

2020

## Artificially-Generated Scenes Demonstrate the Importance of Global Properties during Early Scene Perception

Mavuso Wesley Mzozoyana  
*Wright State University*

Follow this and additional works at: [https://corescholar.libraries.wright.edu/etd\\_all](https://corescholar.libraries.wright.edu/etd_all)



Part of the [Neuroscience and Neurobiology Commons](#), and the [Physiology Commons](#)

---

### Repository Citation

Mzozoyana, Mavuso Wesley, "Artificially-Generated Scenes Demonstrate the Importance of Global Properties during Early Scene Perception" (2020). *Browse all Theses and Dissertations*. 2301.  
[https://corescholar.libraries.wright.edu/etd\\_all/2301](https://corescholar.libraries.wright.edu/etd_all/2301)

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact [library-corescholar@wright.edu](mailto:library-corescholar@wright.edu).

ARTIFICIALLY-GENERATED SCENES DEMONSTRATE THE IMPORTANCE OF  
GLOBAL PROPERTIES DURING EARLY SCENE PERCEPTION

A thesis submitted in partial fulfillment of the  
requirements for the degree of  
Master of Science

by

MAVUSO WESLEY MZOZOYANA  
B.A., KENT STATE UNIVERSITY, 2014

2020

Wright State University

WRIGHT STATE UNIVERSITY  
GRADUATE SCHOOL

August 28, 2019

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Mavuso Wesley Mzozoyana ENTITLED Artificially-Generated Scenes Demonstrate the Importance of Global Properties during Early Scene Perception BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

---

Assaf Harel, Ph.D.  
Thesis Director

---

Eric Bennett, Ph.D.  
Chair, Department of  
Neuroscience, Cell Biology, and  
Physiology

Committee on Final Examination:

---

Assaf Harel, Ph.D.

---

Sherif M. Elbasiouny, Ph.D.

---

Joe Houpt, Ph.D.

---

Barry Milligan, Ph.D.  
Interim Dean of the Graduate School

## ABSTRACT

Mzozoyana, Mavuso Wesley. M.S. Department of Neuroscience, Cell Biology, and Physiology, Wright State University, 2020. Artificially-Generated Scenes Demonstrate the Importance of Global Properties during Early Scene Perception.

During scene perception, studies have shown the importance of the global distribution of a scene. Electrophysiological studies have found these global effects concentrated corresponding to the second positive and first negative peaks (P2 and N1, respectively) of the Event-related potential (ERP) during the first 600 ms of scene perception. We sought to understand in Experiment 1, to what extent early responses to scenes were driven by mid-level global information such as the degree of naturalness or openness in a scene image in the absence of specific low-and high-level information (color and semantic object detail). This was done using artificially-generated stimuli controlling for two global scene properties (GSPs) of spatial boundary and naturalness while minimizing color and semantic object information. Significant effects were observed on the P2 and N1 components as well as the P1 component. However, the question of whether scene perception is dominated by global or local factors had yet to be answered leading to Experiment 2. During Experiment 2, for half the trials scenes were presented in an inverted orientation. We found only an orientation interaction approaching significance corresponding to the P1 time course.

## Table of Contents

Introduction .....	1
Experiment 1.....	10
Method & Results .....	12
Participants .....	12
Stimulus and Apparatus.....	12
EEG Recording.....	13
Procedures .....	14
Data processing and analysis.....	15
Results .....	17
P2 Component.....	18
N1 and P1 Components .....	19
Discussion.....	23
Experiment 2 .....	25
Method & Results .....	28
Participants .....	28
Stimulus, Apparatus, and Procedures.....	28
Data processing and analysis.....	29

Results .....	29
P2 Component.....	30
N1 and P1 Components .....	30
Summary and Discussion .....	34
General Discussion.....	35
Summary and Conclusions.....	46
Limitations and Future Directions.....	48
References.....	50

## List of Figures

Figure	Page
Figure 1: ERP waveforms.....	6
Figure 2: Averaged waveform example.....	17
Figure 3a: P2 Experiment 1 Results .....	19
Figure 3b: N1 Experiment 1 Results: .....	20
Figure 3c: P1 Experiment 1 Results: .....	20
Figure 4: Group-averaged waveforms.....	21
Figure 5: Grand average ERP analysis results .....	22
Figure 6a: P2 Experiment 2 Results: .....	30
Figure 6b: N1 Experiment 2 Results: .....	31
Figure 6c: P1 Experiment 2 Results: .....	32
Figure 7: Grand average ERP analysis. ....	33

## List of Tables

Table	Page
Table 1: Experiment 1 Hypotheses .....	11
Table 2: Experiment 2 Hypotheses .....	27
Appendix Table	
Table 3: Summary of results from Experiments 1 and 2.....	60



## **Statement of Significance**

Previous studies on scene recognition have shown low-level information such as color (Oliva & Schyns, 2000) and high-level information such as object categories (Walther et al., 2009) are important for scene recognition. What has yet to be shown is how these two levels of processing can be bridged to support scene recognition. We suggest another level of processing, mid-level global processing, which takes into account is the global distribution of the scene such as the naturalness, openness, ruggedness, expansion, and roughness of a scene serves this function. We tested this by measuring brain responses to scene stimuli that were stripped of color and rich semantic object information while maintaining the global scene properties of naturalness and spatial boundary. The results of the studies prove the importance of mid-level global information during early-scene perception with naturalness and spatial boundary effects impacting the magnitude of early visually evoked potentials (P1, N1, and P2) even when prominent sources of low- and high-level information are absent from the scene.

## **Acknowledgements**

The author would like to express his gratitude to Dr. Assaf Harel, primary advisor and mentor, for his support and patient guidance during the authors two years spent in his lab.

Also, the author would like to express his gratitude towards the other members of his committee, Dr. Sherif M. Elbasiouny and Dr. Joe Houpt for their time and input given to the authors thesis project

Additionally, the author would like to thank his parents, Mbulelo Mzozoyana and Teri Mzozoyana and grandmother, Mildred Sherron, for their continued financial and emotional support throughout the entirety of his education.

Finally, the author would like to thank fellow graduate student Birken Noesen for taking time to code the experiments the author was then able to conduct and former undergraduate student Jordan Keller for assisting the author in preparing the participants for the experiments as well as assisting with data organization and statistical analysis.

## **Introduction**

In the world of scene perception, a scene is defined “as a 3D environment the observer is embedded in and interacts with...allowing the observer to carry out specific behavioral goals, for example locomotion or motor interaction...” (review by Malcolm et al., 2016, pg. 844). Humans have acquired the ability to recognize and perceive scenes very rapidly. This is due to the ability to extract the gist of a scene. Scene gist- providing a basic-level description of a scene such as a desert or city landscape- has been shown to be recognized very quickly and under rapid presentation rates (Joubert et al., 2007; Potter, 1976; Intraub, 1981) with incredibly high accuracy rates (Rousselet et al., 2005).

This rapid extraction of scene gist is done in the almost complete absence of specific information relating to object shape or identity (Oliva & Torralba, 2001). For example, when a person is watching television, and rapidly flipping through the channels. They can determine each picture’s meaning (e.g. a car chase, the news, cartoons, etc.) despite a brief glimpse of each picture independent of detail (Oliva, 2004). Additionally, the process of identifying the gist of a scene has been shown to occur in the near absence of attention (Li et al., 2002) and requiring very minimal attentional resources (Greene & Li, 2014).

Despite the ease with which humans can perceive scenes, scene processing is a complicated computational feat. Particularly, one challenge is the multiple levels of representations along which scenes can be conceptualized. Specifically, scene perception

has been shown to be ordered hierarchically with low, mid, and high-level representations (Groen et al., 2017). Low-level representation involves basic features such as color, local contour, luminance, and spatial frequency. Mid-level representations are less understood (Peirce et al., 2015) but are thought to represent features and properties such as surfaces, textures, shape features, and 3D depth cues (Welchman et al., 2005; Groen et al., 2017) while high-level representation involves extraction of visual input into categorical representations enabling classification (Groen et al., 2017) such as being able to classify an image as being manmade by visually extracting information typically seen in a manmade scene such as a city landscape or streets with their corresponding street names or the ability to classify an image as being natural by visually extracting information typically seen in natural images such as mountains, grassy fields, or a barren landscape.

Many different studies have been conducted in order to find the level of information that contributes the most to rapid scene recognition. Some studies have suggested that low-level properties are what allows for scene recognition. For example, in a study by Rajimehr et al. (2011) reported that high spatial frequency images of scenes were effective in activating the Parahippocampal place area (PPA), a region in the brain shown to respond more to scene images than other object categories (Epstein & Kanwisher, 1998). Another study by Oliva & Schyns (2000) conducted over three experiments showed in the first two experiments when presented with color-diagnostic and color non-diagnostic scenes, faster recognition occurs with color-diagnostic scenes. The third experiment found in images with low-resolution, the low-resolution scene images that were color-diagnostic had elicited higher performance than low-resolution

color-nondiagnostic scene images. This shows the importance color information has in recognizing scene images.

While the above studies have shown the importance of low-level information for scene recognition, studies have also shown the importance of high-level properties in the recognition of scenes. In a study by Walther et al. (2009) they showed that the brain activity in various areas can be used to decode natural scene categories (forests, mountains, beaches) and that natural scene information is primarily contained within the PPA, retrosplenial cortex (RSC), and the lateral occipital complex (LOC). This suggests these regions store high-level information allowing for the recognition of natural scene images.

Furthering the notion of high-level properties driving scene recognition is a study by Li et al. (2002). When presented with natural scenes participants were able to quickly detect animals and vehicles embedded within the natural scene images while occupied with another task demanding their attention. While under the same conditions they were unable to discriminate between the letters T and L or bisected two-color disks from their mirror images under the same conditions. This shows that high-level visual tasks can proceed in the absence of attention.

