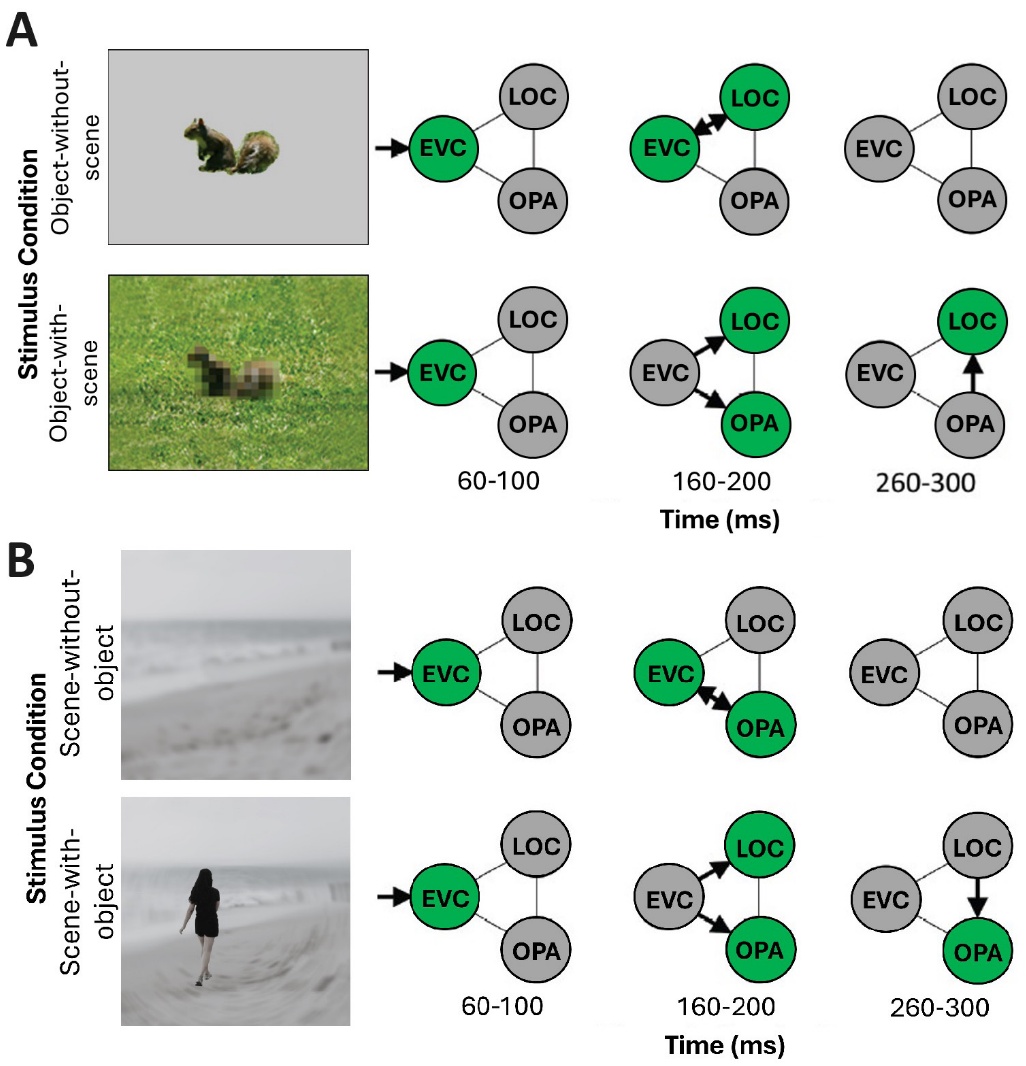
I See an Object, So I Can Make Sense of the Scene: Object Recognition in the LOC Facilitates Scene Recognition

Aaron Schnippe
Supervisors: Marius Peelen & Marco Gandolfo

Radboud University, The Netherlands

Donders Centre for Cognition, Nijmegen, The Netherlands


A
Stimulus Condition
Object-without-scene
Object-with-scene
EVC
LOC
OPA
60-100
160-200
260-300
Time (ms)
B
Stimulus Condition
Scene-without-object
Scene-with-object
EVC
LOC
OPA
60-100
160-200
260-300
Time (ms)

Abstract

Our visual environment can be distinguished into objects and scenes. While objects refer to entities such as the coffee machine on the kitchen counter, scenes refer to the overall setting, such as the “kitchen” itself. Previous research has extensively shown that scenes can facilitate object recognition and that objects can facilitate object recognition. However, whether objects can also facilitate scene recognition remains unclear. Here, we propose that object processing in the lateral occipital cortex (LOC) facilitates scene recognition when scene recognition is ambiguous. When scenes are ambiguous, objects may be integrated into the scene representation as cues to help disambiguate the scene. This would be reflected in increased repetition priming when scene repetition is accompanied by the same disambiguating object, failed priming when object and scene are incongruent, and “illusory priming” when a disambiguating object creates the illusion that an ambiguous scene has been seen before. To test these hypotheses, we will first conduct three behavioural experiments (Experiments 1–3) to establish the conditions under which object-facilitated scene recognition occurs and to identify a robust behavioural marker of object-facilitated scene recognition. In Experiment 1, we will demonstrate that objects facilitate scene recognition even when the scene is presented without objects. In Experiment 2, we will show that repetition priming for scenes is stronger when the same object appears in both presentations of the scene compared to when a different object appears. In Experiment 3, we will test whether presenting the same object with different ambiguous scenes induces illusory scene recognition. Finally, in Experiment 4, we will use transcranial magnetic stimulation (TMS) to test the causal role of the lateral occipital cortex (LOC) and the occipital place area (OPA) in object-facilitated scene recognition. Together, these experiments will provide converging evidence for the existence of object-facilitated scene recognition and will clarify the role of object processing in disambiguating scenes. This will not only advance our understanding of interactions between object and scene processing, but will also contribute to a more complete account of visual recognition when recognition is facilitated by object information.

