Distribution Agreement

Zirui (Ray) Chen March 24, 2022

The global color of a scene is relevant for human visual scene discrimination

By

Zirui (Ray) Chen

Daniel D. Dilks, PhD.

Advisor

Department of Neuroscience and Behavioral Biology

Daniel D. Dilks, PhD.

Advisor

Michael Crutcher, PhD.

Committee Member

Elizabeth Newman, PhD.

Committee Member

March 24, 2022

The global color of a scene is relevant for human visual scene discrimination

By

Zirui (Ray) Chen

Daniel D. Dilks, PhD.

Advisor

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Sciences with Honors

Department of Neuroscience and Behavioral Biology

2022

Abstract

By Zirui (Ray) Chen

Humans are incredibly good at discriminating places, or "scenes", (e.g., to recognize an image as a beach and not a kitchen). But are th precise stimulus features that humans use to discriminate one scene from another. In this study, I hypothesized that the global color of a scene is a stimulus feature that humans use for differentiating visual scenes, but not objects, and predicted that 1) behaviorally, participants will show a more different dissimilarity judgement pattern for color versus grayscale scenes than for color versus grayscale objects (Experiment 1), and 2) neurally, color versus grayscale scenes will be represented more differently than color versus grayscale objects. To test my hypothesis, I asked the participants to behaviorally judge the dissimilarity among a set of scene and object images that are highly variable in global color (e.g., a forest full of red versus green leaves) and the grayscale version of the exact same images (Experiment 1). Consistent with my hypothesis, I found a greater difference between the participants' behavioral dissimilarity judgment for color versus grayscale scenes than color versus grayscale objects. Next, I showed participants the same set of images in the functional magnetic resonance imagining (fMRI) scanner and examined the neural response of the cortical scene and object processing system (Experiment 2). Consistent with my hypothesis, I found that human brains were more sensitive to the differences between color versus grayscale scenes than color versus grayscale objects, providing neural evidence that global color is used for human visual scene processing. Collectively, these findings suggested that global color is a stimulus feature used for human visual scene discrimination.

The global color of a scene is relevant for human visual scene discrimination

By

Zirui (Ray) Chen

Daniel D. Dilks, PhD.

Advisor

A thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Science with Honors

Department of Neuroscience and Behavioral Biology

2022

Acknowledgements

I would like to express my sincerest gratitude to my advisor, Dr. Daniel D. Dilks, for his continued support throughout my undergraduate research experience. I would also like to thank my graduate student mentor, Annie Cheng, for her endless encouragement, patience, and guidance in this process. I would like to acknowledge the members of the Dilks lab, especially Chris Jones, Nathan Trinkl, Rebecca Rennert, and Dr. Yaelan Jung, for their constructive feedback and technical assistance. Finally, I would like to thank Dr. Michael Crutcher, and Dr. Elizabeth Newman for their invested time and participation on my thesis committee.

Table of Contents

Introduction..1

Experiment 1 (Behavioral)...4

    Methods..4

    Results..7

Experiment 2 (fMRI)...9

    Methods..9

    Results..12

Discussion...15

Conclusion..19

Supplement Analysis...20

Supplement Table..23

References...24

Introduction

Humans are incredibly good at discriminating scenes (e.g., recognizing an image as a beach but not a kitchen) – we can recognize a scene in as little as 19 milliseconds of display time (Greene & Oliva 2009). How are we able to achieve such a remarkable feat? In the past two decades, cognitive neuroscientists have found a dedicated cortical area in human visual cortex – the Parahippocampal Place Area (PPA; Epstein & Kanwisher, 1998) – that is selectively involved in visual scene recognition. PPA is considered "scene-selective" as it responds two to four times more to scene over non-scene images, such as common objects (e.g., an apple) or faces. Moreover, PPA has also been suggested to be involved in scene discrimination as it shows distinct response patterns for different kinds of scenes (e.g., a beach versus a kitchen; Walther et al., 2009). However, one fundamental question remains – how exactly do humans discriminate scenes?

In the existing literature, one prominent theory posited that scene discrimination may be achieved by processing the co-occurring objects within an environment (e.g., recognizing a kitchen by seeing a sink and a refrigerator; MacEvoy & Epstein, 2011). While object information can certainly be used for visual scene discrimination, a growing body of behavioral evidence suggests that visual scene discrimination can occur independently of object recognition. For example, humans can recognize a scene as fast as recognizing an object, indicating that objects do not need to be processed prior to scenes (Hagmann & Potter, 2016). Furthermore, humans can also recognize a scene image even when the object details within are blurred out (Oliva & Torralba, 2006). For instance, when presented with the image in Figure 1A, participants

Figure 1. A & B. The buildings in a street were replaced with kitchen furniture and blurred out yet participants nevertheless consistently judged the image as a street (Oliva & Torralba, 2006). **C.** An example of normally colored scenes in Oliva & Schyns, 2000. **D.** An example of abnormally colored scenes twisted from Figure 1C (Oliva & Schyns, 2000).

predominantly recognize the image is a street scene, even though the central orange blobs were in fact kitchen cabinets overlayed on the original scene image (Figure 1B). Such a response pattern demonstrated that with missing or even "wrong" objects, humans can nevertheless use some coarse information in the visual scene stimuli for visual scene discrimination. These studies collectively suggest that objects are not necessary for visual scene discrimination, and raise the question that if not through the objects, then what exactly do we rely on for scene discrimination?

Some existing studies have hinted at the role of color for visual scene discrimination. In one behavioral study (Oliva & Schyns, 2000), the participants were asked to determine whether a presented image matched with a word that was presented right before (e.g., an image of forest and the word "forest"). They found that participants were able to perform the task much faster with normally colored scenes (Figure 1B) than abnormally colored scenes (Figure 1C), suggesting that humans may rely on the global colors for visual scene discrimination. Furthermore, neuropsychological evidence also highlighted the role of color for visual scene processing. Steeves et al. (2004) reported that patient D.F. (Steeves et al., 2004) had profound visual form agnosia – impaired processing of objects and shapes but spared information of color. Intriguingly, while impaired at recognizing grayscale scenes, D.F. could surprisingly classify color scenes as well as neurotypical adults, highlighting the relevance of color to visual scene discrimination.

