distribution of information across the entirety of the scene. At the same time, intermediate-level representations are also economical in their computational demands – a minimal amount of visual processing of scene properties such as structural layout and the potential for action is enough to capture the key essence of a scene (Brady et al., 2017; Greene and Oliva, 2009a; Aude Oliva and Torralba, 2001). This makes global intermediate representations important for scene recognition, as these representations can retain essential information for scene processing across a variety of changes in image appearance. We therefore argue that as long as the scene contains these key sources of information (i.e. GSPs), even when other sources of information are absent from the scene (e.g., color, texture, foreground objects), one should expect scene recognition to be relatively unperturbed, both at the behavioral and neurophysiological level.

Further support for this notion comes from two recent studies. First, Brady, Shafer-Skelton, and Alvarez (2017) examined the role of global ensemble textures (spatial patterns of orientation and spatial frequency information) in rapid scene categorization, and reported that priming effects of scene background on object recognition could be observed even when the scene was unrecognizable (i.e. when the semantic meaning of scenes was eliminated), as long as the global ensemble texture information was preserved (see their Experiment 2). In other words, a scene still exerts its 'semantic' impact, as long as it retains its global spatial properties. Moreover, Brady et al. (2017, Experiment 1) demonstrated that sensitivity to changes in spatial ensemble structure was strongly correlated with the faciliatory effect of the scene background on object recognition, but neither of these measures correlated with a measure of object-based summary statistics (orientation). These two experiments converge on the idea that global scene information is intrinsically linked with scene recognition, and importantly, that it can be distinguished from low- and high-level sources of information. Second, more direct evidence for the central role that GSPs play in scene recognition comes from a series of behavioral adaptation experiments (Greene and Oliva, 2010). The idea at the base of this work is that the environmental regularities captured by GSPs are encoded at early stages of visual processing, and therefore should evoke perceptual aftereffects. Indeed, GSPs like spatial expanse, naturalness, and navigability produced substantial aftereffects. Notably, these aftereffects were also evident across changes in stimulus location (i.e., position-invariant), meaning that the aftereffects are not the result of adaptation to low-level properties of the stimulus. Lastly, a further demonstration of the close link between GSP processing and basic-level scene categorization came from the finding that adaptation to particular values of spatial expanse (open/closed) led to a change in the subjective perception of the scene category (either as a field or a forest, respectively). Thus, in spite of differences in methodology and experimental paradigms, these two behavioral studies combined with our current electrophysiological study all converge on the conclusion that GSPs are encoded during the early stages of scene perception, are heavily involved in the perception of scene gist, and provide a means of bridging across low- and high-level sources of scene information by preserving global spatial information.

While thus far we have focused on P2 as the main time window in which GSP processing is manifest, we also found earlier visually-evoked ERP components to be sensitive to the GSPs contained in the artificially-manipulated scene stimuli. Specifically, significant effects of Naturalness and Spatial Expanse were observed on the amplitude of both P1 and N1 components. The P1 and N1 effects indicate that GSPs may be extracted earlier than 220 ms, putatively as early as 120–150 ms after stimulus onset, although this seems to reflect the impact of GSPs to the extent that they are associated with low-level image features, such as texture. The results of Experiment 1 could not unambiguously determine whether global (i.e., scene category) or local (i.e., texture) information was driving the P1 (and N1) naturalness effect. Teasing apart these two factors in Experiment 2 revealed that the determining factor is the texture itself rather than scene category, for both P1 and N1. Given that the first visually-evoked ERP component observed at lateral occipital electrode sites, P1, is especially sensitive to variations in physical stimulus properties such as contrast and luminance (for a review, see Luck, 2014), we suggest that the P1 sensitivity to the scene's naturalness and spatial expanse may thus reflect the discrimination of GSPs based on low-level image features other than color. Our findings are in line with previous studies reporting early effects of Spatial Expanse and Naturalness, some recording from midline occipital electrodes (Groen et al., 2013; Lowe et al., 2018) and others recording from the same posterior lateral electrodes reported here (Hansen et al., 2018; Harel et al., 2016). A similar account was made by Lowe et al. (2018) who found that the onset of scene naturalness and spatial expanse discrimination starts as early as the P1 time window, with discrimination activity extending to the N1–P2 time window. Lowe at al. suggested that low-level image statistics are essential for scene perception, insofar as they capture the holistic and diagnostic structure of a scene.