Keywords: Scene recognition, objects, object-facilitated scene recognition, TMS, occipital place area, lateral occipital cortex

1 Introduction

Humans can perceive and recognize their visual environment very rapidly and with remarkable accuracy. When you enter your kitchen in the morning, for instance, you can quickly find the coffee mug on the counter, identify the coffee machine, and locate the fridge to get some milk.

1.1 Objects and Scenes

The visual environment can be distinguished into objects and scenes, both of which can be recognized in isolation and are processed in different regions of visual cortex (Park et al., 2011; Peelen and Kastner, 2014; R. A. Epstein and Baker, 2019). Objects are potentially movable entities defined by local properties that are placed in scenes. Scenes, on the other hand, consist of larger-scale and global surfaces or environments and are a view in which objects and surfaces are arranged in a meaningful way (Henderson and Hollingworth, 1999). It is not a single object that makes up a scene, but the context of many different objects together (Kaiser et al., 2019). For example, the coffee machine mentioned in the example above is a single object in the kitchen, however it does not make up the scene of a kitchen by itself, as one may as well put a coffee machine in an office. But seeing the coffee machine as a part of a set of objects, e.g., in combination with a fridge, a stove, an oven, a microwave, and a cutting board, leads to the perception of the scene "kitchen."

1.2 Neural Correlates of Object and Scene Perception

Rapid scene and object perception is supported by a network of brain regions, identified in humans using functional magnetic resonance imaging (fMRI). Scene-selective brain areas include the parahippocampal place area (PPA, R. Epstein and Kanwisher, 1998), the retrosplenial cortex (RSC, R. Epstein and Kanwisher, 1998; Aguirre et al., 1998) and the occipital place area (OPA, Dilks et al., 2013; Grill-Spector, 2003). Transcranial Magnetic Stimulation (TMS) of the OPA leads to impairments in scene categorization (e.g., a kitchen or garden, Ganaden et al., 2013), scene discrimination based on spatial layout (Dilks et al., 2013), and scene-related expectations (Gandolfo and Downing, 2019). Object-selective areas include the object-selective lateral occipital cortex (LOC, Grill-Spector, 2003; Malach et al., 1995) and the posterior fusiform gyrus (pFs, Grill-Spector and Weiner, 2014. Using TMS, Dilks et al., 2013 first showed a double dissociation between the scene-selective OPA and the object-selective LOC in isolated scene and object recognition tasks: The OPA is involved in scene recognition, but not object recognition; the LOC is involved in object recognition but not scene recognition. This finding was further supported by a preregistered replication study (Wischnewski and Peelen, 2021a). Together, these studies show that scene and object information are processed by separable networks of brain regions. Disruption of one node in these networks via TMS can selectively impair scene or object recognition.

1.3 Scene Perception

Different aspects of a scene have been proposed to be crucial for rapid scene recognition including both low-and high-level properties (R. A. Epstein and Baker, 2019). Low-level properties, such as spatial frequency or rectilinearity, have been shown to distinguish between scene and non-scene images within 100 ms after stimulus onset (Bastin et al., 2013) and influence event-related amplitudes even in later stages of scene processing (Groen et al., 2016; Groen et al., 2013; Harel et al., 2016). However, they are not sufficient to explain scene selectivity, as scene selectivity persists even when low-level properties are controlled for (R. A. Epstein and Baker, 2019). High-level properties have also been shown to modulate responses in scene-selective areas, with these areas being sensitive to spatial layout (Kaiser et al., 2019) and scene category (R. A. Epstein and Morgan, 2012), amongst others. More specifically, scene-evoked patterns are driven by the shape of the scene (Greene and Oliva, 2009) and the distance to the scene surfaces (Kravitz et al., 2011). The OPA has been proposed to process spatial structures of scenes (Dilks et al., 2011; Persichetti and Dilks, 2018), environmental boundaries (Julian et al., 2016) and contour junctions (Choo and Walther, 2016). Consequently, TMS-induced disruption of the OPA results in worse scene categorisation (Ganaden et al., 2013) and discrimination of scenes based on their layout (Dilks et al., 2013). On a different note, Oliva and Schyns, 1997) have shown the specific aspects of an image used for scene recognition to be subject to the task at hand. Broad cues, such as the colours of blobs (Oliva and Schyns, 2000) can already distinguish between broad categories (e.g. indoor vs outdoor scenes), whereas finer-grained details are required for distinguishing within categories (e.g. distinguishing Berlin from Amsterdam). Peripheral vision has been shown to be important for scene categorisation tasks (e.g. Thorpe et al., 2001; Tran et al., 2010), with global scene properties being recognized even with eccentricities as high as 70° (Boucart et al., 2013).

1.4 Object Perception

Object-selective areas are activated by a variety of different objects, such as faces (Kanwisher et al., 2002), animals (Chao et al., 1999), or tools (Macdonald and Culham, 2015), but not by pure textures, noise, or highly scrambled objects (Grill-Spector, 2003). Whereas some of these areas are selectively active for specific types of objects, such as the fusiform face area for faces (Kanwisher et al., 2002) or the extrastriate body part area for body parts (Downing et al., 2001), other areas, including the lateral occipital cortex, support more general object recognition (Grill-Spector, 2003). The LOC shows activation in response to objects defined by different low-level properties, such as luminance, texture, or motion (Grill-Spector et al., 1998), as well higher-level aspects such as line drawings (Ishai et al., 2000) and shapes defined by illusory contours (e.g. the Kanizsa triangle) (Mendola et al., 1999). Later, it was suggested that object-selective regions primarily respond to the shape of an object rather than its contours or other low-level features (Kourtzi and Kanwisher, 2001). The influence of higher-level image properties, such as object category, becomes particularly apparent in the body of literature assessing differences between animate and inanimate objects. For example, Proklova et al., 2016 showed differential activation in animate- and inanimate-preferring regions for visually similar objects that only differed in their animacy. Thorat et al., 2019 provided evidence for the animate-inanimate distinction to rely, amongst others, on high-level properties such as the agency of the object. Together, these studies show that object processing, just like scene processing, depends on both low- and high-level image properties.