Building upon these findings, I aimed to directly test the role of global color of a scene for visual scene discrimination. I hypothesized that if the global color of visual stimuli is a feature that humans use for visual scene discrimination, then 1) behaviorally, participants will show a more different dissimilarity judgement pattern for color versus grayscale scenes than for color versus grayscale objects (Experiment 1), and 2) neurally, color versus grayscale scenes will be represented more differently than color versus grayscale objects (Experiment 2). I tested the predictions using both behavior in Experiment 1 and fMRI in Experiment 2.

Experiment 1 (Behavioral)

Methods

Participants

Two hundred participants with normal or corrected-to-normal vision were recruited on Amazon Mechanical Turk. All participants gave informed consent in accordance with the Emory Institutional Review Board.

Stimuli

There were four color conditions of stimuli: color scenes, grayscale scenes, color objects, and grayscale objects. Within each of the conditions, I collected 24 naturalistic images, including two color variations for each of the 12 categories (e.g., a red church image versus a blue church image, or a red apple versus a green apple) (Figure 2A). The color images were then converted into grayscale images for the stimuli in the grayscale conditions. Note that, among the

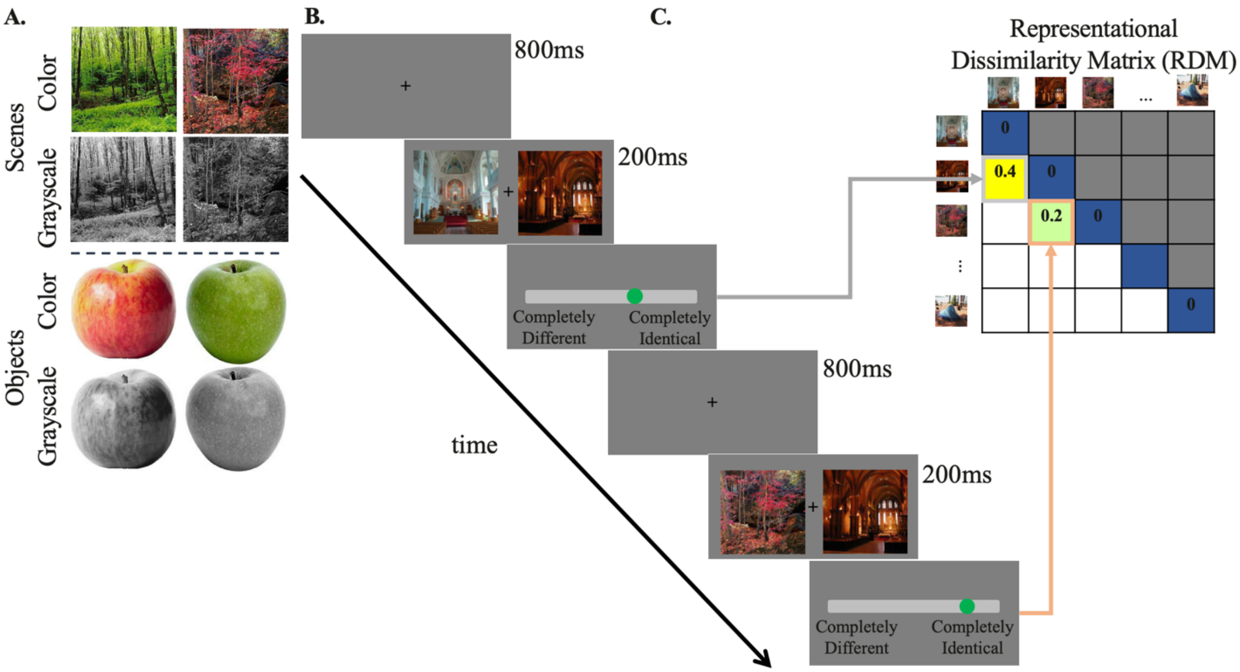

Figure 2. A. Examples from the four stimuli conditions: color scenes, grayscale scenes, color objects, and grayscale objects. **B.** A sample of two trials in Experiment 1. **C.** Demonstration of how a representational dissimilarity matrix (RDM) is created to summarize a participant's dissimilarity judgment pattern on a set of stimuli.

24 scene images, I matched the proportion of natural versus man-made and indoor versus outdoor scenes to prevent any biases that may occur on a particular type of category (Supplementary Table 1); I also matched the proportion of natural versus man-made categories for the object stimuli.

Procedures

To test the effect of global color in behavioral scene discrimination, participants were asked to judge the dissimilarity between two scene images. The experiment started with an instruction, in which the participants were only told to decide how similar or different the two images are without any explicit criteria, so that they wouldn't be biased towards using color, semantic meaning, or any other criterion for their judgement. Within each trial, the participant fixated at a plus sign at the center of the computer monitor for 800ms (Figure 2B), followed by a pair of images displayed horizontally for 200ms. Afterwards, participants had unlimited time to rate the dissimilarity of the images with a continuous slider scale, ranging from "completely different" to "completely identical", which was recorded on the backend as a dissimilarity rating ranging from 1 to 0. The whole experiment consisted of eight practice trials and 276 experimental trials, which consisted of all possible pairwise combination of the 24 stimuli within one stimuli condition. Both trial order and the left-right position of each pair of displayed images were randomized across participants. I used a between-subject design (i.e., each participant only saw the stimuli within one stimuli condition) to ensure that the participants could maintain their attention within the time span of an experiment (around 30-40 minutes).

Data Analysis

I first checked whether the participants were paying attention during the experiment by examining participants' dissimilarity judgement ratings on trials in which they were presented with two identical images. In these trials, if a participant showed an average dissimilarity ratings higher than 0.2, which indicated that one did not consider two identical images as very similar, I excluded them from the analysis. A total of 31 out of 200 participants were excluded.

Next, I summarized each participant's behavioral dissimilarity judgement for each stimulus within a stimuli condition by organizing the participants' dissimilarity judgement scores into representational dissimilarity matrices (RDMs; Figure 2C). In these behavioral RDMs, the value of each cell indicated a participant's dissimilarity judgement score between the image represented by its row position and the image represented by its column position. I then averaged across the behavioral RDMs of the participants within a condition to examine the group response pattern.