Despite evidence showing high and low-level information contributing to rapid scene recognition, there is still no clear explanation of how the two levels of processing are integrated to support scene recognition. Incorporating both low and high-levels of processing or being totally independent of either levels of processing, instead relying on

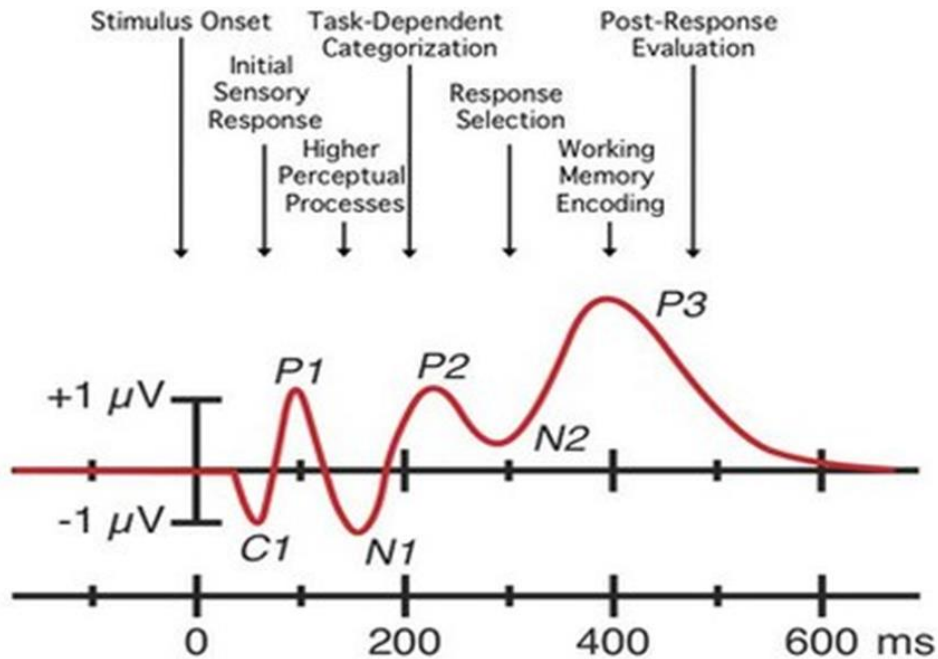
mid-level information to achieve rapid scene recognition. We propose this intermediary is the global distribution of a scene. This is defined as the collection of descriptors applying to the whole image not just a portion of it in which the distribution of textures and edges are encoded in the image as well as their spatial arrangement without segmenting the image (Oliva & Torralba, 2002). Studies conducted have proposed the idea of a spatial envelope (Oliva & Torralba, 2006, 2001) in which details such as an objects shape or identity is not needed for scene perception but instead proposes the holistic representation is essential to scene recognition.

Derived from the spatial envelope are Global scene properties (GSPs) which are “descriptors of the structure and function of real-world scenes” (Hansen et al., 2018, p.3) more appropriately defined as “ecologically relative scene primitives that describe the structural, constancy, and functional aspects of scene surfaces without representing objects and parts” (Greene & Oliva, 2009a, p.163). It is believed GSPs are not only encoded during early scene processing, but they are also categorized rapidly supporting the idea of rapid scene recognition (Greene & Oliva, 2009a, 2009b; Ross & Oliva, 2010). Examples of GSPs include the naturalness, openness, expansion, and navigability of a scene (Greene & Oliva, 2009a, 2009b).

Further proving the existence of GSPs it has been shown particular GSPs correspond to specific high-level, category-selective cortical regions in the brain (for a review, see Malcolm et al., 2016). One of such regions is the previously mentioned Parahippocampal Place Area (PPA) a region within the posterior medial temporal lobe (MTL: Köhler et al., 2002) located within the region of the Anterior collateral sulcus

(aCos). The PPA has been shown to harbor information regarding the spatial expanse of the scene-pertaining to the openness and closeness of a scene- when participants are presented with naturalistic scene images (Park et al., 2011; Kravitz et al., 2011) and computer-generated artificial images (Harel et al., 2013; Lowe et al., 2016). The PPA has also been shown to harbor information relating to the spatial boundary of the scene (Ferrera and Park, 2016). Information about another important property of a scene, naturalness, has been shown to be held in the Lateral Occipital Cortex (LOC) where different response patterns have been observed for manmade and natural scenes (Park et al., 2011).

While fMRI studies have uncovered some of the neural correlates of scene processing, electrophysiological studies have been able to uncover the time course of scene processing by measuring the brain's electrical activity. This is measured by using event-related potentials (ERPs) which are stereotyped brain responses due to the presentation of a stimulus. They are measured using electroencephalography (EEG) and magnetoencephalography (MEG). Consisting of a series of positive and negative voltage deflections corresponding to different cognitive processes taking place during a particular moment in time.



**Figure 1:** Time (ms) on the X-axis and amplitude ( $\mu\text{V}$ ) on the Y-axis. As is shown, ERP's are present corresponding to particular cognitive processes taking place at a particular moment in time (S.J. Luck, 2015; *The ERP Boot Camp: What are ERPs and What Are They Good For?*).

These electrophysiological studies have been crucial in our understanding of the time course of scene processing. For example, a study conducted, using magnetoencephalography (MEG) found a scene-specific component by contrasting responses to faces and scenes, reported a stronger response to scenes than faces between 200-300 ms. after stimulus onset (Sato et al., 1999). Another MEG study reported an earlier scene response peaking between 100-130 ms. after stimulus onset (Rivolta et al., 2012). An electroencephalographic (EEG) study conducted by Harel et al. (2016) using scenes, faces and common objects confirmed the results of the former MEG study, showing a higher P2 response (peaking 220 ms. after stimulus onset) to scenes compared with faces and objects. In a second experiment, Harel et al. (2016) used naturalistic scene images complete with color and semantic category information spanning the global scene



properties of naturalness (manmade/natural), spatial boundary (open/closed), and relative distance (near/far), demonstrating that the amplitudes of the scene-selective P2 and earlier N1 (peaking 170 ms after stimulus onset) components changed as a function of the naturalness and spatial boundary of the scene.

The results of the second experiment done in the Harel et al. (2016) study- henceforth referred to as the 96 scenes study- showed that the P2 and N1 ERP components are sensitive to the GSP dimensions of naturalness and spatial boundary. However, these results were observed using naturalistic scene stimuli rich in low (color) and high-level (semantic category) which could provide an alternative explanation to the observed results. It may stand to reason that the observed results do not reflect the processing of mid-level GSPs, but rather the processing of either low- or high-level information.

The current study was conducted in order to investigate this possibility. Over the course of two experiments, we sought to explore the two following questions.(1) Will GSP sensitivity to mid-level global information be observed using artificially-generated stimuli (eliminating color information and minimizing semantic category information from scene stimuli)? (2) What is the extent to which scene processing is dependent on local or global elements (spatial layout) contained within the scene? This was done by inverting the same scene stimuli as in Experiment 1, eliminating global information while preserving local information.

In both experiments, we presented participants with artificially-manipulated scenes which did not contain color information (they were grayscale images) and had

very little semantic detail (object information informing on the scene category). At the same time, the scenes varied on naturalness and spatial boundary, allowing us to ask whether we will still see the effect of these two GSPs, despite the removal of low- and high-level sources of information. The main difference between the two experiments was that as noted before, during the follow-up experiment, half the scene stimuli were inverted. This allowed us to examine the extent to which the observed results in Experiment 1 are dependent on local/global information.

In both experiments, the P2 component was the key component of interest in furthering the understanding the role GSPs of naturalness and spatial boundary play when color and semantic category information are eliminated. Additionally, the earlier N1 component, well as the even earlier P1 component (peaking 120 ms post stimulus onset) were examined to determine their sensitivities to spatial boundary and naturalness. Research has shown the importance of global information (Oliva & Torralba, 2001, 2006) particularly of naturalness (Loschky et al., 2010; Harel et al., 2016) and spatial boundary (Harel et al., 2016) during early scene processing. However, this effect has yet to be seen using computer-generated stimuli stripped of much semantic information informing about the identity of a scene as well as the absence of color information.

Stimuli used in the present study differed from stimuli in the 96 scenes study. In the 96 scenes study participants viewed highly detailed diverse real-world images complete with color information of differing scene categories (churches, hallways, living rooms, deserts, mountains, etc.) spanning three global scene properties: spatial expanse, relative distance, and naturalness. While participants in the present studies viewed

“impoverished” artificially-generated scene stimuli deprived of low- and high-level scene information with only global mid-level scene information available for processing.

If mid-level naturalness and spatial boundary effects were seen despite the removal of low and high-level information, it could be concluded that GSPs are processed by mid-level global information. Particularly with naturalness and spatial boundary acting as a bridge between low and high-level scene information. Taken together the present study could add to the 96 scenes study and show the effect of GSPs on the P2 component is evident across different types of scene stimuli during early scene processing.

Also, the N1 component was analyzed since its amplitude has previously been shown to be modulated by the naturalness of the scene (Harel et al. 2016). The 96 scenes study showed the N1 to have a sensitivity to the naturalness of the scene, with manmade scenes eliciting a stronger response than natural scenes. While the N1 component has been shown to be sensitive to scene-properties, face perception studies have shown the N1 to be a face-specific component (Bentin et al., 1996; Bötzel et al., 1995; Blau et al., 2007). Studies of the N1 have shown a larger response is elicited when presented with face stimuli as opposed to non-face stimuli (Rossion & Jacques, 2008). Results from face perception studies show the N1 plays a role in global processing (Caharel et al., 2013), and also show a face inversion effect (Rossion et al., 1999). With the notion that global processing forms the basis of scene perception, we thought it important to include the N1 as a component of interest.

In addition, the P1 component was analyzed. Past research has shown the P1 component is sensitive to spatial attention (Hillyard, Vogel, & Luck, 1998). In the 96 scenes study, scene-specific effects on the P1 were not shown but given the P1's role during attention it was important to include the component in the analysis for the present study. Analyzing this component in the current study could also reveal whether global scene processing begins earlier than was thought previously.

### **Experiment 1**

At the level of the P2 component, we formulated the null hypothesis ( $H_0$ ) to be if P2 GSP effects are driven by low-level such as color information and high-level semantic information then stripping scene stimuli of this information would result in the diminishment of those P2 effects. Contrary to the null hypothesis we formulated our actual hypothesis also known as the alternative hypothesis ( $H_a$ ) to be if P2 GSP effects are driven by GSPs and not by color or semantic information then the removal of such information from scene stimuli would result in no changes to the P2 effects.

Similar predictions were made about the N1 component as well. We predicted the null hypothesis ( $H_0$ ) to be that if the N1 GSP effects are driven primarily by color and semantic information, then stripping the scene stimuli of this information will also result in the diminishment of those N1 effects. While our alternative hypothesis ( $H_a$ ) states that if N1 GSP effects are driven by GSPs and not by color or semantic information, then removal of information from the stimuli will not result in any changes to the N1 effects.

We did not hypothesize there would be any effects on the P1 component. This is because it has been shown that at the level of the P1 component, low-level features such

as contrast, spatial frequency, and color are dominant (Fu et al., 2005; Roission & Caharel, 2011). However, given the differences in stimuli between the experiments a scene-selective effect might still be observed at the P1 level. To summarize the above hypotheses, we present them in a schematic format in Table 1, below.