1.5 Scene–Object Interactions

In our everyday life, objects and scenes are rarely separated from another and therefore object and scene perception are tightly linked. For example, object recognition is improved when an object is embedded in a congruent scene as compared to an incongruent scene (Boyce et al., 1989): a priest is better recognized when he is shown in a church, as opposed to seeing him in a football stadium (Davenport and Potter, 2004). This effect persists even if the scene is void of any structure and only retains its textures (Lauer et al., 2018). Further, objects are perceived to be sharper when placed in a congruent scene than when presented in an incongruent scene or in isolation (Lupyan, 2015; Lupyan, 2017; Rossel et al., 2022) and an ambiguous object is better recognized when presented in its scene as compared to being presented in isolation (Brandman and Peelen, 2017). Thus, there is a facilitatory effect of scenes on object recognition. Võ, 2021 highlighted the importance of "scene grammar" for context effects on object recognition. Like grammar used in language, scene grammar describes that naturalistic scenes and the object within follow certain rules and regularities, which are then used to form quicker predictions about the visual environment for easier and quicker recognition. These regularities are of both semantic nature, meaning the congruency between the object and the scene it is in, and syntactic nature, describing positional regularities of the object within a scene. A piece of cheese in a bathroom is a semantic violation, as objects of the category "cheese" do not belong to the scene category "bathroom." On the other hand, a toilet roll in the shower is a syntactic violation. While the object "toilet roll" does belong in the bathroom scene, it is usually not positioned in the shower. Both semantic and syntactic violations hinder scene-facilitated object recognition (Võ, 2021).

1.6 Theories of Scene-Object Interactions

Three different accounts have been proposed to explain the nature of the interactions between scene and object processing (Peelen et al., 2024): A "scene-first" account proposes that scenes are processed before objects and bias object recognition, respectively. This framework is supported by studies demonstrating the use of coarse information in early visual processing, followed by the use of finer details in later stages to refine the coarse input (Schyns and Oliva, 1994). Further, Hochstein and Ahissar, 2002 propose that the initial conscious percept of visual input is a categorical scene interpretation, that is then refined through feedback connections. Contrary to the scene-first account, the "object-first" account argues that the combination of object features into objects occurs before these objects are then combined to build the scene. In support of this framework, Liu et al., 2009 were able to show object selectivity in the human brain as early as 100 ms after stimulus onset and Crouzet et al., 2012 found that animal detection in an image precedes scene categorization. More recently, Peelen et al., 2024, proposed a "parallel processing" account supported by both behavioural and electrophysiological work: In rapid image recognition tasks, participants were equally likely to describe the semantics of objects and scenes (Fei-Fei et al., 2007), indicating no preference of object over scene recognition or the other way around. On a temporal scale, high-level representations of both objects and scenes peak at around 200 ms after stimulus onset (Carlson et al., 2013; Cichy et al., 2014; Kaiser et al., 2019), further supporting the parallel processing account.

1.7 Scene-Facilitated Object Recognition

If processed in parallel, how do object and scene representations facilitate each other? Henderson and Hollingworth, 1999 proposed that scene and object information is integrated at post-perceptual stages. By contrast, the contextual-facilitation model (Bar, 2004) proposed that contextual predictions are sent back to modulate activity in object-selective regions. Brandman and Peelen, 2017 investigated these theories emphasizing the facilitatory effect of scenes on object recognition. Participants not only performed better in categorizing an ambiguous object when the object was placed in its original scene, but it was also found that the recognition accuracy in this object-in-scene condition was even higher than the sum of the categorization accuracies of the object and the scene alone. This finding demonstrated that more information about the object can be extracted when it is placed in a scene than the combination of information that the object and scene provide separately. MEG recordings further showed that the scene-based facilitation of object processing peaks nearly 100 ms after the processing of intact objects, namely at 320 ms after stimulus onset. In line with the contextual-facilitation model, this finding suggests a feedback loop between scene- and object-selective areas.

Using brain stimulation, Wischnewski and Peelen, 2021b assessed the causal neural mechanisms behind this proposed feedback loop. In an object recognition task, participants were presented photographs containing a single object in one of two conditions: in the "isolated object" condition, the object was cropped out of the image and presented on a neutral background and in the "context-based" condition, the object was pixelated in its original scene. To avoid confusion with conditions in the current research project, the isolated object condition will henceforth be referred to as the "object-without-scene" condition, and the context-based condition will be referred to as the "object-with-scene" condition. Chronometric TMS was applied over the early visual cortex (EVC), LOC and OPA at an early (60-100 ms after stimulus onset), middle (160-200 ms after stimulus onset) or late time point (260-300 ms after stimulus onset). In line with feedforward processing of intact object features (Cichy et al., 2014), stimulation of the scene-selective OPA in the object-without-scene condition was not expected to have an effect on object classification accuracy. Stimulation of the object-selective LOC was hypothesized to decrease accuracy selectively at the middle time point, reflecting object processing, and EVC stimulation was expected to selectively decrease accuracy at the early time point. In the object-with-scene condition, stimulation of the EVC at the early time point and stimulation of the LOC at the middle time point were again expected to decrease classification accuracy, reflecting object processing. Stimulation of the OPA at a middle time point was expected