Finally, to test my prediction, I correlated the group behavioral RDMs of color and grayscale scenes, as well as color and grayscale objects (Figure 3A), seperately. Since I used a between-subjects design, I could not compute a pairwise correlation between the color versus grayscale RDMs at the subject level. Thus, instead, to compare the behavioral RDMs color versus grayscale scenes and color versus grayscale objects – I transformed the groupwise Pearson correlation coefficients (r) between color and grayscale scenes, as well as color and grayscale objects, into Fisher's z's (Equation 1) and assessed the differences by calculating $z_{observed}$ (Equation 2).

$$z = \frac{\ln\left(\frac{1+r}{1-r}\right)}{2} \qquad \text{Equation 1}$$

$$z_{observed} = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}} \qquad \text{Equation 2}$$

Results

If the global color of visual stimuli is a feature that humans use for visual scene discrimination, then the participants will show a more different dissimilarity judgement pattern for color versus grayscale scenes than for color versus grayscale objects. Consistent with my prediction, I observed a numerically lower correlation between RDMs of color and grayscale scenes ($r = .81, p < .0001, Fisher's\ z = 1.13$; Figure 3B) than between RDMs of color and grayscale objects ($r = .96, p < .0001, Fisher's\ z = 2.00$), indicating that the dissimilarity judgement patterns for color versus grayscale scenes is more different than that for color versus grayscale objects.

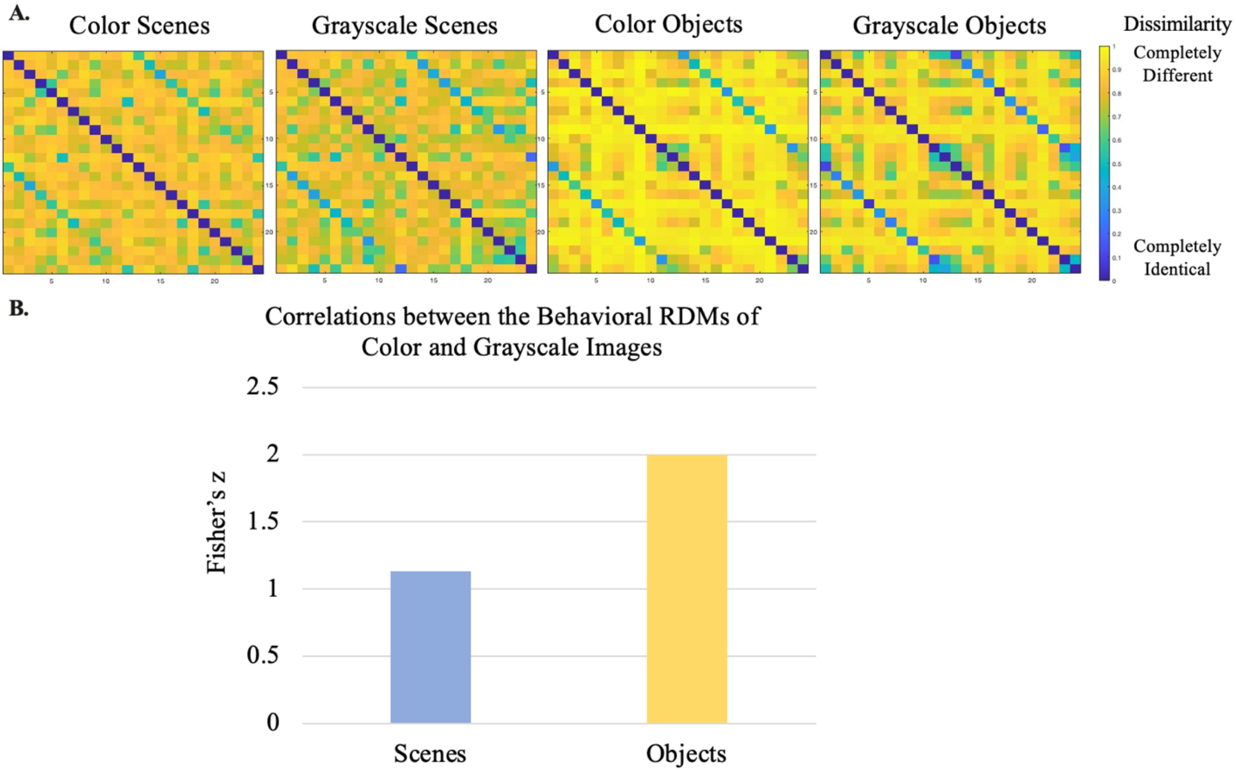

Figure 3. A. The group behavioral RDMs for the four stimuli conditions, which capture the behavioral pattern of the participants' dissimilarity judgement. Dissimilarity = The participants' averaged dissimilarity judgment score. **B.** Fisher's z value between the participants' behavioral dissimilarity judgment. The correlation between color and grayscale scenes is lower than that between color and grayscale objects, suggesting that the participants' behavioral dissimilarity judgment of color scenes versus grayscale are significantly more different than color versus grayscale objects.

Next, to directly compare the two correlations, I transformed the Pearson correlation coefficients (r) into Fisher's z's and assessed the difference. Consistently, I found a significant difference between the two correlations ($z_{observed} = -10.25, p < .001$), indicating that the participants' behavioral dissimilarity judgment for color versus grayscale scenes are indeed significantly more different than that of color versus grayscale objects.

Finally, I also further examined whether the difference between participants' behavioral dissimilarity judgment for color versus grayscale scenes were truly driven by color – not by such potential confounds as luminance or semantic information, and I found that color was indeed the driving feature. See Supplementary Analysis for details.

Experiment 2 (fMRI)

Methods

Participants

Four participants (two female) with normal or corrected to normal vision were recruited from the Emory University. All participants gave informed consent in accordance with the Emory Institutional Review Board.

Stimuli

I used the same stimuli as those in Experiment 1 (Figure 2A).