**Table 1: Experiment 1 Hypotheses:**

If null hypothesis ( $H_0$ ) true	Color and Semantic information removal: ↓P2, ↓N1, -P1
If alternate hypothesis ( $H_A$ ) is true	Color and Semantic information removal: ↑P2, ↑N1, -P1

## **Method & Results**

### **Participants**

Seventeen undergraduate and graduate students from Wright State University (14 males,  $M = 22.5$  years) participated in this experiment. It was not intentional to have a study that was almost exclusively male, during the recruitment process, the majority of people who responded and participated in the experiment were males. Three participants were excluded from the final analyses due to their EEG artifacts (eye blinks, eye movements, muscle activity, and skin potentials) being more than 20% of the trials, resulting in us not knowing if the ERP signals were due to the experimental manipulation or the abundance of artifacts in the participants data. The final sample size is in line with current ERP literature and follows common practices in the field of ERP (for current recommendations and guidelines, see Luck 2014; Picton et al., 2000). All participants were compensated monetarily using checks in the amount of \$20.

### **Stimuli and Apparatus**

The stimulus set consisted of 576 individual artificially generated grayscale scene images (Originally used in Lowe et al., 2016) spanning two global scene properties (GSPs): Naturalness (Natural/Manmade) and Spatial Boundary (Open/Closed) creating four total conditions: Manmade Closed, Manmade Open, Natural Closed, and Natural Open with 144 variable images per condition and twelve unique layouts and textures per image. While Spatial Boundary and Naturalness can be considered as continuous

variables (Zhang, Houpt, & Harel, 2018), here, for sake of simplicity, we opted to use only stimuli that represent the distinct qualities of these dimension. For example, natural scenes were unambiguously natural (open desert and closed caves) and manmade scenes did not contain any element that would signify them as natural. We simply used open rooftops and indoor spaces (rooms). This allowed us (unlike studies using naturalistic images) to actively control and manipulate the factors of spatial boundary and naturalness. Moreover, the unique 12 layout and 12 textures within each condition allowed for substantial variation in stimulus appearance, culminating in 144 unique exemplars per condition, which by far exceeds the recommended number of trials per experimental condition (see below). Images sized 500 x 500 pixels were presented on a Dell LCD monitor on 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](http://www.neurobs.com)).

### **EEG Recording**

Two electrodes were placed on the right and left mastoids as well as an electrode to the tip of the nose. To control for eye movements and blinks, two pairs of electrooculography (EOG) electrodes were used: one pair attached to the external canthi and one pair attached to the superior and inferior regions of the right eye. The nasion-inion and head circumference anatomical measurements were taken to locate the vertex site and fit them for an electro-cap (ECI), respectively. The electrode holes were soaked

in electrolyte ([www.parklabs.com](http://www.parklabs.com)) followed by the application of 64 Ag-AgAl pin-type active electrodes (Biosemi ActiveTwo) according to the 10-20 system.

All electrodes were grounded to the Common Mode Sense (CMS) active electrode and the Driven Right Leg (DRL) passive electrode placed between the PO3 and PO4 electrodes. The signal was amplified by an ActiveTwo AD-Box digitized at 100 microvolts ( $\mu\text{V}$ ) and recorded using Biosemi software running on an HP desktop computer. Both EEG and EOG electrodes were sampled at 512 Hz with a 24-bit resolution with an active input range of -262 mV to +262 mV per bit. The application of a 51 Hz low-pass filter was applied to prevent aliasing effects. The digitized EEG was saved and processed off-line.

## **Procedures**

Participants were escorted to the preparation room where they were told the specifics of the study and signed an informed consent form. Once prepped for the study, subjects were taken to the electromagnetically-shielded experimental room. The study spanned twelve blocks for a total of 65 minutes. An individual block lasted 5.5 minutes and during that time scene stimuli were pseudo randomly presented for 500 ms with a jittered interstimulus interval (ISI) between them ranging from 750 ms to 1250 ms. Stimuli were presented twice throughout the study, creating a total of 1152 trials, 288 trials per experimental condition. The reasoning for such a large amount of trials per participant is that since ERPs are so small a large amount of trials is needed in order to get an accurate ERP measurement (Luck, 2014). Notably, whereas the standard practice is a minimum of 50 trials per condition, here we use five times as much, with more 250



trials per condition. Participants performed an orthogonal fixation cross task (Harel et al., 2016) in which a fixation cross would appear simultaneous to the presentation of images and subjects were instructed to fixate on the cross and report whether the fixation cross arm lengthened horizontally or vertically with each trial by button press using a mouse attached to the computer monitor displaying the stimuli.

This task was used to measure scene representations while trying to minimize any task effects (Kravitz et al., 2008). Participants were instructed to keep movements- eye blinks, crossing of legs and feet, arm movements, etc.- to a minimum due to these movements creating electrical activity masking the minute signals of interest. Once done with a block, participants were able to move freely in order to ready themselves for the next block. Participants were then able to initiate the next block by using button press. Upon completion, participants were taken to a washroom where they could wash the electrolyte gel from their hair and were given monetary compensation in the form of a check in the amount of \$20.

### **Data processing and analysis**

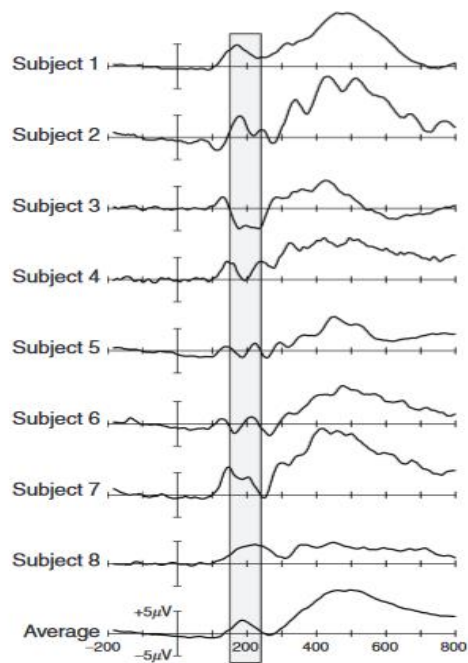
The raw data was imported into the computer software program Brain Vision Analyzer 2 (Brain Products GmbH, Munich, Germany). The data was processed first by using a 70 Hz low-pass filter (24 dB/oct), a 1 Hz high-pass filter (24dB/oct), and a 60 Hz Notch filter and referenced to the tip of the nose. Ocular artifacts were removed using an ocular correction Independent Component Analysis (ICA) procedure (Jung et al., 1998). Any remaining artifacts exceeding  $\pm 100$  mV in amplitude or a change of 100 mV over a 50 ms time span were rejected.

The processed data of each participant was then segmented into epochs ranging from -200 ms before stimulus onset to 800 ms after stimulus onset representing each of the four conditions (Natural Open, Natural Closed, Manmade Open, Manmade Closed) and their main effects (Natural, Manmade, Open, Closed) followed by a 17 Hz low-pass filter for all the conditions. Within the segments of each participant, the peaks of the P1, N1, and P2 components were detected for each of the posterior lateral electrode nodes.

The peaks for each component of interest: P1, N1, and P2 were defined for each subject in each experimental condition: The first most positive peak (P1 component) lasting from 80 ms -130 ms, the most negative peak (N1 component) lasting from 140 ms -190 ms, and the second most positive peak (P2 component) lasting from 200 ms - 270 ms. Analysis was restricted to the posterior lateral electrodes on the right (P6, P9, PO7, P10) and left (P5, P7, P9, PO7) hemispheres where scene effects have been shown to be most pronounced (Harel et al., 2016). The detected peaks representing each condition were exported into Microsoft Excel to be properly formatted and then imported into IBM SPSS Statistics ([www.ibm.com](http://www.ibm.com)) where data analysis took place.

## Results

The grand-average waveforms were taken by averaging all the averaged waveforms of individual subjects for the main effects of spatial boundary (open/closed) and naturalness (natural/manmade) and simple effects of manmade and natural scenes. Peak amplitudes of the P1, N1, and P2 components were analyzed using a repeated measures Analysis of Variance (ANOVA) with four independent variables: Hemisphere (left/right), Laterality (lateral/medial), Naturalness (natural/manmade), and spatial boundary (open/closed).



**Figure 2:** The waveforms reflect 8 randomly selected subjects from the 20 individuals who contributed to the grand average. The bottom waveform is the grand average across subjects. The high-lighted portion of the figure corresponds to the time period the P2 wave is measured. The activity within this time window varies considerably across the individual subject waveforms. For some of the subjects, the first positive wave peaks prior to the beginning of the window (e.g., subjects 3, 4, and 7), and one subject's waveform is entirely negative during this window (subject 3). (Kappenman & Luck, 2012; *ERP Components: The Ups and Downs of Brainwave Recordings*, p.13)

## P2 Component

Grand-average waveforms are depicted in Figure 1. The significant results are reported in Figure 2.

There was a significant effect of Spatial Boundary [ $F(1,13) = 8.38$ ,  $MSE = 1.45$ ,  $p < 0.01$ ] with closed scenes ( $M_{\text{closed}} = 2.00$  mV,  $SE = 0.51$ ) evoking a higher response than open scenes ( $M_{\text{open}} = 1.31$ ,  $SE = 0.55$ ). Also in line with our predictions, there was a significant effect of Naturalness [ $F(1,13) = 5.57$ ,  $MSE = 2.67$ ,  $p < 0.04$ ], with manmade scenes ( $M_{\text{manmade}} = 2.04$  mV,  $SE = 0.60$ ) eliciting a higher response than natural scenes ( $M_{\text{natural}} = 1.31$  mV,  $SE = 0.47$ ).

There was a significant interaction of Hemisphere and Spatial Boundary [ $F(1,13) = 6.28$ ,  $MSE = 0.20$ ,  $p < 0.03$ ] with closed scenes ( $M_{\text{closed}} = 2.57$  mV,  $SE = 0.57$ ) evoking a greater response than open scenes ( $M_{\text{open}} = 1.70$  mV,  $SE = 0.60$ ) in the right hemisphere ( $t(13) = 3.26$ ,  $p < 0.003$ ). While in the left hemisphere ( $t(13) = 2.06$ ,  $p < 0.03$ ) closed scenes ( $M_{\text{closed}} = 1.42$  mV,  $SE = 0.52$ ) eliciting a higher response than open scenes ( $M_{\text{open}} = 0.98$  mV,  $SE = 0.56$ ). This shows the effects of Spatial Boundary are more pronounced in the right hemisphere relative to the left hemisphere which is in line with other studies (Harel et al., 2016 & Hansen et al., 2018) showing hemispheric asymmetries with the right hemisphere eliciting a higher response than the left hemisphere.

Lastly, analysis of the P2 component shows there was a significant effect of Hemisphere [ $F(1,13) = 6.86$ ,  $MSE = 3.60$ ,  $p < 0.02$ ] with the right hemisphere ( $M_{\text{right}} = 2.14$  mV,  $SE = 0.57$ ) evoking a higher response than the left hemisphere ( $M_{\text{left}} = 1.0$  mV,  $SE = 0.53$ ).