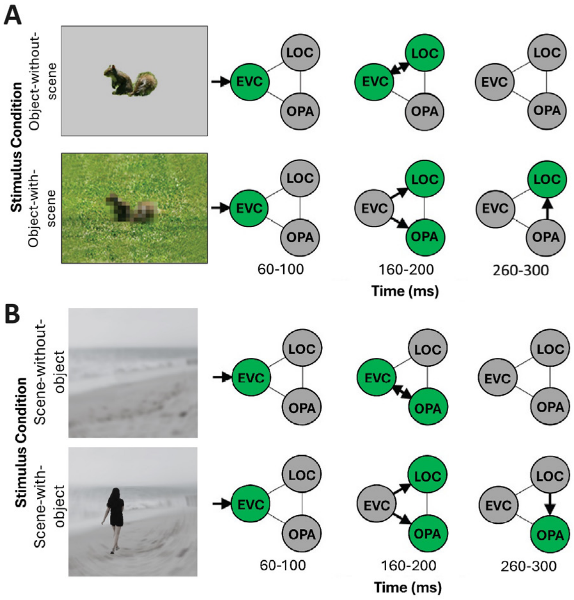


Figure 1: **Feedback loops between scene- and object-selective areas:** *Note.* (A) Schematic processing routes of isolated and context-based object recognition (adapted from Wischnewski and Peelen, 2021b). Isolated object recognition (top row), i.e. an object placed on a neutral background, is initially supported by EVC early (60-100 ms), followed by higher-level object processing in the LOC (160-200 ms) and feedback processing to the EVC (160-200 ms). Context-based object recognition (bottom row), i.e. an object placed in a scene, is initially supported by EVC (60-100 ms), followed by the LOC and OPA (160-200 ms), reflecting object and scene processing, respectively. Contextual disambiguation of the object then occurs in the LOC at 260-300 ms. (B) Proposed schematic processing routes of isolated and object-facilitated scene recognition. Isolated scene recognition (top row), i.e. a scene without an object, is initially supported by the EVC (60-100 ms), followed by higher-level scene processing in the OPA (160-200 ms) and feedback processing to the EVC (160-200 ms). Object-facilitated scene recognition (bottom row), i.e. a scene with an object in it, is initially supported by the EVC (60-100 ms), followed by the OPA and LOC (160-200 ms), reflecting object and scene processing, respectively. Object-based disambiguation of the scene then occurs at 260-300 ms in the OPA.

to decrease accuracy, reflecting scene processing. Crucially, stimulation of the LOC at the late time point was also expected to decrease context-based object recognition, as scene-based expectations of object category were hypothesized to reach the LOC at this late time point. Indeed, it was found that, in the object-without-scene condition, the EVC supports object recognition at the early time point and the LOC supports object recognition at the middle time point. In the object-with-scene condition, EVC was found to support context-based object recognition at an early time point and the LOC was found to support it at the middle time point. The OPA selectively supported context-based object recognition at the middle time point, and importantly, the LOC was also found to be involved in context-based object recognition at the late time point. These results propose a feedback loop between object-selective LOC and scene-selective OPA looking as follows. Shortly after stimulus onset (60-100 ms), the visual input is represented in the early visual cortex. This information is then sent to the scene- and object-selective areas, which process their respective input (160-200 ms). Object-selective areas recognize the object, scene-selective areas recognize the scene. After the scene-selective areas have disambiguated their input, they send their output to the object-selective areas (260-300 ms), where that information is then used to disambiguate the object (see Figure 1A).

1.8 Object-Facilitated Scene Recognition

While contextual object recognition has been widely explored (e.g. Biederman et al., 1982; Brandman and Peelen, 2017; Brandman and Peelen, 2019; Rossel et al., 2022), the opposite interaction between scenes and object has received less attention (Peelen et al., 2024). Using behavioral measures, previous studies demonstrated a facilitatory effect of congruent objects on scene recognition (Davenport and Potter, 2004; Davenport, 2007), even when objects and scenes were presented separately (Leroy et al., 2020). Another body of literature demonstrated an effect of objects on scene recognition in which incongruent objects decreased scene classification accuracy (Joubert et al., 2007; Furtak et al., 2022). However, to the best of our knowledge, no clear behavioral evidence that objects can disambiguate an otherwise ambiguous scene exists to date. Using fMRI, Brandman and Peelen, 2017 examined an interaction by which objects facilitate the neural representation of scene layout. In a scene-categorization task, participants were presented with degraded scenes that were difficult to recognize on their own, but easily disambiguated when an object was included. It was found that, in the left hemisphere, objects also help scene disambiguation in the scene-selective regions. More specifically, activation in the left OPA was higher during object-facilitated scene recognition than activation in the right OPA. This finding was consistent with another study that provided neural evidence that scene selective regions process scenes in conjunction with their embedded objects (Aminoff and Durham, 2023). Later, it was shown that this facilitatory effect of objects on scene categorization showed an identical temporal profile as scene-facilitated object recognition starting from 300 ms after stimulus onset (Brandman and Peelen, 2023). This gives rise to the assumption that the feedback loop between scene- and object-selective areas for scene-facilitated object recognition found by Wischnewski and Peelen, 2021b also works the other way around for object-facilitated scene recognition, further extending the contextual-facilitation model. That is, after object-selective areas recognize the object at around 160-200 ms, that information is sent to the scene-selective areas, where it helps disambiguating the scene at 300 ms (see Figure 1B). Causal neural evidence arguing for the involvement of both object- and scene-selective regions in object-facilitated scene recognition is required to show the existence of a feedback loop between object- and scene-selective areas in object-facilitated scene recognition, but it is still lacking to date. Therefore, the current research project aims to explore the causal neural mechanisms behind object-facilitated scene recognition and to provide a novel behavioural marker for this effect. Further, we investigate which aspects of objects and scenes are of importance for the object-facilitatory effect to take place.

1.9 Current Research Project

In a behavioral study (Experiment 1), we first establish that the scene category (i.e., whether a scene is indoors or outdoors) of an ambiguous scene can be inferred from the presence of an object. In

Experiment 2, we assess the importance of scene structure for this object-facilitatory effect to take place. In a third behavioural experiment (Experiment 3), we investigate the importance of object position for object-facilitated scene recognition. Finally, we assess the causal involvement of the LOC and OPA during object-facilitated scene recognition using TMS (Experiment 4).
Together, these experiments will provide both behavioural and neural evidence for object-facilitated scene recognition. To the best of my knowledge, the current research is the first research project to assess the causal neural mechanisms of object-facilitated scene recognition and provide a novel behavioural marker for it.

2 Experiment 1

In Experiment 1, we assessed whether objects, whose object category cannot fully explain scene category can facilitate the categorisation of ambiguous scenes. We expect scene classification accuracy to be higher when the object was placed in the ambiguous scene than when the object was shown on a neutral background or when the ambiguous scene was shown without the object.