Procedures

I used a region of interest (ROI) approach in which I localized the cortical ROIs with the Localizer runs, and then used an independent set of Experimental runs to investigate the neural responses to the visual stimuli in these ROIs. The ROIs included the scene-selective PPA and the Lateral Occipital Cortex (LO; Grill-Spector et al., 2001) as a control region, which is known to be "object selective" as it responds two to four times more to object over scrambled-object images. For the Localizer runs, participants saw images of faces, objects, scenes, and scrambled objects in a block, during which they were asked to perform a one-back repetition detection task to maintain attention (i.e., to press a button whenever an image is continuously repeated twice). For the Experimental runs, I used a fast event-related paradigm (Figure 4). Each run consisted of six fixation blocks (four to six seconds each) and three repetitions of the 24 images within the same stimuli condition. Each image was presented for one second, followed by a fixation screen for a pseudo-randomly jittered inter-stimulus interval ranging from three to five seconds. To

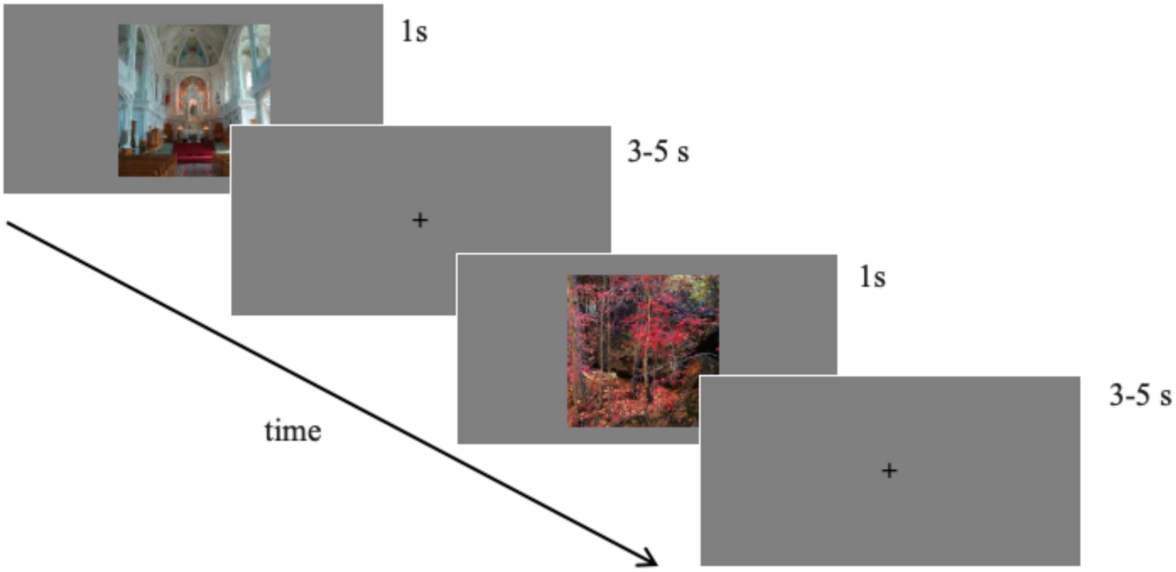

Figure 4. A sample of two trials in Experiment 2. The inter-stimulus-interval was pseudo-randomly jittered, ranging from three to five seconds. Participants were asked to press a button to indicate whether they like or dislike an image following the presentation of each image

maintain attention, the participants were asked to press one of two buttons to indicate whether they liked or disliked the image following the presentation of each image. The order of the images in each run was randomized. Each participant completed two Experimental runs per stimuli condition. To prevent the participants' knowledge of the images' color from biasing their neural response to the grayscale stimuli, participants completed the Experimental runs for the grayscale conditions before the color conditions. The entire session for each participant lasted for approximately 90 minutes.

The fMRI scanning sessions were completed on a 3T Siemens Trio scanner at the Facility for Education and Research in Neuroscience at Emory University. Functional images were acquired with a 32-channel head matrix coil and a gradient echo single-shot planar imaging sequence (28 slices, repetition time = 2s, echo time = 30ms, voxel size = 1.5 x 1.5 x 2.5 mm, 0.2mm interslice gap). Slices were oriented approximately parallel to the calcarine sulcus in order to fully cover the occipital and temporal lobes and most of the parietal and frontal lobes. The whole-brain, high-resolution anatomical images were also acquired for each participant.

Data Preprocessing

I analyzed the fMRI data using the FSL software (Smith et al., 2004) and custom MATLAB code. Preprocessing included skull-stripping (Smith, 2002), linear trend removal, three-dimensional motion correction, and temporal smoothing with FSL's MCFLIRT tool. After preprocessing, I fit the preprocessed data with a double-gamma function to estimate the hemodynamic response function to each of the stimulus conditions.

To define the ROIs, I first estimated the location of PPA and LO within each individual using a probabilistic parcel of these ROIs (Julian et al., 2012). Next, among the voxels within these probabilistic parcels of each ROI, I selected the top 100 voxels that showed the strongest scenes > objects response from the independent Localizer runs and defined them as PPA, and the top 100 voxels that showed the strongest object > scrambled objects response from the independent Localizer runs and defined them as LO.

Data Analysis

To examine the neural response of an ROI, I extracted the multi-voxel activity pattern of the ROIs to each stimuli condition in the Experimental runs. Next, I conducted the representational dissimilarity analysis on the multi-voxel activity pattern of each ROI. Specifically, within each condition (e.g., color scenes), I correlated the multi-voxel activity pattern of an ROI to one image with that to another image for all possible pair-wise combinations of the stimuli. Then, I subtracted the Pearson correlation coefficients from one to obtain a measure of pairwise dissimilarity, which was then organized into the neural RDMs (Figure 6A). As such, in these neural RDMs, the value of each cell was the dissimilarity between the ROI's multi-voxel response to the image represented by its row position and that to the image represented by its column position. Finally, I averaged the RDMs of each ROI across both hemisphere and across the two Experimental runs of the same stimuli conditions, such that for each subject, I acquired a group RDM for each ROI (i.e., PPA and LO), and for each of the four stimuli conditions.