It is important to note that we do not argue in the present work that the early GSP-related responses (i.e., P2) are driven *exclusively* by global information. Rather, we consider GSPs as a way of bridging the large gap

[2] Parenthetically, it is interesting to note that the recent rise in multivariate decoding approaches puts the directionality difference in the naturalness effect into a broader context. To yield significant classification accuracy, such decoding approaches (e.g. linear discriminant analysis; Parra et al., 2005) require establishing a difference between the patterns of neural activity in response to two or more conditions. This means that a difference, in general, is an important source of information regardless of the direction of the difference (Mur et al., 2009).

between the scene's low-level physical properties (e.g. image contrast) on the one hand, and higher-level observer-based ecological properties (e.g. navigability affordances), on the other. Both ends of the representational continuum have been demonstrated to contribute to scene understanding (for recent reviews and discussion, see Malcom et al., 2016; Groen et al., 2017), which ultimately leads to the following question: How does the visual system make use of both types of scene information? How do the two (or more) sources of information get integrated? We suggest that to understand this question, one must consider scene understanding as a process that is achieved through multiple routes; oftentimes these multiple routes to scene understanding are indeed related, but oftentimes, they traverse standard definitions of what constitutes low- or high-level information. This is because in scene recognition, in contrast to other visual domains, such as object recognition or word recognition, the different sources of information cannot be easily teased apart. For example, beach scenes have a higher likelihood to be dominated by low spatial frequencies, due to the presence of a prominent horizontal boundary and large homogeneous sections (sky and beach), resulting in an open spatial layout (Groen et al., 2017). Thus, given that these sources of information often co-vary and evince many self-similarities, it seems ineffective to reduce scene recognition to just a single scene dimension (be it low-level or high-level, for that matter), as this provides at best a partial, incomplete picture of scene understanding. Instead, one should study the joint combinations of multiple factors, which together determine the nature of scene recognition (as well as the circumstances under which they become diagnostic for the specific task at hand; see Lowe et al., 2016). We suggest that GSPs serve this exact function, as they capture both ends of the information spectrum: on the one hand, they are related to several summary image statistics, but on the other hand they are closely related to behavior and reflect high-level ecological properties of organization. At the neural level, this means that for an ERP component such as the P2 to be considered GSP-sensitive, it needs to be partially based on the extraction of low-level visual information (e.g., for naturalness, some image statistic diagnostic of the difference between natural and man-made scenes), but not at a local spatial scale. That is, GSPs may be sensitive to low-level visual features, but only at a broader spatial scale (i.e., global!). Thus, a GSP-sensitive component has to be sensitive to some image properties (importantly, "low-level" image-properties should not be equated with local features), while at the same time invariant to other image properties as long as the global layout of the scene is retained. These mutliple contributions from visual features at all stages of the visual-processing hierarchy simply reflect the complex nature of scene perception.

Indeed, the Harel et al. (2016) study (see above) showed that while summary image statistics (contrast and Fourier spectrum measures) could explain a substantial amount of the variance of the P2 amplitude, adding behavioral ratings of the spatial expanse and naturalness of the same scene images (i.e. participants' subjective perception of the GSPs) contributed a substantial amount of explanatory power, with GSP behavioral ratings explaining a notable amount of the P2 variance above and beyond the image statistics alone. This directly supports our idea that while low-level image statistics are correlated with GSPs, image statistics alone are not sufficient to explain the early neural responses to scenes. Further support for this idea comes from a recent computational neuroimaging study (Lescroart et al., 2015). The authors aimed to determine the nature of representations in scene-selective areas PPA, OPA and RSC by applying voxel-wise modeling to BOLD fMRI responses to scene images, testing three competing models of scene recognition: (1) using 2D Fourier power features, (2) using 3D spatial features (e.g. relative distance), or (3) using abstract features (e.g. scene category). Critically, the authors found – consistent with the Harel et al. (2016) findings - that the response variance explained by these three models is largely shared, and that the individual models actually explain very little unique variance. The implication of this finding is that one cannot a-priori consider one level of representation as more privileged than the other, and further, it emphasizes the idea suggested above that scene information is extracted at multiple levels. Future research is required to determine the exact nature of the interplay between the different sources of scene information. A promising direction would be to manipulate low-level (and intermediate- and high-level, for that matter) visual information at both local and global spatial scales and examine their impact on the P2 response to manmade and natural scenes. Of additional importance will be electrophysiological studies revealing how these factors unfold over time. For example, a potential study could examine the extent to which the presence or absence of diagnostic, semantically-related objects modulates the early scene-evoked responses (for a fMRI analog, see Harel et al., 2013).