2.1 Methods

2.1.1 Participants

Sixty-one participants participated in the experiment (33 females, 27 males; mean age 36.27 years; SD = 13.24) and were recruited via Prolific (Prolific Inc.). Eleven participants were excluded during the analysis due to performance at chance level, so that the final sample size consisted of 50 participants (27 females, 23 males; mean age 35.17 years; SD = 12.64). The sample size was chosen due to the exploratory nature of this study and the corresponding power considerations, A sensitivity analysis showed that this sample is sufficient to detect a small to medium effect size (partial $\eta 2 = 0.03$) with a power of $\beta = .800$. Participants gave informed consent before the study and received financial compensation for their participation (£4). This study was approved by the Radboud University Ethics Committee (ECSW – 2022-079).

2.1.2 Stimuli

The stimulus set consisted of degraded scenes that were considered ambiguous on their own but were more easily recognizable when an object was placed in the scene. We ensured that the scene did not contain objects other than the specified object. The main experiment included 48 photographs of outdoor scenes (e.g. beaches, gardens, forests) and 48 photographs of indoor scenes (e.g. kitchens, living rooms, malls). The photographs were cropped and cleaned to include one dominant foreground object. After temporarily cropping out the object, the scenes were desaturated and ambiguated. Each image was saved in three conditions (see Figure 2A): the degraded scene including the object ("scene-with-object" condition), the degraded scene without the object ("isolated scene" condition), and the

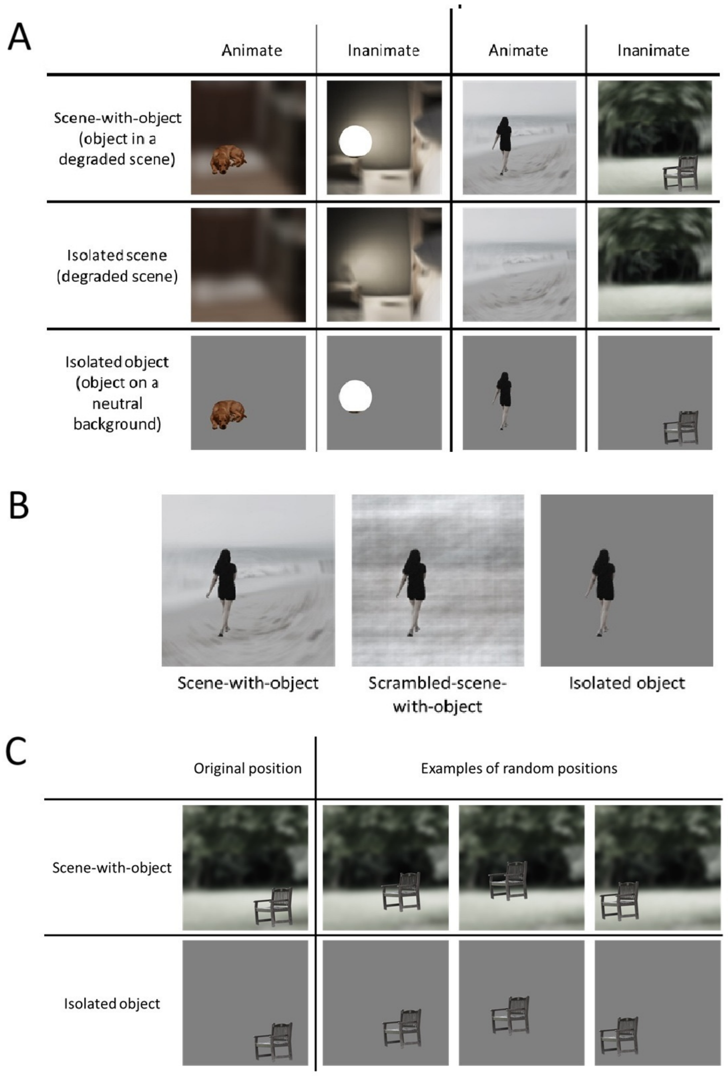


Figure 2: **Sample stimuli of Experiments 1-3:** *Note.* (A) Sample stimuli for Experiment 1. Stimuli could show indoor or outdoor scenes. The objects within the scenes were either animate or inanimate and belonged to one of six categories (cat, dog, human, chair, lamp, plant). Each stimulus was created in three conditions: A degraded scene including the non-degraded object; a degraded scene without the object; the non-degraded object on a neutral background. All stimuli were desaturated to reduce colour cues. (B) Sample stimuli for Experiment 2. Object- and scene-categories were identical to Exp 1, but the isolated scene condition was replaced with a condition depicting the non-degraded object in the phase-scrambled scene. (C) Sample stimuli for Experiment 3. The same stimuli as in Experiment 1 were used with the following changes. The object in each stimulus was either shown at its original position in the scene, or at the position that another object has in its respective scene ("random position").

object on a neutral gray background ("isolated-object" condition). Thus, the final stimulus set contained 288 images across all conditions.

2.1.3 Procedure

On each trial, participants were presented with a fixation cross (500 ms), followed by the stimulus presentation (50 ms), a blank screen (500 ms), and a two-second two-alternative forced choice response

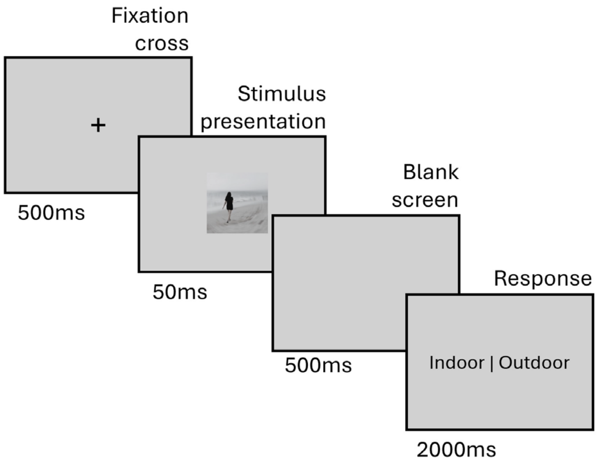


Figure 3: **Trial Procedure of Experiment 1-3:** *Note.* Procedure of Experiments 1-3. After a fixation cross (500 ms), the stimulus was presented for 50ms. Then a blank screen appeared (500 ms), followed by a 2000 ms response window (the indoor/outdoor labels are presented for illustrative purposes, but were not presented in the actual experiment). The participant responded using the "f" or "j" key on the keyboard. The next trial began after a response was recorded or the 2000 ms passed.