**Figure 3a: Experiment 1 Results**

<b>P2</b>	Effect	df	MSE	F	Significance
Spatial Expanse	Closed > Open	1,13	1.45	8.38	0.01
Naturalness	Manmade > Natural	1,13	2.67	5.57	0.04
Hemisphere x Spatial Boundary	RH: Closed > Open LH: Closed > Open	1,13	0.20	6.28	0.03
Hemisphere	RH > LH	1,13	3.60	6.86	0.02

### **N1 and P1 Components**

Significant effects were also found in the other two components- N1 and P1 components. For the N1 component there was a significant effect of Spatial Boundary [ $F(1,13) = 5.79$ ,  $MSE = 0.90$ ,  $p < 0.04$ ] with closed scenes ( $M_{\text{closed}} = -3.46$  mV,  $SE = 0.67$ ) evoking a more negative response than open scenes ( $M_{\text{open}} = -3.03$ ,  $SE = 0.60$ ).

Approaching significance is an interaction of Hemisphere and Naturalness [ $F(1,13) = 4.26$ ,  $MSE = 0.42$ ,  $p < 0.06$ ] with manmade scenes ( $M_{\text{manmade}} = -3.76$  mV,  $SE = 0.63$ ) evoking a greater negative amplitude than natural scenes ( $M_{\text{natural}} = -3.26$  mV,  $SE = 0.75$ ) but only in the right hemisphere ( $t(13) = 1.55$ ,  $p < 0.08$ ) with no effect in the left hemisphere ( $t(13) = 0.006$ ,  $p < 1.00$ ).

At the level of the P1 component there was a significant effect of Naturalness [ $F(1,13) = 7.45$ ,  $MSE = 1.51$ ,  $p < 0.02$ ] with natural scenes ( $M_{\text{natural}} = 3.10$  mV,  $SE = 0.44$ ) evoking a higher response than manmade scenes ( $M_{\text{manmade}} = 2.47$  mV,  $SE = 0.53$ ). A significant effect of Spatial Boundary [ $F(1,13) = 15.31$ ,  $MSE = 1.12$ ,  $p < 0.002$ ] was

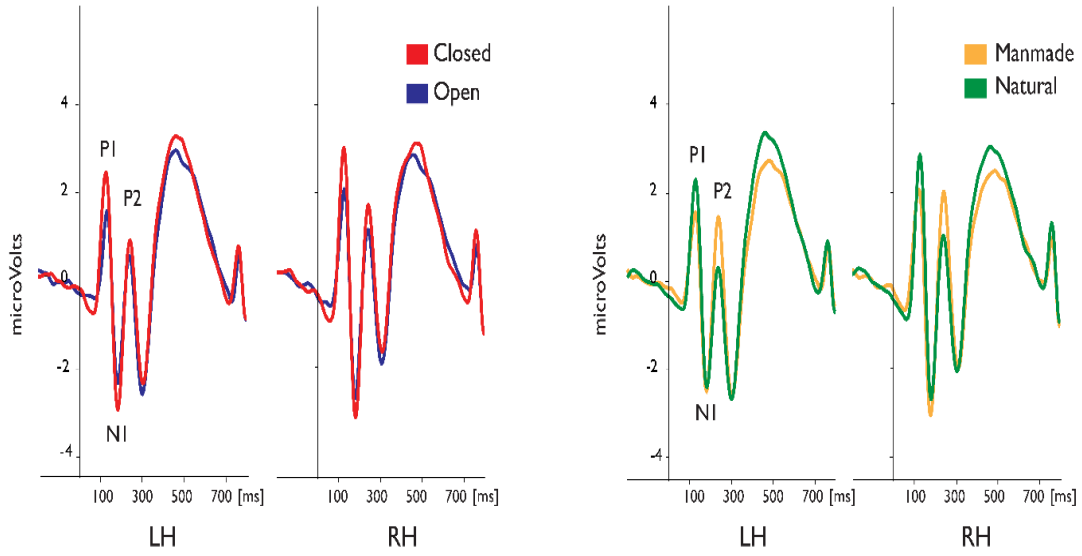
observed with closed scenes ( $M_{\text{closed}} = 3.18$  mV,  $SE = 0.50$ ) evoking a higher response than open scenes ( $M_{\text{open}} = 2.39$ ,  $SE = 0.48$ ).

**Figure 3b: Experiment 1 Results:**

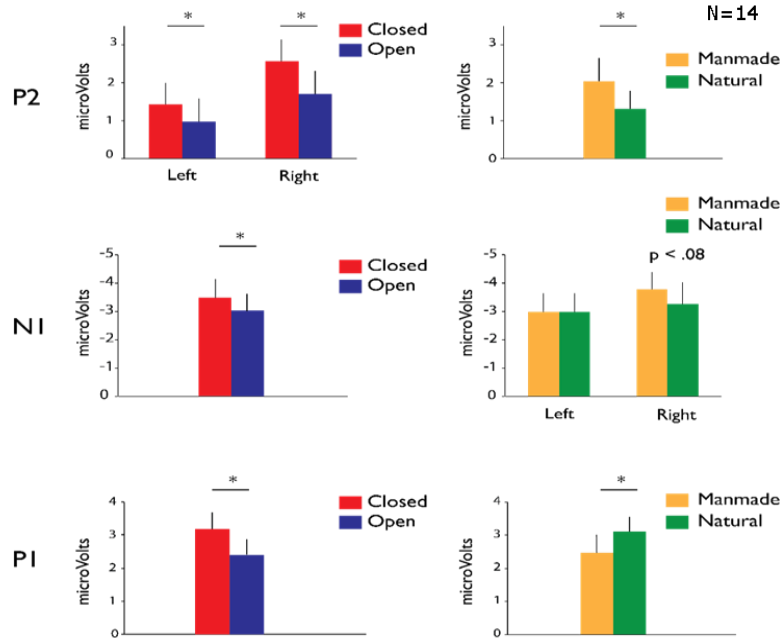
<b>N1</b>	Effect	df	MSE	F	Significance
Spatial Expanse	Closed > Open	1,13	0.90	5.79	0.04
Hemisphere x Naturalness	RH: Manmade > Natural LH: n.s.	1,13	0.42	4.26	0.06 , $\alpha=0.05$

**Figure 3c: Experiment 1 Results:**

<b>P1</b>	Effect	df	MSE	F	Significance
Spatial Boundary	Closed > Open	1,13	1.12	15.31	0.002
Naturalness	Natural > Manmade	1,13	1.51	7.45	0.02



**Figure 4:** 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 extent (closed vs. open; red and blue, respectively). Right: Naturalness (man-made vs. natural; yellow and green, respectively).



**Figure 5:** Grand average ERP analysis results. Top row: Mean P2 peak amplitudes in response to closed and open scenes (red and blue, respectively) presented separately for the left and right hemispheres (LH, RH, respectively) and in response to manmade and natural scenes (yellow, green, respectively). Middle row: 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 (LH, RH, respectively). Bottom row: 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. Reported here are the significant effects only. 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.



## Discussion

Results from the experiment show the effects of GSPs being concentrated during the P2 and N1 ERP components with some effects occurring during the earlier P1 ERP component. During the P2 component, the amplitude was modulated by the naturalness and spatial boundary of the scene with manmade scenes eliciting a more positive amplitude than natural scenes and closed scenes eliciting a more positive amplitude than open scenes respectively. Also, there was a hemispheric effect with the right hemisphere of the brain eliciting a higher amplitude than the left hemisphere.

At the N1 time window, a spatial boundary effect and an interaction of hemisphere and naturalness was observed with closed scenes eliciting a more negative amplitude than open scenes while manmade scenes elicited a more negative amplitude than natural scenes, but this effect was restricted to the right hemisphere. The P1 component was shown to harbor naturalness and spatial boundary effects with manmade scenes eliciting a more positive amplitude than natural scenes and closed scenes eliciting a more positive amplitude than open scenes respectively.

The above results allow us to reject the null hypothesis ( $H_0$ ) and accept the alternative hypothesis ( $H_A$ ), namely, that stripping off low-level information (color) and high-level semantic object information from the scenes while preserving mid-level global information such as naturalness and spatial boundary would have no impact on the amplitudes of the components of interest (P2, N1, and P1).

However, the scenes used in Experiment 1 also contain local (low-level) information which might explain the observed results. For example, manmade scenes also differ from natural scenes based on their local textural information in addition to the changes in their global layout. Thus, there is still a possibility that the results we observed are due to local processing that is still occurring with these scenes. This led us to conduct Experiment 2, which investigated the role that local and global sources of information play in driving the early visually-evoked responses to scenes.

## **Experiment 2**

While Experiment 1 allows us to determine if scene perception is driven by mid-level global information (serving as a bridge between low and high-level information), the extent to which scene perception is driven by the global or local distribution of scene information is unknown. As noted above, previous studies have suggested the global representation is the primary source of information for scene perception. Asking the question of whether scene perception is driven by global or local factors we conducted another experiment. This new study-termed Experiment 2- was near identical to Experiment 1, with the same orthogonal fixation task in which participants were instructed with each trial to report whether the fixation cross situated in the center of the computer screen lengthened horizontally or vertically by button press.

However, unlike Experiment 1, for half the trials (576 trials) orientation of the images were inverted 180°, the rationale for this being that scene inversion should result in the degradation of global features (Spatial Boundary in particular) and instead rely more on local factors (orientation, spatial frequency, color, etc.) begging the question of whether scene perception is globally or locally driven. Participants' brain responses were measured using EEG studying the time course of scene processing, the stimulus set used in Experiment 2 was identical to what was used in Experiment 1. Consisting of 576 artificially generated images (originally used by Lowe et al., 2016) spanning the GSPs of Naturalness (Manmade/Natural) and Spatial Boundary (Open/Closed) with color and

semantic category information stripped away from the images (discussed in introduction for Experiment 1).

The present study parallels work furthering the understanding of facial processing. It has been shown using electrophysiological measures that facial processing is primarily concentrated around 170 ms post-stimulus onset pertaining to the level of the N1 component (Roisson and Jacques, 2011). Examining the differences in the neural processes involved in facial recognition as compared to object recognition has led to studies being conducted in which participants are presented inverted human faces as compared to upright faces and upright and inverted objects (Rossion & Delvenne et al., 1999; Roisson & Gauthier et al., 1999). These studies showed the N1/N170 face-selective component is higher in amplitude and has a more delayed latency when participants were presented with inverted human faces as compared to upright human faces (Rossion & Delvenne et al., 1999).

The explanation for this face inversion effect is that the processing of holistic information normally seen during the presentation of upright facial stimuli (Tanaka & Farrah, 1993; Rhodes, 1993; Moscovitch et al., 1997) is disrupted when inverted facial stimuli are presented. Thus, inversion causes the degradation of holistic processing mechanisms (Rhodes, 1993; Rhodes et al., 1993; Young et al., 1989) while preserving low-level features such as individualized facial characteristics (Moscovitch et al., 1997).