In summary, we show here that the early electrophysiological responses to artificially-generated scenes carry robust information about the global properties of scenes, distinguishing scenes based on their spatial expanse and naturalness. At the level of P1 and N1, this is likely driven by lower-level visual information (e.g., texture, summary image statistics), but the extraction of GSPs at the P2 component is likely independent of lower-level sources of visual information and instead reflects intermediate-level features. The GSP effects observed with such 'impoverished' artificially-generated scenes resemble in their magnitude and latency the same effects previously observed with realistic, real-world scenes, particularly for spatial expanse. Thus, our findings provide support for the notion that global scene information is central to scene recognition, and further emphasize the utility of ERPs as a means to elucidate the temporal dynamics underlying the extraction of GSPs in scene perception and recognition.

CRediT authorship contribution statement

Assaf Harel: Conceptualization, Methodology, Formal analysis, Writing - original draft, Writing - review & editing, Visualization, Supervision. **Mavuso W. Mzozoyana:** Investigation, Formal analysis. **Hamada Al Zoubi:** Investigation, Formal analysis. **Jeffrey D. Nador:** Software, Investigation, Formal analysis. **Birken T. Noesen:** Software, Investigation. **Matthew X. Lowe:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Visualization. **Jonathan S. Cant:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Visualization.

References

Andrews, T.J., Watson, D.M., Rice, G.E., Hartley, T., 2015. Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. J. Vis. 15 (7), 3.

Bar, M., 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. J. Cognit. Neurosci. 15 (4), 600–609.

Bar, M., Kassam, K.S., Ghuman, A.S., Boshyan, J., Schmidt, A.M., Dale, A.M., Halgren, E., 2006. Top-down facilitation of visual recognition, 103(2). In: Proceedings of the National Academy of Sciences of the United States of America, pp. 449–454.

Bonner, M.F., Epstein, R.A., 2017. Coding of navigational affordances in the human visual system, 114(18). In: Proceedings of the National Academy of Sciences, pp. 4793–4798.

Brady, T.F., Shafer-Skelton, A., Alvarez, G.A., 2017. Global Ensemble Texture Representations Are Critical to Rapid Scene Perception.

Choo, H., Walther, D.B., 2016. Contour junctions underlie neural representations of scene categories in high-level human visual cortex. Neuroimage 135, 32–44.

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. Retrieved from. http://www.sciencedirect.com/science/article/pii/S1053811916300076.

Çukur, T., Huth, A.G., Nishimoto, S., Gallant, J.L., 2016. Functional subdomains within scene-selective cortex: parahippocampal place area, retrosplenial complex, and occipital place area. J. Neurosci. 36 (40), 10257–10273.

Dilks, D.D., Julian, J.B., Paunov, A.M., Kanwisher, N., 2013. The occipital place area is causally and selectively involved in scene perception. J. Neurosci. 33 (4), 1331–1336.

Fabre-Thorpe, M., 2011. The characteristics and limits of rapid visual categorization. Front. Psychol. 2.

Ferrara, K., Park, S., 2016. Neural representation of scene boundaries. Neuropsychologia 89, 180–190.