window (Figure 3). Before the recorded trials, participants were given instructions and completed a practice block of 12 trials (4 trials for each of the three conditions) using stimuli that were not used in the main experimental block. During the practice, participants received feedback after each trial ("correct" or "incorrect"). Participants did not receive feedback during the experiment, but every 24 trials a break screen presenting the average response time and accuracy of their last 24 trials appeared. For all participants, the stimuli were fit to a size of 10x10 cm, irrespectively of the monitor's resolution.

2.1.4 Design

The dependent variables of this within-subjects design were accuracy, reaction time and LISAS (Linear Integrated Speed-Accuracy Score; Vandierendonck, 2017; Vandierendonck, 2018) and the independent variable was object presence. Object presence included three conditions: scene-with-object, isolated scene, and isolated object. Each participant completed two blocks of 144 trials each for a total of 288 trials. Stimuli were presented in random order within each block. To avoid familiarity and carry over effects within images, the stimuli were split such that of any given image, participants were presented with either the isolated object and isolated scene condition, or with the scene-with-object condition. Therefore, one stimulus set contained 48 stimuli in the scene-with-object condition and the remaining 48 stimuli in both the isolated object and isolated scene condition. These two sets of 144 stimuli each were counterbalanced across participants.

2.1.5 Analyses

Three separate repeated-measures ANOVAs using a Greenhouse-Geisser sphericity correction were conducted to assess the effect of stimulus condition on accuracy, reaction time and LISAS. Recognition accuracy is defined as the percentage of correct answers given. In the reaction time analysis, only reaction times of correct trials were considered. The linear integrated speed-accuracy score is defined as:

$$LISAS = RT_C + PE \times \frac{S_{RT}}{S_{PE}}$$

Where RT_C is the average reaction time in correct trials, PE is the proportion of errors (defined as $PE = 1 - \frac{n_{\text{correct trials}}}{n_{\text{all trials}}}$), SRT refers to the standard deviation of the correct RTs and SPE refers to the standard deviation of the proportion of errors (Vandierendonck, 2017; Vandierendonck, 2018). Lower LISAS indicates better performance. We excluded participants performing at chance level from the analyses via a binomial test. If a one-sided binomial test comparing their accuracy with 50% was insignificant (at α=.05), that participant performed at chance level and was subsequently removed from the analyses.

2.2 Results

There was a main effect of object presence (F(1.857, 90.997) = 54.008, p < .001, partial η2 = .524), indicating a large effect (Figure 4). Post-hoc pairwise comparisons with a Bonferroni correction revealed that there were significant differences in accuracy between all conditions. When the object was presented in the ambiguous scene (scene-with-object condition), participants achieved better scene classification performance (μ = 0.718, SD = 0.084) than in the isolated-scene condition (μ = 0.644, SD = 0.082, t = 6.818, p_{adj} < .001). The effect size, calculated using Cohen's d, indicated a large effect (d = 0.95). Accuracy in the scene-with-object condition was also higher than accuracy in the isolated object condition (μ = 0.605, SD = 0.060, t = 11.616, p_{adj} < .001) and the effect size indicated a large effect (d = 1.45). In addition, accuracy in the isolated-scene condition was higher than in the isolated object condition (t = 3.130, p_{adj} = .009) with medium effect size (d = 0.50).

3 Experiment 2

In Experiment 2, we investigated whether the facilitatory effect of the object is based on the object disambiguating the scene layout, or whether object and scene information contribute independently to this effect. To test this, the scene was phase-scrambled, and performance was compared between the object placed in a scrambled scene, the object in the ambiguous scene and the object on a neutral background. Phase-scrambling preserves some low-level properties, amongst which colours and luminosity (Oliva and Schyns, 1997; Oliva and Schyns, 2000), and the overall distribution of spatial frequencies, but, importantly, it disrupts the spatial arrangement of these frequencies. This makes the scene structure unrecognizable to the participant. As no scene structure remains after phase-scrambling, we expect to find no facilitatory effect of object presence if the effect is based on the object disambiguating the scene layout. However, if object and scene contribute independently to this effect, we expect the facilitatory effect to persist in the scrambled scene by combining the retained low-level properties of the scene with the object information.

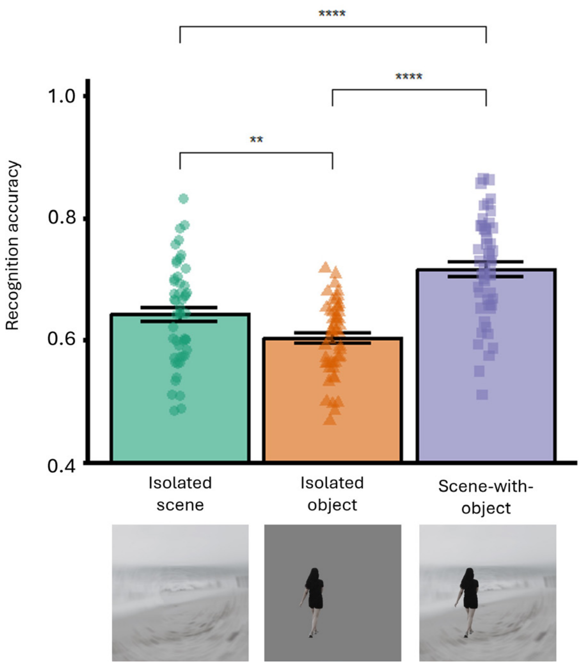


Figure 4: **Scene Recognition:** *Note.* Data are presented as mean recognition accuracy, error bars indicate SEM, points represent individual participants. Scene recognition accuracy was highest in the scene-with-object condition and lowest in the isolated object condition. $**\ p \leq .01$; $****\ p \leq .0001$