We predicted the null hypothesis ( $H_0$ ) would be that inverting the normally upright stimuli 180° for half the trials would not result in the degradation of the global distribution of the scene. Thus, preserving the naturalness and spatial boundary effects

seen during Experiment 1, with no orientation effects corresponding to the ERP components of interest (P1, N1, P2).

Alternatively, we predicted our actual hypothesis ( $H_1$ ) to be inversion of the stimuli for half the trials would result in the degradation. of the global scene properties of naturalness and Spatial Boundary. This would be concentrated around the P2 and N1 time points. At the level of the P1 component, we did not predict any diminishment of P1 naturalness and spatial boundary effects this is due to the P1 primarily processing low-level features which are not impacted by scene inversion. `

**Table 2: Experiment 2 Hypotheses**

Null hypothesis ( $H_0$ )	<ul style="list-style-type: none"> <li>• No effect of inversion on P2, N1, and P1 Components.</li> <li>• <math>\uparrow</math>P2, <math>\uparrow</math>N1, <math>\uparrow</math>P1 Naturalness and Spatial Boundary effects.</li> </ul>
Alternate hypothesis ( $H_A$ )	<ul style="list-style-type: none"> <li>• <math>\uparrow</math>P2 and <math>\uparrow</math>N1 Scene inversion effects.</li> <li>• <math>\downarrow</math>P2 and <math>\downarrow</math>N1 Naturalness and Spatial Boundary effects.</li> <li>• No diminishment of P1 Naturalness and Spatial Boundary effects.</li> </ul>

## **Method & Results**

### **Participants**

Twenty-one undergraduate and graduate students from Wright State University (12 females, M=19.8 years) participated in this experiment (see comment above regarding recommended sample size). All participants were recruited from the Wright State, but none participated in Experiment 1 to avoid unequal exposure to the experimental stimuli. Four participants were excluded from the final analyses either due to EEG artifact rejection being more than 20% of the trials or problems arising during the study resulting in the participant exiting early. All participants signed an informed written consent form according to the guidelines of the Institutional Review Board (IRB) of Wright State University. All were right-handed had normal or corrected-to-normal acuity with no history of neurological diseases. All were compensated monetarily using checks in the amount of \$30 or were given course credit in order to fulfill a requirement for an introductory psychology course in which they were enrolled.

### **Stimuli, Apparatus, and Procedures**

The stimuli, materials, and tasks were the same as in Experiment 1. However, for half the trials- 576 trials- the stimuli normally in an upright orientation were inverted 180° to test whether the global or local distribution of the scene is primarily responsible for scene recognition. The experimental session consisted of twelve blocks, in which scene stimuli were presented pseudo randomly for 500 ms., with a jittered interstimulus

interval (ISI) between them lasting 750-1250 ms. Stimuli were presented twice for a total of 1152 trials.

### **Data processing and analysis**

The raw data was processed identical to the way it was in Experiment 1. Once processed (as discussed in methods for Experiment 1) the data for each participant was segmented into epochs representing eight conditions (Natural Open, Natural Closed, Manmade Open, Manmade Closed, Natural Open Inverted, Natural Closed Inverted, Manmade Open Inverted, Manmade Closed Inverted) ranging from -200 ms. before stimulus onset to 800 ms. after stimulus onset. Peak detection was also identical to what was done in Experiment 1.

### **Results**

Four participants were eliminated from the data analysis due to artifacts consisting over 20% of their raw data. Analysis was restricted to the posterior lateral electrodes on the right (P6, P9, PO7, P10) and left (P5, P7, P9, PO7) hemispheres where scene effects are most pronounced (Harel et al., 2016).

Peak amplitudes of the P1, N1, and P2 components were analyzed using a four-way repeated measures ANOVA with four independent variables: Hemisphere (left/right), Spatial Boundary (open/closed), Naturalness (natural/manmade), and Orientation (upright/inverted).

## P2 Component

Similar to the first experiment, during the P2 timepoint there was a significant effect of there was a significant effect of Spatial Boundary [ $F(1,16) = 6.41$ ,  $MSE = 2.09$ ,  $p < .022$ ] with closed scenes ( $M_{\text{closed}} = 3.73$  mV,  $SE = 0.77$ ) eliciting a higher response than open scenes ( $M_{\text{open}} = 3.29$  mV,  $SE = 0.72$ ). A significant effect of Naturalness [ $F(1,16) = 10.57$ ,  $MSE = 3.49$ ,  $p < .005$ ] was observed with manmade scenes (3.88 mV,  $SE = 0.75$ ) eliciting a higher response than natural scenes (3.14 mV,  $SE = 0.76$ ). Also, there was a significant effect of hemisphere [ $F(1,16) = 19.33$ ,  $MSE = 17.22$ ,  $p < .000$ ] with the right hemisphere ( $M_{\text{right}} = 4.61$  mV,  $SE = 0.86$ ) eliciting a higher response than the left hemisphere ( $M_{\text{left}} = 2.40$  mV,  $SE = 0.71$ ). However, there was no observed significant effect of orientation or any observed interaction with orientation (all  $p > 0.05$ ).

**Figure 6a: Experiment 2 Results:**

<b>P2</b>	Effect	df	MSE	F	Significance
Spatial Expanse	Closed > Open	1,16	2.09	6.41	0.022
Naturalness	Manmade > Natural	1,16	3.49	10.56	0.005
Hemisphere	RH > LH	1,16	17.22	19.33	0.000

## N1 and P1 Components

In contrast to what was observed in Experiment 1, during the N1 timepoint no significant effect of any of the main effects were shown (all  $p > 0.05$ ) not even a significant effect of Spatial Boundary as was seen in Experiment 1 and neither were there any significant interactions of hemisphere and naturalness (all  $p > 0.05$ ).



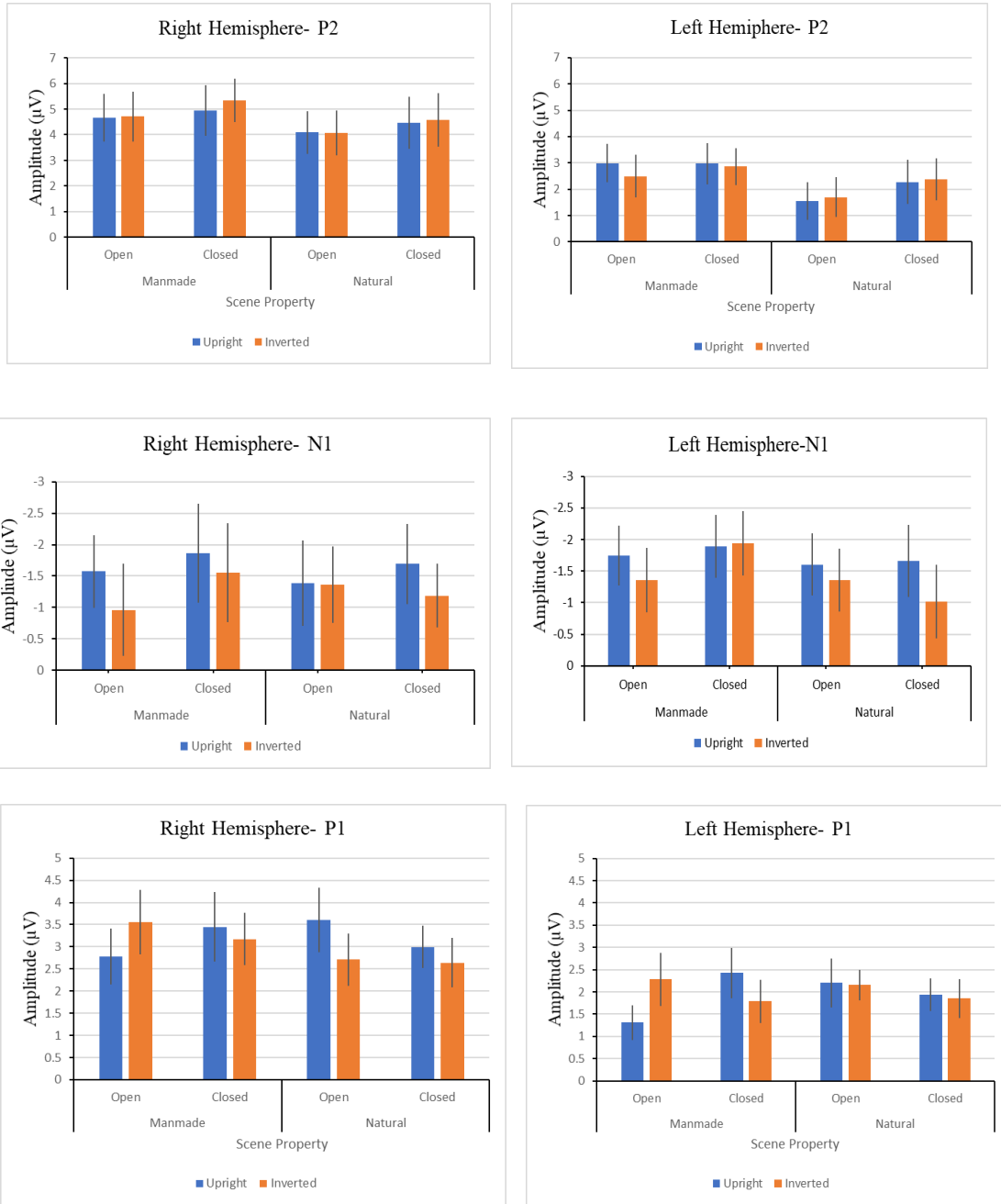
During the P1 time window there was a significant effect of Hemisphere [ $F(1,16) = 7.53$ ,  $MSE = 11.33$ ,  $p < 0.01$ ] with the right hemisphere eliciting a higher response than the left hemisphere. Also, there appeared to be significant interactions of (1) Naturalness and Orientation [ $F(1,16) = 7.93$ ,  $MSE = 0.66$ ,  $p < 0.01$ ], (2) Hemisphere and Spatial Boundary and Orientation [ $F(1,16) = 5.79$ ,  $MSE = 1.99$ ,  $p < 0.03$ ], and (3) Spatial Boundary and Naturalness and Orientation [ $F(1,16) = 4.3$ ,  $MSE = 0.20$ ,  $p < .055$ ]. However, follow-up analyses using a three-way repeated measures ANOVA with Spatial Expanse, Naturalness, and Orientation as independent variables revealed none of the above effects to be statistically significant (all  $p > 0.05$ ). However, there was a Naturalness by Orientation interaction that was approaching significance [ $F(1,16) = 4.21$ ,  $MSE = 0.87$ ,  $p < 0.055$ ,  $\alpha = 0.050$ ]. Manmade inverted scenes ( $M_{\text{inverted}} = 2.49$  mV,  $SE = 0.46$ ) evoked a greater response than manmade upright scenes ( $M_{\text{upright}} = 2.35$  mV,  $SE = 0.46$ ), while in contrast natural upright scenes ( $M_{\text{upright}} = 2.47$  mV,  $SE = 0.40$ ) evoked a greater response than natural inverted scenes ( $M_{\text{inverted}} = 2.17$  mV,  $SE = 0.37$ ).