Results

I first validated the data from the Experimental runs by examining whether the scene and object selective ROIs demonstrated the known scene selectivity (i.e., scenes > objects) or object selectivity (i.e., objects > scenes in addition to objects > scrambled objects), respectively. To do so, I examined whether PPA shows a greater response to the scene condition, and LO shows a greater response to the object condition. Consistent with the previously known scene and object selectivity of these regions, I found that PPA's response to the color and grayscale scenes was

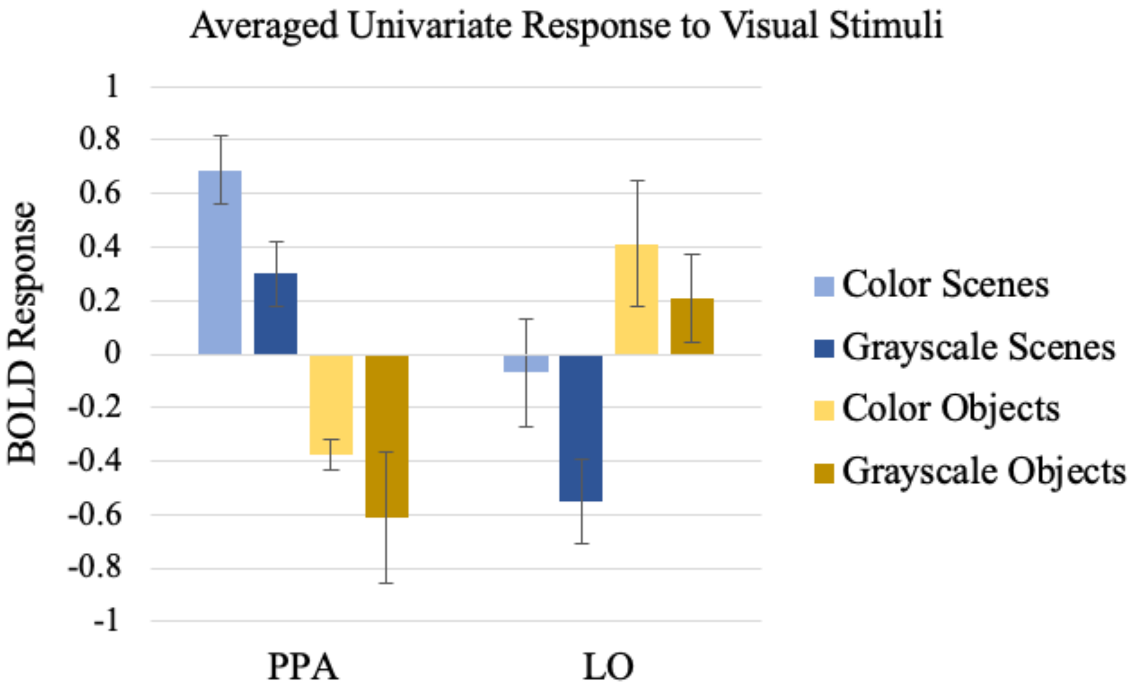

Figure 5. Neural response of PPA and LO to the different stimuli conditions ($N = 4$). PPA's response to color and grayscale scenes was higher than that to the color and grayscale objects; LO's response to the color and grayscale object was higher than that for color and grayscale scenes. I also observed PPAs' responses to color scenes to be higher than those to grayscale ones

higher than that to the color and grayscale objects ($t(9) = 6.43, p < .0001, r = .91$; Figure 5), and LO's response to the color and grayscale object was higher than that for color and grayscale scenes ($t(9) = -3.22, p < 0.01, r = .73$). Having validated the data, interestingly, I also observed that PPAs' response to color scenes was marginally higher than that to grayscale ones ($t(9) = 1.79, p = .10, r = .51$). By contrast, there was no difference between color versus grayscale objects in LO ($t(9) = .75, p = .47, r = .24$). These evidence already suggested that PPA is more sensitive to the difference between color and grayscale stimuli than LO.

Given the univariate results, I next directly tested my prediction by comparing the neural RDMs for color versus grayscale scenes and objects, respectively (Figure 6B). If the global color of visual stimuli is a feature that humans use for visual scene discrimination, then the neural