Greene, M.R., Fei-Fei, L., 2014. Visual categorization is automatic and obligatory: evidence from Stroop-like paradigm. J. Vis. 14 (1), 14. https://doi.org/10.1167/14.1.14. Retrieved from.
Greene, M.R., Oliva, A., 2009a. Recognition of natural scenes from global properties: seeing the forest without representing the trees. Cognit. Psychol. 58 (2), 137–176.
Greene, M.R., Oliva, A., 2009b. The briefest of glances the time course of natural scene understanding. Psychol. Sci. 20 (4), 464–472.
Greene, M.R., Oliva, A., 2010. High-level aftereffects to global scene properties. J. Exp. Psychol. Hum. Percept. Perform. 36 (6), 1430.
Groen, I.I.A., Ghebreab, S., Lamme, V.A.F., Scholte, H.S., 2012. Spatially pooled contrast responses predict neural and perceptual similarity of naturalistic image categories. PLoS Comput. Biol. 8 (10), e1002726 https://doi.org/10.1371/journal.pcbi.1002726.
Groen, I.I.A., Ghebreab, S., Lamme, V.A.F., Scholte, H.S., 2016. The time course of natural scene perception with reduced attention. J. Neurophysiol. 115 (2), 931–946.
Groen, I.I.A., Ghebreab, S., Prins, H., Lamme, V.A.F., Scholte, H.S., 2013. From image statistics to scene gist: evoked neural activity reveals transition from low-level natural image structure to scene category. J. Neurosci. 33 (48), 18814–18824.
Groen, I.I.A., Greene, M.R., Baldassano, C., Fei-Fei, L., Beck, D.M., Baker, C.I., 2018. Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. Elife 7, e32962.
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. Phil. Trans. Biol. Sci. 372 (1714).
Gronau, N., Izoutcheev, A., 2017. The necessity of visual attention to scene categorization: dissociating "task-relevant" and "task-irrelevant" scene distractors. J. Exp. Psychol. Hum. Percept. Perform. 43 (5), 954.
Hansen, B.C., Jacques, T., Johnson, A.P., Ellemberg, D., 2011. From spatial frequency contrast to edge preponderance: the differential modulation of early visual evoked potentials by natural scene stimuli. Vis. Neurosci. 28 (3), 221–237. https://doi.org/10.1017/S095252381100006X.
Hansen, B.C., Johnson, A.P., Ellemberg, D., 2012. Different spatial frequency bands selectively signal for natural image statistics in the early visual system. J. Neurophysiol. 108 (8), 2160–2172. https://doi.org/10.1152/jn.00288.2012.
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.
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 (5). ENEURO-0139.
Harel, A., Kravitz, D.J., Baker, C.I., 2013. Deconstructing visual scenes in cortex: gradients of object and spatial layout information. Cerebr. Cortex 23 (4), 947–957.
Hochstein, S., Ahissar, M., 2002. View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36 (5), 791–804.
Hollingworth, A., 2004. Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. J. Exp. Psychol. Hum. Percept. Perform. https://doi.org/10.1037/0096-1523.30.3.519. Hollingworth, Andrew: University of Iowa, Department of Psychology, 11 Seashore Hall E., Iowa City, IA, US, 52242-1407, andrew-hollingworth@uiowa.edu: American Psychological Association.
Hollingworth, A., 2005. The relationship between online visual representation of a scene and long-term scene memory. J. Exp. Psychol.: Learning, Memory, and Cognition. Hollingworth, Andrew. https://doi.org/10.1037/0278-7393.31.3.396. Department of Psychology, The University of Iowa, 11 Seashore Hall E, Iowa City, IA, US, 52242-1407, andrew-hollingworth@uiowa.edu: American Psychological Association.
Intraub, H., 1981. Rapid conceptual identification of sequentially presented pictures. J. Exp. Psychol. Hum. Percept. Perform. 7 (3), 604.
Joubert, O.R., Rousselet, G.A., Fize, D., Fabre-Thorpe, M., 2007. Processing scene context: fast categorization and object interference. Vis. Res. 47 (26), 3286–3297.
Jung, T.-P., Humphries, C., Lee, T.-W., Makeig, S., McKeown, M.J., Iragui, V., Sejnowski, T.J., 1998. Extended ICA removes artifacts from electroencephalographic recordings. Adv. Neural Inf. Process. Syst. 894–900.
Kauffmann, L., Chauvin, A., Guyader, N., Peyrin, C., 2015. Rapid scene categorization: role of spatial frequency order, accumulation mode and luminance contrast. Vis. Res. 107, 49–57. https://doi.org/10.1016/j.visres.2014.11.013.
Konkle, T., Brady, T.F., Alvarez, G.A., Oliva, A., 2010. Scene memory is more detailed than you think. Psychol. Sci. 21 (11), 1551–1556. https://doi.org/10.1177/0956797610385359.
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. J. Neurosci. 31 (20), 7322–7333.
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. Front. Comput. Neurosci. 9.
Li, F.F., VanRullen, R., Koch, C., Perona, P., 2002. Rapid natural scene categorization in the near absence of attention, 99(14). In: Proceedings of the National Academy of Sciences, pp. 9596–9601. Retrieved from. http://www.pnas.org/content/99/14/9596.abstract.