**Figure 6b: Experiment 2 Results:**

<b>N1</b>	<b>Effect</b>	<b>df</b>	<b>MSE</b>	<b>F</b>	<b>Significance</b>
Spatial Expanse	–	–	–	–	n.s.
Naturalness	–	–	–	–	n.s.
Hemisphere	–	–	–	–	n.s.
Orientation	–	–	–	–	n.s.

Figure 6c: Experiment 2 Results:

<b>P1</b>	Effect	df	MSE	F	Significance
Hemisphere	RH > LH	1,16	11.33	7.53	0.01
Naturalness x Orientation	Manmade Inverted > Manmade Upright Natural Upright > Natural Inverted	1,16	0.87	4.21	0.055 $\alpha=0.05$



**Figure 7:** Grand average ERP analysis results. Top row: Mean P2 peak amplitudes in response to upright and inverted scenes (blue and orange, respectively) for manmade open/manmade closed and natural open/natural closed images (presented respectively) across the right and left hemispheres (RH, LH, respectively). Middle row: Mean N1 peak amplitudes in response to upright and inverted scenes (blue and orange, respectively) for manmade open/manmade closed and natural open/natural closed images (presented respectively) across the right and left hemispheres (RH, LH, respectively). Bottom row: Mean P1 peak amplitudes in response to upright and inverted scenes (blue and orange, respectively) for manmade open/manmade closed and natural open/natural closed images (presented respectively) across the right and left hemispheres (RH, LH, respectively). All data are plotted for the posterior lateral sites.

## Summary and Discussion

During the P2 time point significant effects of naturalness and spatial boundary were observed with manmade scenes eliciting a higher amplitude than natural scenes and closed scenes eliciting a higher amplitude than open scenes. However, no effects of orientation were seen. At the level of the N1 component no significant effects of naturalness and spatial boundary were found as well as a lack of any orientation effects. While at the level of the P1 component there appeared to be multiple significant interactions, follow-up analyses of these interactions revealed no significant effects except for a naturalness by orientation interaction approaching significance.

The results of this study allow us to accept the null hypothesis ( $H_0$ ) stating that scene inversion would not degrade global information contained within the scene. Thus preserving the naturalness and spatial boundary effects observed in the components of interest. We are then allowed to reject the actual hypothesis ( $H_A$ ) stating scene inversion would degrade global scene information causing a diminishment of spatial boundary and naturalness effects concentrated around the P2 and N1 time points. However, most interesting is the complete lack of naturalness and spatial boundary effects at the level of the N1 component. Additionally, the orientation interaction approaching significance at the level of the P1 component was another unexpected finding.

## General Discussion

The current work investigated the extent to which mid-level scene properties - particularly of naturalness and spatial boundary - drive scene perception even when low-level color information and high-level semantic category information are absent or minimized. We then asked whether the observed results were due to the processing of global or local information within the scene. To address these questions, we conducted two experiments. In Experiment 1, participants observed gray-scale scene stimuli varying on the GSPs of naturalness and spatial boundary, with much detail informing about scene category removed. In Experiment 2, participants observed the same stimuli, but now half of the scenes were in an inverted orientation, the rationale being that global processing would be disrupted upon inversion<sup>1</sup>, while local information would be preserved.

Operationally, Experiment 1 investigated whether the ERP signature responses to the GSPs of spatial boundary and naturalness would be seen in images that were grayscale and lacking in semantic information. The P2 and N1 components were analyzed following previous literature showing the GSP effects are concentrated during those time windows (Harel et al., 2016). The P1 component was also investigated to

---

<sup>1</sup>In a study done by Loschky et al. (2015) able to show that for terrestrial scenes there was an orientation effect and it was most maximal when the scenes were rotated at 135°.

observe possible earlier effects. In Experiment 1, across all three components, a spatial boundary effect was observed, with closed scenes eliciting a higher response than open scenes. In addition, a naturalness effect was also observed across the three components with manmade scenes eliciting a higher response than natural scenes during the P2 time window, and natural scenes eliciting a higher response than manmade scenes during the P1 time window. During the N1 time window, manmade scenes elicited a more negative response than natural scenes, but this was restricted to the right hemisphere.

While previous studies have shown GSP effects are seen during early scene perception, these effects were seen using naturalistic stimuli complete with color and semantic category information. What Experiment 1 was able to show was that despite the scenes being stripped of certain types of low-level and high-level information, GSP effects were still observed on the early evoked potentials. In other words, despite an impoverished stimulus set, naturalness and spatial boundary effects were still observed arguably due to the processing of mid-level global information.

The results of Experiment 1 can be compared to the results of the 96 scenes and Hansen et al. (2018) studies using naturalistic scene images, suggesting that the global structure of the scene is sufficient to activate the processes responsible for the early, visually-evoked potentials. All three studies show that at the level of the P2 component closed scenes elicit a higher response than open scenes. Differences between the results for the three studies were also observed. For example, while a spatial boundary effect was seen at the P2 timepoint for both studies, for the 96 scenes study this effect was restricted to only natural scenes. Yet another difference lies at the naturalness effect on the N1

component. Even though a naturalness effect was seen in all three studies, with manmade scenes evoking a more negative response than natural scenes observed during Experiment 1, the opposite pattern was manifest in the Harel et al. (2016) and Hansen et al. (2018) studies, with natural scenes evoking a more negative response than manmade scenes. Arguably, this could be attributed to differences in stimulus quality between the artificial and naturalistic scene images. Further, the naturalness effect in Experiment 1 was restricted to the right hemisphere while in the 96 scenes study the effect was seen across both hemispheres. In spite of the small differences in the results of the current experiment and the other two studies, together, all three studies converge on the idea that global scene properties (GSPs) are essential in driving the early visually-evoked responses to scenes. Notably, Experiment 1 shows that the impact of these global properties can be observed independently of color information and rich semantic detail.

It has been shown humans can determine scene gist within 100-120 milliseconds of stimulus presentation (Potter, 1976), even under impoverished viewing conditions such as low spatial frequency (Schyns & Oliva, 1994). This leads to the question of how this feat is accomplished? What are the underlying neural mechanisms which allow us to categorize scenes rapidly and efficiently? A prominent suggestion has been Oliva & Torralba's (2001) notion of the spatial envelope and its subsequent corollary of GSPs (Greene & Oliva, 2009) suggesting a set of qualities could be used to describe the dominant structure of a scene fast and accurately. We show here, utilizing the high temporal resolution of EEG as a measurement, that indeed global scene information can be extracted as early as 100- 220 ms post stimulus onset. As noted above, our findings

corroborate previous findings by Harel et al. (2016) and Hansen et al. (2018) suggesting that naturalness and spatial boundary are encoded during the early stages of scene processing. Using an impoverished stimulus set, devoid of color and high-level semantic detail, we could still observe the effects of naturalness and spatial boundary during the early stages of scene processing. Thus, irrespective of the specific appearance of the scene images, our brain computes very early on the global properties of its surrounding visual environment.

Other studies have shown the importance of global representations as the drivers of rapid scene recognition. In a study done by Oliva & Torralba (2008) participants had to rank 500 natural scene images along seven global properties (openness, expansion, mean depth, temperature, transience, concealment, and navigability). Oliva & Torralba then asked to what extent were these global properties used during rapid scene recognition. This was done by having participants perform a rapid scene recognition task where they were presented with a target category (desert, field, lake, mountain, etc.) and were instructed to respond as quickly and accurately as possible to whether each image- either target or distractor image- represented the target category. The results of the experiment show that with short reaction time, global property information is used to identify whether the images presented represent the target category. The second part of this study used a naïve Bayes classifier (Murphy, 2006; [www.cc.gatech.edu](http://www.cc.gatech.edu))- an algorithm used to determine the probability between events- was used to determine the extent to which one could use global property information to predict rapid scene recognition. The results showed no significant difference in performances between the



classifier and human observers. This shows that global property information can accurately predict human performance at rapid scene recognition. In order to determine whether another representation could accurately predict rapid scene recognition two models of local representation were used: the local semantic concept model (Vogel & Schiele, 2007) and prominent object model (Biederman, 1981). While it was shown that local representation is important when determining a scene identity, it is the global representation that is most important during rapid scene recognition.

Another study by Brady et al. (2017) proposed rapid scene recognition is achieved by looking at the global texture ensemble defined as “simple spatial patterns of orientation information” (Brady et al., 2017, p.1160). Devised over three experiments, the first experiment was conducted to establish a correlation between spatial ensemble processing and rapid scene recognition by determining whether the same sensitivity to changes of high spatial frequency Gabor elements (Alvarez & Oliva, 2009) would be seen during an object recognition task. The object recognition task first used by Davenport & Potter (2004) measures scene perception by having participants recognize objects that appear either in informative backgrounds (a desk in an office) or uninformative backgrounds (a desk in a field). The only difference between conditions is the scene background. Once the connection between spatial ensemble processing and rapid scene recognition was established, over the following two experiments the same object recognition task was used to determine whether global ensemble texture is relevant for scene representations and if it is equally useful in the activation of object recognition systems. The results show while preserving global ensemble texture information in

scenes is enough for rapid scene perception, preserving the same information insufficient for object recognition.

The results of Experiment 1 allow us to reject the null hypothesis ( $H_0$ ) that the diminishment of color and semantic information from the scenes results in diminished spatial boundary and naturalness effects seen during the P2 and N1 timepoints and instead accept the actual hypothesis ( $H_a$ ) that removal of color and semantic information from scenes would not result in diminishment of spatial boundary and naturalness effects during the P2 and N1 time windows. Also, spatial boundary and naturalness effects were found at the earlier P1 time window as well.

While the study was able to show the importance of GSPs during early scene perception, questions regarding the manipulations done in the study need to be addressed. For example, our study involved the elimination of color information. But is color information an important enough dimension during scene perception that its elimination would drastically change how one perceives scenes? Studies addressing this question have been met with differing results. The role of color in scene perception in general and the processing of GSPs in particular is still debated. On the one hand, a study done by Delorme et al. (2000) suggested that color information is not essential for rapid recognition of natural scenes. They asked both human and monkey participants to categorize food vs. non-food images and animal vs. non-animal scene images when color was removed and found that accuracy and latency information was unaffected by color elimination. In contrast, work by Oliva & Schyns (2000) which had participants view color-diagnostic and non-color-diagnostic stimuli and name these images as quickly and

accurate as possible showed that color information is important when the scenes are diagnostic of the scene category. Reaching a similar conclusion, Castelhana & Henderson (2008) presented participants with either color or monochrome photographs varying in their duration and demonstrated that color information was essential to activate the mechanisms responsible for scene gist.