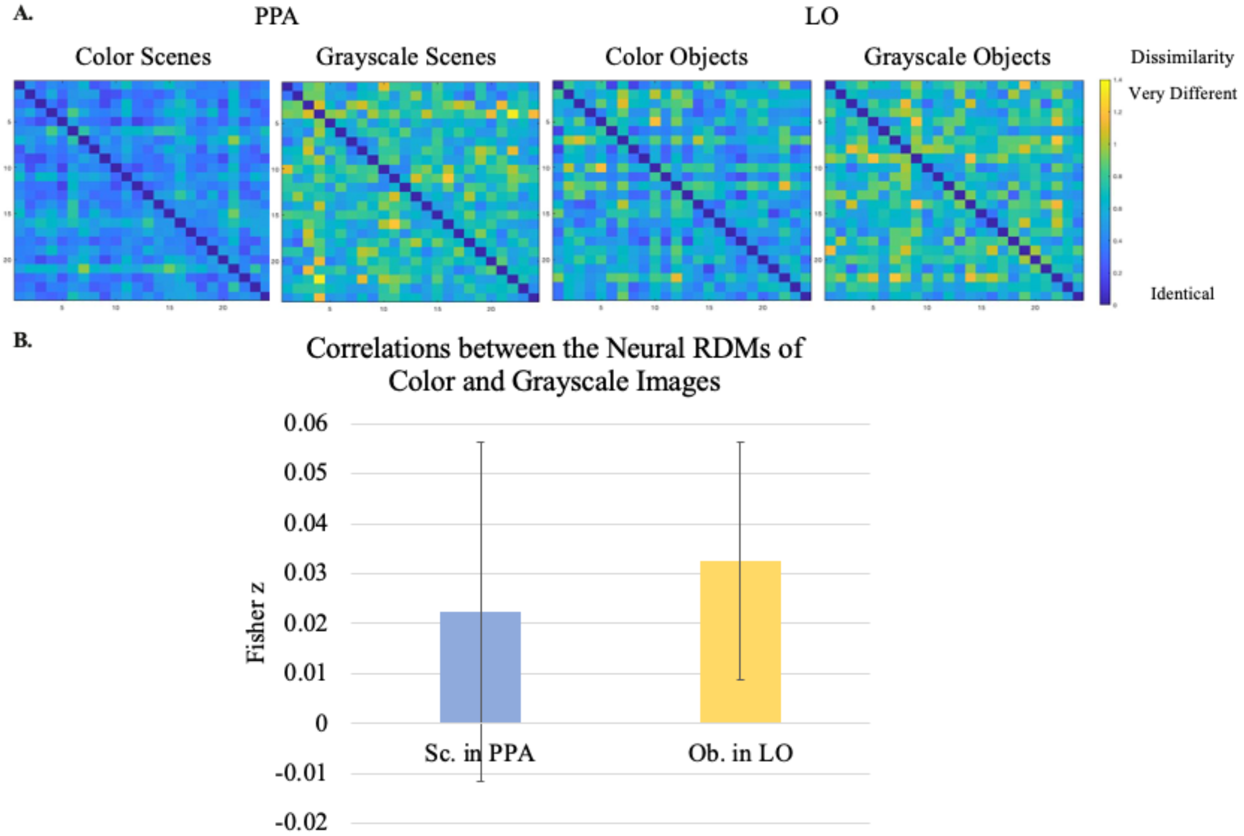

Figure 6. A. The averaged neural RDMs of PPA and LO to Color versus Grayscale Scenes and Objects, distinctively. Dissimilarity = 1 – Pearson correlation coefficient. **B.** Fisher's z values between the participants' PPA RDMs of color and grayscale scenes and that between their LO RDMs of color and grayscale objects. No significant difference was found, presumably due to lack of participants.

representation of color versus grayscale scenes in PPA should be more different than that of color versus grayscale objects in LO. Thus, I correlated the RDMs of PPA for color and grayscale scenes, and those of LO for color and grayscale objects. However, I did not observe significant difference between correlations between the PPA RDMs of color and grayscale scenes (Fisher's z: $M = .016, SE = .043\ , N = 4$; Figure 6B) than those between LO RDMs of color and grayscale objects (Fisher's z: $M = .025, SE = .021\ , N = 4$), even though the average value for scenes was lower than that for objects.

Discussion

In this study, I investigated whether the global color of visual stimuli is a feature that humans use for visual scene discrimination. In Experiment 1, I found that the participants' behavioral dissimilarity judgment for color versus grayscale scenes were more different than for color versus grayscale objects. In Experiment 2, I found that the neural representation of color versus grayscale scenes in the scene-selective PPA were more different than that of color versus grayscale objects in the object selective LO. Together, these results support my hypothesis that the global color of visual stimuli is used for humans' visual scene discrimination.

My findings that color is relevant for visual scene behavioral discrimination in Experiment 1 is consistent with previous findings for a potential role of global color in scene discrimination (Oliva & Schyns, 2000). Moreover, while Oliva & Schyns (2000) suggested that colors only mediate recognition of certain "color-diagnostic" scene categories – namely natural, outdoor scenes with prevalent color schemes (e.g., ocean is usually blue) – our study suggest that the effect of global color for human visual scene discrimination is general across natural and manmade, and indoor and outdoor scenes. One possible explanation for the difference between our findings is that Oliva & Schyns' (2000) manipulation of color (i.e., flip the red and blue of the images) for the manmade and indoor scenes might not be as drastic as that for natural and outdoor scenes, making the color-manipulated scenes to seem relatively similar to the original image, thus resulting in the null effect in manmade and indoor scenes. Furthermore, while Oliva & Schyns (2000) measured scene discrimination indirectly based upon the participants' accuracy in matching between images and names, I directly measured the participants' dissimilarity judgment pattern among the scenes. Hence, my study is likely more sensitive to detect the effect of global color in visual discrimination than was previously undetected.

In addition, our findings that color versus grayscale scenes were represented differently in PPA in Experiment 2 also shed light onto global color as a property that differentiates neural representations of visual scenes, providing insights for a recent attempt in understanding what underlies the representational dissimilarity of visual scenes in the human visual cortex (Groen et al., 2018). Furthermore, our findings for a greater response to color versus grayscale scene is also consistent with previous neuropsychological evidence from D.F. (Steeves et al., 2004) – who had a similar level of PPA response to color scenes as neurotypical adults but less PPA response to grayscale scenes – thus providing complementary evidence that color is also used for cortical scene processing in neurotypical populations. However, given our small sample size (N=4) in Experiment 2 due to difficulties in recruiting participants for in-person experiment during COVID, more data is necessary to verify the effects observed in Experiment 2.

Beyond our main findings that color matters for scene discrimination, one might be surprised by our findings that color does not play a crucial role in object recognition. Moreover, our findings may also seem to be at odds with some previous evidence for colors affecting object recognition (Biederman & Ju, 1988; Davidoff & Ostergaard, 1988). Upon further examination, however, these previous studies revealed a discrepancy in the effect of color for object recognition: colors mediate people's performance in the objects naming tasks, during which the participants were given an image and asked to name the objects, but colors had no effect on their performance in the object verification tasks, during which they verified if the given object image matched with the given name of an object category. Such a discrepancy may result from two potentially distinct cognitive process – "recognition" of individual object versus “discrimination”. In both the dissimilarity judgement task tested in Experiment 1 and the object verification task reported in the previous, it is plausible that only the “discrimination” process is involved, which may explain why these two tasks produced consistent evidence that color is not used for object discrimination. Thus, our current findings does not contradict with the existing literature concerning the role of color for object recognition.

The current study is limited in two ways. Firstly, the lack of participants in Experiment 2 limited the conclusion we could draw from the study. To test more rigidly whether PPA’s response to color versus grayscale scenes is significantly more different than that of LO for color versus grayscale objects, I will continue to collect data from more participants in the following months. Before then, it is too early to draw any conclusions based on the multi-voxel pattern activity due to the highly-noisy nature of fMRI data. Nevertheless, the univariate findings that PPA is more sensitive to the difference between color and grayscale stimuli than LO, provides promising neural evidence that color is used for scene discrimination. Secondly, the scene and object stimuli are not matched in their visual complexity, which may interact with the effect of color for visual scene discrimination. Thus, one fruitful future direction is to match the distribution of spatial frequency – which is a measure of image complexity – and other possible image statistics between the naturalistic scene and object stimuli.