Loschky, L.C., Larson, A.M., 2010. The natural/man-made distinction is made before basic-level distinctions in scene gist processing. Vis. Cognit. 18 (4), 513–536.
Lowe, M.X., Gallivan, J.P., Ferber, S., Cant, J.S., 2016. Feature diagnosticity and task context shape activity in human scene-selective cortex. Neuroimage 125, 681–692.
Lowe, M.X., Rajsic, J., Ferber, S., Walther, D.B., 2018. Discriminating scene categories from brain activity within 100 milliseconds. Cortex 106, 275–287.
Luck, S.J., 2014. An Introduction to the Event-Related Potential Technique, second ed. MIT press.
Malcolm, G.L., Groen, I.I.A., Baker, C.I., 2016. Making sense of real-world scenes. Trends Cognit. Sci. 20 (11), 843–856.
Mur, M., Bandettini, P.A., Kriegeskorte, N., 2009. Revealing representational content with pattern-information fMRI—an introductory guide. Soc. Cognit. Affect Neurosci. 4 (1), 101–109. https://doi.org/10.1093/scan/nsn044.
Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., Peyrin, C., 2014. Coarse-to-fine categorization of visual scenes in scene-selective cortex. J. Cognit. Neurosci. 26 (10), 2287–2297. https://doi.org/10.1162/jocn_a_00643.
Mzozoyana, M., Lowe, M., Groen, I., Cant, J., Harel, A., 2017. Artificially-generated scenes demonstrate the importance of global scene properties for scene perception. J. Vis. 17 (10), 312. https://doi.org/10.1167/17.10.312.
Oliva, A., Schyns, P.G., 2000. Diagnostic colors mediate scene recognition. Cognit. Psychol. 41 (2), 176–210. https://doi.org/10.1006/cogp.1999.0728S0010-0285(99)90728-4 ([pii]).
Oliva, Aude, Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. 42 (3), 145–175.
Oliva, Aude, Torralba, A., 2006. Chapter 2 Building the gist of a scene: the role of global image features in recognition. Prog. Brain Res. 155, 23–36. https://doi.org/10.1016/S0079-6123(06)55002-2.
Park, S., Brady, T.F., Greene, M.R., Oliva, A., 2011. Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. J. Neurosci. 31 (4), 1333–1340.
Park, S., Konkle, T., Oliva, A., 2014. Parametric coding of the size and clutter of natural scenes in the human brain. Cerebr. Cortex 25 (7), 1792–1805.
Parra, L.C., Spence, C.D., Gerson, A.D., Sajda, P., 2005. Recipes for the linear analysis of EEG. Neuroimage 28 (2), 326–341.
Peyrin, C., Mermillod, M., Chokron, S., Marendaz, C., 2006. Effect of temporal constraints on hemispheric asymmetries during spatial frequency processing. Brain Cognit. 62 (3), 214–220.
Potter, M.C., 1976. Short-term conceptual memory for pictures. J. Exp. Psychol. Hum. Learn. Mem. 2 (5), 509.
Potter, M.C., Levy, E.I., 1969. Recognition memory for a rapid sequence of pictures. J. Exp. Psychol. 81 (1), 10.
Ross, M.G., Oliva, A., 2010. Estimating perception of scene layout properties from global image features. J. Vis. 10 (1), 2. https://doi.org/10.1167/10.1.2. Retrieved from.
Rousselet, G., Joubert, O., Fabre-Thorpe, M., 2005. How long to get to the "gist" of real-world natural scenes? Vis. Cognit. 12 (6), 852–877.
Schyns, P.G., Oliva, A., 1994. From blobs to boundary edges: evidence for time-and spatial-scale-dependent scene recognition. Psychol. Sci. 5 (4), 195–200.
Stansbury, D.E., Naselaris, T., Gallant, J.L., 2013. Natural scene statistics account for the representation of scene categories in human visual cortex. Neuron 79 (5), 1025–1034.
Tanaka, J., Weiskopf, D., Williams, P., 2001. The role of color in high-level vision. Trends Cognit. Sci. 5 (5), 211–215.
Tversky, B., Hemenway, K., 1983. Categories of environmental scenes. Cognit. Psychol. 15 (1), 121–149.
Velisavljević, L., Elder, J.H., 2008. Visual short-term memory of local information in briefly viewed natural scenes: configural and non-configural factors. J. Vis. 8 (16), 8. https://doi.org/10.1167/8.16.8. Retrieved from.
Võ, M.L.-H., Henderson, J.M., 2009. Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. J. Vis. 9 (3), 24.
Võ, M.L.-H., Wolfe, J.M., 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. Psychol. Sci. 24 (9), 1816–1823.
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. J. Neurosci. 29 (34), 10573–10581.
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, 108(23). In: Proceedings of the National Academy of Sciences, pp. 9661–9666.
Walther, D.B., Shen, D., 2014. Nonaccidental properties underlie human categorization of complex natural scenes. Psychol. Sci. 25 (4), 851–860. https://doi.org/10.1177/0956797613512662.
Wilder, J., Dickinson, S., Jepson, A., Walther, D.B., 2018. Spatial relationships between contours impact rapid scene classification. J. Vis. 18 (8), 1. https://doi.org/10.1167/18.8.1.
Wilder, J., Rezanejad, M., Dickinson, S., Siddiqi, K., Jepson, A., Walther, D.B., 2019. Local contour symmetry facilitates scene categorization. Cognition 182, 307–317.