Given these conflicting results, how then, can we evaluate the role that color information plays in driving the ERP response to GSPs? Future studies examining GSPs effect on ERP correlates of scene perception should present color and non-color images to determine the importance of color information on scene perception. One example of such a study is Lowe et al. (2018). They presented participants images of color photographs and colorless line drawings with the task of remembering the images presented. Following this memorization phase, participants were required to recall whether an image presented was new or old during the test-phase. This study shows that rapid scene recognition can occur when either the color photographs or the colorless line drawings are presented. This suggests that since rapid scene recognition can occur with color as well as with non-color scene images. Low-level color information is not needed for rapid scene recognition to occur. Similarly, we were able to show processing of mid-level global information still occurs in spite of the fact color information was eliminated from the scene stimuli.

Another question of concern was whether high-level effects associated with removal of high-level semantic information were really removed. While much of the detail relating the identity of the scene was stripped from the images there remained still

some objects that informed about the identity of the scene. For example, for some of the manmade scenes despite having been stripped of much object information, there remained a city landscape containing skyscrapers acting as possible objects enabling scene perception. I am not sure if it is entirely possible to completely rid the scene of high-level object information when eliminating object information from a scene without losing its essence as a unique semantic category, that is, as a unique, namable categorical identity.

Even though we were able to show the importance of GSPs during the early stages of scene processing in Experiment 1, one may argue that the obtained results were due to the processing of specific local features, particularly texture cues, rather than the global spatial layout. Thus, in Experiment 2 we asked whether early scene perception is dominated by global or local information. Experiment 2 was identical to Experiment 1 with one major difference: In Experiment 2, the orientation of the normally upright stimuli was inverted 180° for half the trials (576 trials) so participants were viewing half the stimuli upside down. The rationale for this being that when stimuli are upside down global information contained within the scene is disrupted while the local information contained within the scene is preserved. During the P2 time point, we observed significant effects of the GSPs of naturalness and spatial boundary with manmade scenes eliciting a higher amplitude than natural scenes and closed scenes eliciting a higher amplitude than open scenes, respectively. However, no orientation effects were seen. At the level of the N1 component no significant effects of naturalness and spatial boundary were found as well as a lack of any orientation effects. This not only goes in contrast to

previous studies showing significant effects of spatial boundary and naturalness seen during the N1 component time point (Harel et al., 2016; Hansen et al., 2018). It also goes against N1 component inversion studies showing a larger more delayed N1 amplitude when shown inverted facial stimuli compared to upright facial stimuli (Rossion & Delvenne et al., 1999; Rossion & Gauthier et al., 1999; Rossion & Gauthier, 2002; Sedeh & Yovel, 2010). Interestingly, at the level of the P1 component there appeared to be multiple significant interactions, follow-up analyses of these interactions revealed no significant effects except for a naturalness by orientation interaction approaching significance.

Since the GSP effects on the P2 and N1 amplitude were not impacted by inversion and given that both the upright and inverted scenes contained the same local featural information, the current results suggest that surprisingly, at least with the current set of scenes, GSP-evoked responses not only reflect the extraction of global spatial layout, but also the extraction of local scene information. Note, however, that while no inversion effects were found at the P2 and N1 time windows, there was an interaction involving an effect of inversion at the P1 time window (approaching significance). This effect could be due to attentional mechanisms that have been shown to occur as early as 75 msec post stimulus onset corresponding to the start of the P1 component (Luck et al., 1990). Speculatively, it might be possible that the inverted natural scenes appeared so unusual, that they captured the attention of the participants, thereby leading to an enhancement of the evoked response to these stimuli. A future study should be conducted specifically looking at scene inversion and attention to determine if a significant interaction really

occurs at the P1 level and getting a better understanding of the processes taking place during the P1 component allowing for the processing of inverted scenes.

Kauffmann et al. (2015) showed that visual analysis follows a coarse-to-fine (C-t-F) processing sequence with low spatial frequencies (LSF) more rapidly being processed than high spatial frequencies (HSF). This was shown by exposing participants to stimuli in which two images of differing spatial frequencies and semantic information were superimposed with different presentation times. When the presentation time was 30 milliseconds, perception of the images was dominated by LSF information and when the presentation time was 150 milliseconds, perception of the images were dominated by HSF information. This view is in line with the global precedence hypothesis (Navon, 1977), which suggests that the processing of the global structure comes first followed by the more fine-grained analysis of the local details in addition to global recognition interfering with local recognition but not vice versa. According to Navon this means that when perceiving a scene, it is “decomposed rather than built up” (Navon, 1977, p.354). How do the results of Experiment 2 correspond to this theory? Putting aside the possibility of attentional effects on P1 (see above), the sensitivity of the P1 to inversion might reflect an earlier global (‘coarse’) response while the N1 and P2 components (and their insensitivity to inversion) could reflect the processing of both global and local information at those later time windows. While this suggestion is tentative and requires further research, it highlights the utility of using ERPs to study the time course of scene perception, and its constituent visual processes.

Comparing the results of Experiment 2 to Experiment 1 show slightly mixed results. At the level of the P2 component, the observed effects (irrespective of scene orientation) were mostly consistent across both Experiment 1 and Experiment 2. Surprisingly, at the level of the N1 component no significant effects were found despite having seen spatial boundary and naturalness effects during Experiment 1. At the level of the P1 component there appeared to be significant interactions between GSPs and orientation, but follow-up analyses revealed none of these interactions to be significant (apart from the approaching significance interaction between naturalness and orientation). These results differ from what was shown in Experiment 1 with significant effects of both spatial boundary and naturalness occurring at the P1 time window.

Note, however, that it is possible that the results observed in Experiment 2 are not due to the manipulation of the stimuli itself but from other extraneous factors. For example, participants viewed stimuli that were both in the upright and inverted orientation. Just as in Experiment 1, these images were used for a total of 1176 trials, however, unlike Experiment 1 participants did not view the same stimuli twice, due to differences in orientation. This might explain the lack of GSP effects observed at the N1 and P1 timepoints, due to reduced statistical power. A future study should seek to double the number of trials to 2352, so each stimulus is presented twice as was done in Experiment 1. This could increase the likelihood of effects being seen due to the increased statistical power.

In addition, the lack of expertise with inverted scene stimuli could have contributed to the lack of effects. Whether in the real world or observing an image of a scene, objects, people, etc. people are used to looking at things in an upright configuration. Having participants view stimuli in an inverted configuration is something they are not used to doing thus resulting in the lack of effects seen. Future scene inversion studies should factor in expertise and have participants do multiple sessions of a scene inversion study.

### **Summary and Conclusions**

Over the course of both experiments, we have presented participants with scene stimuli in upright and inverted orientations. The scene stimuli were stripped of low-level color information and high-level semantic scene category information, allowing us to examine the independence of the mid-level global scene property from these factors. In prior experiments using naturalistic scene images the GSP effects shown to occur on the early ERP components could have co-occurred with non-global information contained within the scenes, thus explaining the observed results. For example, natural scenes contain a lot of green patches (i.e., grassy fields) compared to manmade scenes. Also, manmade scenes may be inferred by specific objects contained within the scene that are not present in natural scenes. We reasoned that if non-global factors really drive the GSP effects, then the GSP effects of spatial expanse and naturalness would be reduced or altogether eliminated when using an ‘impoverished’ stimulus set. The results show, however, robust GSP effects of spatial boundary and naturalness occurring 220 ms post stimulus onset corresponding to the P2 component. These effects were seen when stimuli



were either in the upright or inverted orientations. The results show us that by 220 ms the brain extracts the global spatial layout during rapid scene recognition, utilizing both global and local information.

Unlike the P2 component, GSP effects were less consistent at the N1 level. Spatial boundary and naturalness effects were observed when viewing the stimuli in the upright orientation but when viewing stimuli in the inverted orientation the GSP effects we observed earlier were eliminated, as well as any other effects. It is possible the N1 component is not the best index to use for categorical scene perception, as the N1 component is a face-sensitive component, showing a higher response to faces than to scenes in contrast to the P2, which is higher in response to scenes than it is to faces (Harel et al., 2016).

Like the P2 and N1 components during Experiment 1, naturalness and spatial boundary effects were seen at the P1 time window. However, unlike the other two components, at the P1 component there was an orientation interaction nearing significance. As previously mentioned, attentional mechanisms have been shown to occur at the time point corresponding to the P1 time point (Hillyard & Anllo-Vento, 1998; Luck et al., 1990). However, the observed results seen at the P1 component could be due to the fast, initial processing of coarse global information contained within the scene. More research needs to be done on the P1 component to have a better understanding of the cognitive processes taking place allowing for the processing of global and local information.

Together, the current work sheds light on the underlying processes at the basis of the electrophysiological responses to upright and inverted artificially-generated scenes. We conclude that during early scene perception both global and local information are important when rapidly perceiving scenes and suggest using ERPs as a powerful tool to understand the time course of global and local scene processing during early scene perception.

### **Limitations and Future Directions**

In Experiment 1, of the seventeen participants, fourteen (82%) were males. This was not intentional but was instead due to the participants who were available and signed up to be in the experiment. Possible gender differences should have been explored due to previous ERP studies showing gender differences. For example, in a study done by Pfabigan et al. (2014) it was shown women have enhanced P1 amplitudes compared to men after viewing facial stimuli. While another study showed men have higher amplitudes than women on the P1 and P3b components but women presented higher amplitudes in the N1 than men (Vaquero et al., 2009). This suggests in the future more effort should be placed in attracting a more even distribution of male and female participants in order to reduce any kind of gender effects possibly causing the results observed.

In Experiment 2, while we were able to find mid-level global effects of spatial boundary and naturalness on the P2 and P1 components, we failed to find orientation effects at the level of the P2 and N1 components. These lack of orientation effects could be due to the orientation of the stimuli in our study. In our study, participants viewed

stimuli in a 180° upside-down orientation but it has been shown that when an image is orientated 180° perceptually the image has been completely reversed but the orientation bias (e.g. the predominant edge orientations in the scene) have remained the same (Loschky et al., 2015) explaining the lack of orientation effects observed at the P2 and N1 components. Future EEG orientation studies should take this into account and experiment with stimuli that have been rotated 45°, 90°, or 135° in order to determine if orientation effects will be observed.