If data from more participants for Experiment 2 indeed supports that color is represented in PPA and is thus used for scene discrimination, future works may look into the other two scene selective regions in human cortex – the Retrosplenial Complex (RSC; Maguire, 2001) and the Occipital Place Area (OPA; Dilks et al., 2013). Unlike PPA, RSC and OPA are suggested to be involved in scene navigation rather than scene discrimination (Dilks et al., 2022). Intuitively, global color may not be necessary for the scene navigation systems; thus, these two cortical regions might not represent global color. Investigation into these two regions will provide insights into how the differential functions of these scene selective regions are built upon the differential information extracted from the same stimuli.

Conclusion

This study investigated whether the global color of visual stimuli is a feature that humans use for visual scene discrimination. In Experiment 1, I found that global color is used for behavioral discrimination of scenes but not objects. In Experiment 2, despite limited number of participants, I found preliminary neural evidence that global color is relevant to human visual scene discrimination, as PPA is more sensitive to the difference between color and grayscale stimuli than LO. Collectively, these findings suggest that global color is a stimulus feature used for human visual scene discrimination. This study then sheds light on the relevance for global color for visual scene discrimination and provides further insight into how humans achieve the remarkable cognitive ability of visual scene discrimination.

Supplement Analysis

In Experiment 1, I further tested if the participants based their behavioral dissimilarity judgements on the global color of the scene stimuli. To do so, I created the RDM prediction models that quantified the color dissimilarity between the stimuli and examined the similarity between the behavioral RDM and the predictions models. Specifically, I first converted each image into the CIELAB color space (Figure S1A; International Commission on Illumination, 1976), in which the two color dimensions of interest are the red-green opponency and the yellow-blue opponency. I selected this color space as it was suggested to capture how humans naturally perceive color and lightness in early visual processing. In addition, the two color dimensions are independent of the third dimension, perceived lightness, which enabled me to control for the confound of general lightness to the global colors' relevance to human scene discrimination. For each image, I averaged the values in these two dimensions across all pixels to obtain the global values for both red-green opponency and yellow-opponency. Then, I compared the values between each image pair and organized them into the color prediction models (Figure S1B), in which the value of each cell indicated the Euclidean distance between the averaged value of the particular dimension across all pixels in the image represented by its row position and that by its column position. Finally, I correlated these color prediction models with the participants' averaged behavioral RDMs to observe how similar they are, or how much variance of the behavioral RDMs can be explained by the color differences that the prediction models summarized.

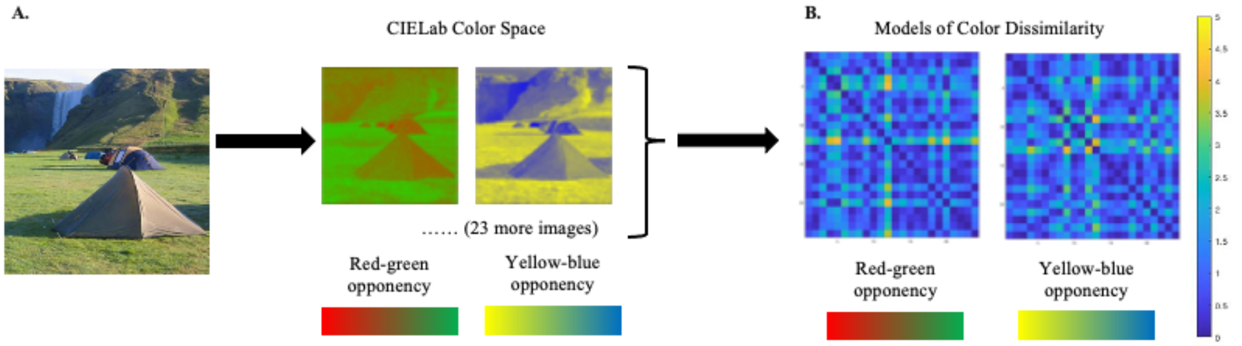

Figure S1. A. Demonstration of converting an image into the color dimensions of interest: red-green opponency and yellow-blue opponency. **B.** The color prediction models that capture the 24 color scene images' similarity relationship in terms of the image statistics of interest, respectively. Dissimilarity = Euclidean distance of the corresponding averaged image statistic among the stimuli.

If the global color of visual stimuli is a feature that humans use for behavioral scene discrimination, then in addition to the main findings of Experiment 1, the participants' behavioral pattern should also be predictable by the color differences among the scene images. That is, the color prediction models should correlate with the behavioral RDM for color scenes but not the other stimuli conditions. Hence, I correlated the two color prediction models with the behavioral RDMs of each participant who saw color scenes. For both color dimensions, the Fisher's z's were significantly higher than zero (red-green opponency: $t(38) = 7.60, p < .0001$; yellow-blue opponency: ; $t(38) = 9.64, p < .0001$; Figure S2), indicating that the differences in red-green opponency and yellow-blue opponency among the color scenes are significantly correlated with the participants' behavioral dissimilarity judgement of these scenes.