3.1 Methods

3.1.1 Participants

Fifty-eight participants participated in the experiment (32 females, 26 males; mean age 36.45 years; SD = 13.76). After removing participants performing at chance, the final sample consisted of 51 participants (29 females, 22 males; mean age 37.10 years; SD = 14.32). Since we had no prior expectations about the effect size, we aimed for a final sample size of at least n = 50. A sensitivity analysis showed that this sample size is sufficient to detect a small to medium effect size (partial $\eta 2 = .03$) with a power of $\beta = .80$ for a one-way repeated measures analysis of variance with three factors. Participants gave informed consent before the study and received financial compensation for their participation (4£). This study was approved by the Radboud University Ethics Committee (ECSW – 2022-079).

3.1.2 Stimuli

A subset of 66 stimuli of the stimuli used in Experiment 1 were used in this experiment with the following changes. The isolated scene condition was replaced with a phase-scrambled scene containing the object (“scrambled-scene-with-object” condition). In this condition, the object was temporarily cropped out and the scene was then degraded using phase scrambling (Thomson, 1999). Contrary to the scene-with-object condition, the phase-scrambled scene was not desaturated and retained low-level properties of the scene. The final images (198 in total) included a degraded scene with an object, the object on a neutral background and the phase-scrambled scene with an object (see Figure 2B) and had an equal number of indoor and outdoor scenes.

3.1.3 Procedure

The same procedure as in Experiment 1 was followed.

3.1.4 Design

The same design was used as in Experiment 1, with the following differences. A subset of 66 scenes of Experiment 1 were chosen. Instead of the isolated scene condition, participants were presented with the scrambled-scene-with-object condition. To avoid familiarity and carry-over effects within images, the stimuli were split such that of any given image, participants were presented with either the scene-with-object condition, the scrambled-scene-with-object condition, or the isolated object condition, but not with multiple conditions of the same image. This resulted in three sets of stimuli, each consisting of 22 stimuli in the scene-with-object, scrambled-scene-with-object and isolated object condition, and thus 66 stimuli in total. Participants completed four blocks of 66 trials each for a total of 264 trials.

3.1.5 Analyses

The same analyses as in Experiment 1 were conducted.

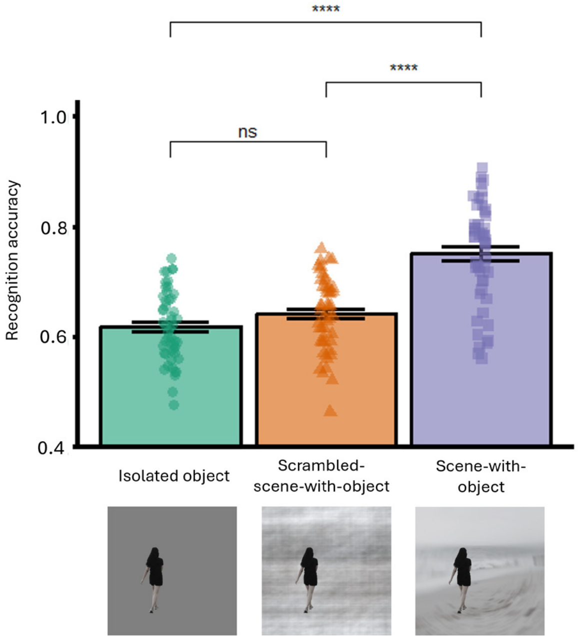


Figure 5: **Scene Recognition:** *Note.*Data are presented as mean recognition accuracy, error bars indicate SEM, points represent individual participants. Scene recognition accuracy was highest in the scene-with-object condition and lowest in the isolated object condition. $****\ p \leq .0001$

3.2 Results

Accuracy was statistically significantly different between the different image conditions (F(1.996, 99.997) = 81.277, p < .001, partial η2 = .619), indicating a large effect. Post-hoc pairwise comparisons with a Bonferroni correction revealed that there were significant differences in accuracy between the scene-with-object (μ = 0.752, SD = 0.091) and scrambled-scene-with-object condition (μ = 0.643, SD = 0.066, t = 9.895, p_{adj} > .001). The effect size was large, with a Cohen's d of 1.38. Accuracy in the scene-with-object condition was higher than in the isolated object condition (μ = 0.619, SD = 0.064, t = 11.679, p_{adj} > .001) with large effect size (d = 1.67). No differences between the isolated object and scrambled-scene-with-object condition were found (t = 2.141, p_{adj} = .112, d = 0.30) (Figure 5), indicating that removing scene structure through phase scrambling inhibits object-facilitated scene recognition as much as replacing the scene with a neutral background.

4 Experiment 3

In Experiment 3, we investigated whether the object-facilitatory effect is based on the object facilitating the construction of the scene layout. The position of each object was changed to the position of a randomly chosen other object (e.g., a dog in an outdoor scene was placed at the position of a chair in an indoor scene). This allows us to explore whether the effect of object-facilitated scene recognition is primarily driven by a semantic effect (e.g., a dog is in that scene so it must be outdoor) or whether the objects original position in the scene contributes to its facilitatory effect. This can be framed using the scene grammar framework. An object-based facilitation on scene judgments could be explained by a semantic (e.g., a dog is usually outdoor) and/or a syntactic relationship between the scene and the object (e.g., a dog in a given spatial position and orientation supports the disambiguation based on the position in which it appears in the scene rather than solely due to its object category). In this experiment, we disrupted the syntactic relationship between the scene and the object by randomly swapping the position of the objects on the scene, while keeping the semantic relationship intact (presenting the very same object that belongs to the photograph). If the object-to-scene syntactic relationship contributes to object-based facilitation effects, the facilitation should be removed or reduced when the object originally belonging to that scene is at a position different from its original location. On the other hand, if the relationship is purely semantic, changing the position of the object should not change the presence or magnitude of the object-facilitatory effect.