In both Experiments 1 and 2, we were able to show the importance of the global scene properties of naturalness and spatial boundary during early scene perception particularly corresponding to the P2 timepoint as both naturalness and spatial boundary effects were observed during that timepoint (220 ms). However, what has yet to be seen is if these global effects would be seen when doing a texture discrimination task in which textures naturally found in natural scenes were swapped with textures naturally found in manmade scenes and vice versa. Doing this kind of manipulation would preserve many of the low-level features while disrupting the higher-level features of the texture appearance.

## References

1. Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *PNAS*, 106, 7345-7350.
2. Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996) Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci*, 8(6): 551-565.
3. Biederman, I. (1981). Do background depth gradients facilitate object identification. *Perception*, 10, 573-578.
4. Biederman, I. (1987). Recognition-by-Components: A Theory of Human Image Understanding. *Psychological Review*, 94, 115-147.
5. Blau, V.C., Maurer, U., Tottenham, N., & McCandliss, B.D. (2007). The face-specific N170 component is modulated by emotional facial expression. *Behavioral and Brain Functions*, 3, 3-13.
6. Bötzel, K., Schulze, S., Stodieck, S. R.G. (1995). Scalp topography and analysis of intercranial sources of face-evoked potentials. *Exp Brain Res*, 104, 135-143.
7. Brady, T. F., Shafer-Skelton, A., & Alvarez, G. (2017). Global Ensemble Texture Representations are Critical to Rapid Scene Categorization. *Journal of Experimental Psychology*, 43, 1160-1176.

8. Caharel, S., Leleu, A., Bernard, C., Viggiano, M-P., Lalonde, R., & Rebai, M. (2013). Early holistic face-like processing of Arcimboldo paintings in the right occipito-temporal cortex: Evidence from the N170 ERP component. *International Journal of Psychophysiology*, 90, 157-164.
9. Castelhamo, M. S., & Henderson, J. M. (2008). The Influence of Color on the Perception of Scene Gist. *Human Perception and Performance*, 34, 660-675.
10. Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, 15, 559-564.
11. de Graef, P., Christiaens, D., & D'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, 52, 317-329.
12. Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: a study in monkeys and humans. *Vision Research*, 40, 2187-2200.
13. Dering, B., & Donaldson, D. I. (2016). Dissociating Attention Effects from Categorical Perception with ERP Functional Microstates. *PLoS ONE*, 11, 1-18.
14. Fergus, R., Perona, P., & Zisserman, A. (2003). Object Recognition by Unsupervised Scale-Invariant Learning. *IEEE Computer Society Conference on Computer and Vision Pattern Recognition*.
15. Friedman A. (1979) Framing pictures: The role of knowledge in automatized encoding and memory of gist. *Journal of Experimental Psychology*, 108, 316-355.

16. Fu, S., Caggiano, D. M., Greenwood, P. M., & Parasuraman, R. (2005). Event-related potentials reveal dissociable mechanisms for orientating and focusing visuospatial attention. *Brain Res Cogn Brain Res*, 23, 341-353.
17. Greene, M. & Li, F. F. (2014). Visual categorization is automatic and obligatory: Evidence from a Stroop-like paradigm. *Journal of Vision*, 14, 1-11.
18. Greene, M. R. & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58, 137-176.
19. Greene, M. R. & Oliva, A. (2009b). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychol Sci.*, 20, 464-472.
20. Greene, M. R., & Oliva, A., (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58, 137-176.
21. Groen, I. I. A., Silson, E. H. & Baker, C. I. (2017). Contributions of low- and high-level *Journal of Vision*, 15, 1-9.
22. 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.
23. 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, 1-18.

24. Harel, A., Kravitz, D., & Baker, C. I. (2013). Beyond perceptual expertise: revisiting the neural substrates of expert object recognition. *Frontiers in Human Neuroscience*, 7, 1-12.
25. Hillyard, S. A., Vogel, E. K. & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society B.*, 353, 1257-1270.
26. Intraub, H. (1981). Rapid Conceptual Identification of Sequentially Presented Pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 604-610.
27. Joubert, O. R., Rousset, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast Categorization and object interference. *Vision Research*, 47, 3286-3297.
28. 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 recording. *Advances in Neural Information Processing Systems*, 10, 894-900.
29. Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, 107, 49-57.
30. Köhler, S., Crane, J., & Milner, B. (2002). Differential Contributions of the Parahippocampal Place Area and the Anterior Hippocampus to Human Memory for Scenes. *Hippocampus*, 12, 718-723.

31. 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. *The Journal of Neuroscience*, 31, 7322-7333.
32. Kravitz, D. J., Vinson, L. D., & Baker, C. I. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences*, 12, 114-122.
33. Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nature Neuroscience*, 4, 533-539.
34. Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *PNAS*, 99, 9596-9601.
35. Loschky, L.C., & Larson, A.M. (2010). The natural/man-made distinction is made prior to basic-level distinctions in scene gist processing. *Visual Cognition*, 18, 513-536.
36. 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.
37. Luck, S. J. (2012). Event-related potentials. *APA Handbook of Research Methods in Psychology*, 1, 1-18.
38. Malcolm, G. L., Groen, I. I. A., & Baker, C. I. (2016). Making sense of Real-World Scenes. *Trends in Cognitive Sciences*, 20, 843-856.



39. Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is Special about Face Recognition? Nineteen Experiment on a Person with Visual Object Agnosia and Dyslexia but Normal Face Recognition. *Journal of Cognitive Neuroscience*, 9, 555-604.
40. Murphy, K. P. (2006). Naïve Bayes Classifiers. 1-8.
41. 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. *Journal of Cognitive Neuroscience*, 26, 2287-2297.
42. Navon, D. (1977). Forest Before Trees: The Precedence of Global Features in Visual Perception. *Cognitive Psychology*, 9, 353-383.
43. Oliva, A. & Schyns, P. G. (2000). Diagnostic Colors Mediate Scene Recognition. *Cognitive Psychology*, 41, 176-210.
44. Oliva, A. & Torralba A. (2001). Modeling the shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42, 145-175.
45. Oliva, A. & Torralba A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, 155, 23-36.
46. Oliva, A. (2005). Gist of the Scene. *Neurobiology of Attention*, 251-256.

47. Park, S., Brady, T. F., Greene, M., & Oliva, A. (2011). Disentangling Scene Content from Spatial Boundary: Complementary Roles for the Parahippocampal Place Area and Lateral Occipital Complex in Representing Real-World Scenes. *The Journal of Neuroscience*, 31, 1333-1340.
48. Peirce, J. W (2015). Understanding mid-level representations in visual processing. Properties to neural processing of visual scenes in the human brain. *Philosophical Transactions Royal Society B* 372: 20160102.
49. Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G.A., Ritter, W., Ruchkin, D.S., Rugg, M.D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37(2), 127-152.
50. Potter, M. (1976). Short-Term Conceptual Memory for Pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 2, 509-522.
51. Rhodes, G. (1993). Configurational coding, expertise, and the right hemisphere advantage for face recognition. *Brain and Cognition*, 22, 43-63.
52. Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition*, 47, 25-57.
53. Riesenhuber, M. & Poggio, T. (1999) Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019-1025.
54. Rivolta, D., Palermo, R., Schmalzl, L., & Williams, M. A. (2012). An early category-specific neural response for the perception of both places and faces. *Cogn Neurosci*, 3, 45–51.

55. Ross, M. G. & Oliva, A. (2010). Estimating perception of scene layout properties from global image features. *Journal of Vision*, 10, 1-25.
56. Rossion, B. & Jacques, C. (2011). The N170: understanding the time-course of face perception in the human brain. *The Oxford Handbook of Event-Related Potential Components*, Oxford University Press, 115-142.
57. Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, 51, 1297-1311.
58. Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysical face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage*, 39, 1959-1979.
59. Rossion, B., Delvenne, J.-F., Debatisse, D., Goffeaux, V., Bruyer, R., Crommelinck, M., & Guérit, J.-M. (1999). Spatio-temporal localization of the face inversion effect: an event-related potentials study. *Biological Psychology*, 50, 173-189.
60. Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (1999). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *NeuroReport*, 11, 69-74.
61. Rousselet, G. A., Joubert, O. R., Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-word natural scenes? *Vision Cognition*, 12, 852-877.

62. Sato, N., Nakamura, K., Nakamura, A., Sugiura, M., Ito, K., Fukuda, H., & Kawashima, R. (1999). Different time course between scene processing and face processing: a MEG study. *Neuroreport*, 10, 3633–3637.
63. Schyns, P.G. & Oliva, A. (1994). FROM BLOBS TO BOUNDARY EDGES: Evidence for time- and spatial-Scale- Dependent Scene Recognition. *Psychological Science*, 5, 195-200.
64. Schyns, P.G. & Oliva, A. (1997). Flexible, diagnostically-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, 26, 1027-1038.
65. Tanaka, J. W. & Curran, T. (2001). A NEURAL BASIS FOR EXPERT OBJECT RECOGNITION. *Psychological Science*, 12, 43-47.
66. Tanaka, J. W. & Farah, M. (1993). Parts and Wholes in Face Recognition. *The Quarterly Journal of Experimental Psychology*, 46A, 225-245.
67. Vogel, J., & Schiele, B. (2007). Semantic scene modeling and retrieval for content-based image retrieval. *International Journal of Computer Vision*, 72, 133-157.
68. 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, 10573-10581.
69. Welchman, A. E., Deubelius, A., Conrad, V., Bühlhoff, H. H., & Kourtzi, Z. (2005) 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience*, 8, 820-827.

70. Yamins, D. L. K & DiCarlo, J. J (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19, 356-365.
71. Young, A. W., Hellawell, D., & Hay, D. C., (1987). Configural information in face perception. *Perception*, 16, 747-775.
72. Zhang, H., Houpt, J. W., & Harel, A. (2019). Establishing reference scales for scene naturalness and openness. *Behavior research methods*, 51(3), 1179-1186.

## Appendix

**Table 3:** Summary of results from both Experiments 1 and 2.

<b>P2</b>	<b>Experiment 1</b>	<b>Experiment 2</b>
Spatial Expanse	Closed > Open	Closed > Open
Naturalness	Manmade > Natural	Manmade > Natural
Hemisphere	RH > LH	RH > LH
Hemisphere x Spatial Boundary	RH: Closed > Open LH: Closed > Open	n.s.

<b>N1</b>	<b>Experiment 1</b>	<b>Experiment 2</b>
Spatial Expanse	Closed > Open	n.s.
Naturalness	n.s.	n.s.
Hemisphere	n.s.	n.s.
Hemisphere x Naturalness	RH: Manmade > Natural LH: n.s.	n.s.

<b>P1</b>	<b>Experiment 1</b>	<b>Experiment 2</b>
Spatial Boundary	Closed > Open	n.s.
Naturalness	Natural > Manmade	n.s.
Hemisphere	n.s.	RH > LH
Naturalness x Orientation	n.s.	Manmade Inverted > Manmade Upright Natural Upright > Natural Inverted