But might the correlations be driven by some confounding features (i.e., lightness or semantic categories) that the color prediction models captured by chance? If so, then I should observe a similar degree of correlation between the color prediction models and the behavioral RDMs for grayscale scenes. Contrary to this alternative hypothesis, however, for both color dimension, the Fisher's z's were not significantly above zero (red-green opponency: $t(35) = -1.22, p = .11$; yellow-blue opponency: $t(35) = -2.10, p < .05$. Thus, the correlations

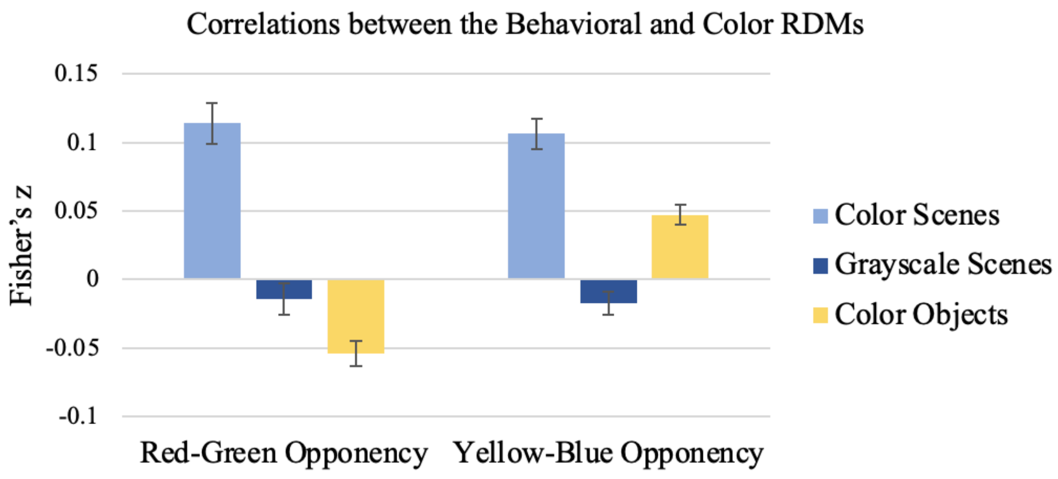

Figure S2. Fisher's z value between the prediction models and participants' dissimilarity judgments (i.e., the behavioral RDM) on color scenes and the predictions models. There is a positive and significant correlation between participants behavioral RDM and the two color prediction models, suggesting that the global color of visual stimuli is used for human behavioral discrimination of visual scenes. As the control, all four correlations between the behavioral RDM of grayscale scenes and color objects and the two color prediction RDMs were significantly lower than those for color scenes, suggesting that the found effect of color was not confounded by other features, and that the effect is specific for scene discrimination rather than any visual discrimination task.

between the color prediction models and color scene RDM is unlikely to be driven by these confounding features that the grayscale scenes share with the color ones.

Finally, to test whether color is specifically relevant for visual scene discrimination and not just any kinds of visual discrimination tasks, I directly compare the Fisher's z correlation between the two color models and the behavioral RDM of color scenes and objects. I found the correlation between the two color models and the behavioral RDM of color scenes to be significantly higher than that of color objects (red-green opponency: $t(83) = 10.03, p < .0001, r = .74$; yellow-blue opponency: $t(83) = 4.74, p < .0001, r = .46$), indicating that the global color of visual stimuli is used for human behavioral discrimination of visual scenes.

Scenes	Objects
Amusement Park (N, O)	Apple (N)
Bedroom (M, I)	Basketball (M)
Campsite (N, O)	Bonsai (N)
Cavern (N, I)	Cabbage (N)
Church Interior (M, I)	Cooler (M)
Field (N, O)	Donut (M)
House (M, O)	Globe (M)
Industrial Area (M, O)	Ice Cream (M)
Kitchen (M, I)	Jeans (M)
Street (N, O)	Leaf (N)
Swimming Pool (M, O)	Pepper (N)
Woods (N, O)	Pumpkin (N)

Supplementary Table 1. The selected 24 stimuli categories of scenes and objects. N – natural, M – man-made; I – indoor, O – outdoor.

References

Davidoff, J. B., & Ostergaard, A. L. (1988). The role of colour in categorial judgements. *The Quarterly journal of experimental psychology. A, Human experimental psychology*, 40(3), 533–544. https://doi.org/10.1080/02724988843000069

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 33(4), 1331–6a. https://doi.org/10.1523/JNEUROSCI.4081-12.2013

Dilks, D. D., Kamps, F. S., & Persichetti, A. S. (2022). Three cortical scene systems and their development. *Trends in cognitive sciences*, 26(2), 117–127. https://doi.org/10.1016/j.tics.2021.11.002

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive psychology*, 20(1), 38–64. https://doi.org/10.1016/0010-0285(88)90024-2

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392(6676), 598–601. https://doi.org/10.1038/33402

Greene, M. R., & Oliva, A. (2009). The briefest of glances: the time course of natural scene understanding. *Psychological science*, 20(4), 464–472. https://doi.org/10.1111/j.1467-9280.2009.02316.x

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision research, 41(10-11), 1409–1422. https://doi.org/10.1016/s0042-6989(01)00073-6

Groen, I. I., Greene, M. R., Baldassano, C., Fei-Fei, L., Beck, D. M., & Baker, C. I. (2018). Distinct contributions of functional and deep neural network features to representational similarity of scenes in human brain and behavior. *eLife*, 7, e32962. https://doi.org/10.7554/eLife.32962

Hagmann, C. E., & Potter, M. C. (2016). Ultrafast scene detection and recognition with limited visual information. *Visual cognition*, 24(1), 2–14. https://doi.org/10.1080/13506285.2016.1170745

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, 60(4), 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

MacEvoy, S. P., & Epstein, R. A. (2011). Constructing scenes from objects in human occipitotemporal cortex. *Nature neuroscience*, 14(10), 1323–1329. https://doi.org/10.1038/nn.2903

Maguire E. A. (2001). The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scandinavian journal of psychology*, 42(3), 225–238. https://doi.org/10.1111/1467-9450.00233

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive psychology*, 41(2), 176–210. https://doi.org/10.1006/cogp.1999.0728

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. https://doi.org/10.1016/S0079-6123(06)55002-2

Smith S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, 17(3), 143–155. https://doi.org/10.1002/hbm.10062

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23 Suppl 1, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Steeves, J. K., Humphrey, G. K., Culham, J. C., Menon, R. S., Milner, A. D., & Goodale, M. A. (2004). Behavioral and neuroimaging evidence for a contribution of color and texture information to scene classification in a patient with visual form agnosia. *Journal of cognitive neuroscience*, 16(6), 955–965. https://doi.org/10.1162/0898929041502715

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 29(34), 10573–10581. https://doi.org/10.1523/JNEUROSCI.0559-09.2009