4.1 Methods

4.1.1 Participants

104 participants participated in the experiment (32 females, 70 males, 1 identified as other, 1 preferred not to say; mean age 36.74 years; SD = 11.27) and were recruited via Prolific. Four participants were excluded during the analysis due to performance at chance level so that the final sample consisted of

100 participants (31 females, 67 males, 1 identified as other, 1 preferred not to say; mean age 36.89 years; SD = 11.25). We expected an effect of changing the position of an object on its facilitatory effect on scene recognition to be small and therefore aimed for a sample size of n = 100. A sensitivity analysis on a paired two-tail contrast revealed that a sample size of n = 100 is sufficient to detect an effect of d = .25 with a power of β = .80. Participants gave informed consent before the study and received financial compensation for their participation (£4). This study was approved by the Radboud University Ethics Committee (ECSW – 2022-079).

4.1.2 Stimuli

A subset of 64 stimuli from Experiment 1 were used. In addition to the existing conditions, variations of the stimuli in the scene-with-object and isolated object condition were created in which the object was placed at the position of any other object in its original scene ("random position," see Fig. 2C).

4.1.3 Procedure

The procedure was identical to the procedure of Experiment 1.

4.1.4 Design

This experiment used a 3x2 within-subjects design. The independent variables were object presence (scene-with-object condition, isolated object condition, isolated scene condition) and object position (original position, random position). The dependent variables were accuracy, reaction time and LISAS. Each participant completed two blocks of 192 trials each for a total of 384 trials. Within a block, each stimulus was repeated twice. The scene-with-object and isolated object condition were shown once with the object at its original position and once with the object at a random position. To avoid familiarity and carry-over effects within images, participants were presented with either the isolated object and isolated scene condition, or with the scene-with-object condition. Therefore, one stimulus set contained 32 stimuli in the scene-with-object condition with the object at the original position, the same 32 stimuli in the scene-with-object with the object at a random position, and the remaining 32 stimuli once in the isolated object condition with the object at its original position and once at a random position and twice in the isolated scene condition. These two sets of 192 stimuli each were counterbalanced across participants. To omit order effects, the order of stimuli within each block was randomized for each participant.

4.1.5 Analyses

A 3x2 repeated measures analysis of variance was conducted, with accuracy as the dependent variable and stimulus condition and object position as the independent variables. Despite no object being present in the isolated scene condition, it was dummy-coded such that stimuli in the isolated scene condition were shown once in the random object position condition and once in the original object position condition. Since stimuli in the isolated scene condition were coded such that they could have the object at the original or random position, despite no object being present in the stimuli, the isolated scene condition was removed for three subsequent 2x2 repeated measures analyses of variance, predicting accuracy, reaction time and LISAS from stimulus condition and object position. Finally, a repeated measures ANOVA was conducted, predicting the object-facilitatory effect on accuracy from object position. The object facilitatory effect was defined as the difference in accuracy between the scene-with-object and the isolated object condition. Participants performing at chance level were removed from the analyses via a binomial test. If a one-sided binomial test comparing their accuracy with 50% was insignificant (at α = .05), that participant performed at chance level and was subsequently removed from the analyses.

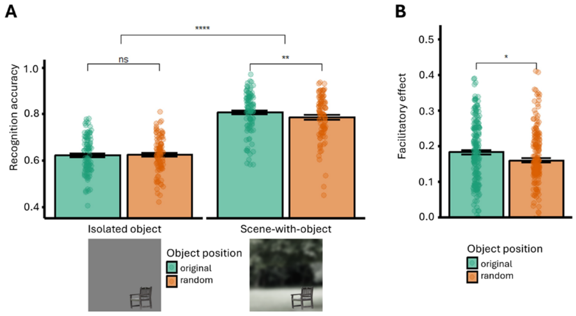


Figure 6: **Results of Experiment 3:** *Note.* . Data are presented as mean recognition accuracy, error bars indicate SEM, points represent individual participants. ns = not significant; $*p\leq.05$; $**p\leq.01$; $****p\leq.0001$ (A) Scene recognition accuracy overall was higher in the scene-with-object condition than in the isolated object condition. In the scene-with-object condition, accuracy was higher when the object was at its original position. (B) The facilitatory effect that the object has on scene classification accuracy, computed as the difference in accuracy between the isolated object and object-with-scene condition, was significantly higher when the object was at its original position than when it was at a random position.

4.2 Results

A significant interaction effect between stimulus condition and object position on accuracy (F(1,99) = 6.097, p = .029, partial $\eta 2$ = 0.036) was found. This interaction was also significant when removing the isolated scene condition (F(1,99) = 6.097, p = .015, partial $\eta 2$ = 0.058). Therefore, the effect of object position was analysed at each stimulus condition. P-values were adjusted using the Bonferroni multiple testing correction method. In the scene-with-object condition, accuracy was higher with the object at the original position (μ = 0.808, SD = 0.088) than at a random position (μ = 0.787, SD = 0.094, t = 2.798, p = .006). The effect size, as measured by Cohen's d, was d = .25, indicating a small effect. No differences based on object position were found in the isolated object condition ($_{\text{original}}$ = 0.624, SD_{original} = 0.080, $_{\text{random}}$ = 0.626, SD_{random} = 0.078, t = 0.401, p = .689, d = .02) and the isolated scene condition ($_{\text{original}}$ = 0.721, SD_{original} = 0.105, $_{\text{random}}$ = 0.718, SD_{random} = 0.103, t = 0.431, p = .667, d = .03) (Figure 6A). This means that accuracy in the scene-with-object position was higher when the object was at its original position, but accuracy in the isolated object condition did not differ depending on object position. Further, the object facilitatory effect that an object has when placed in the scene is stronger when the object is at the position it originally had in the photograph (μ= 0.184, SD= 0.089) than when it is placed at the position of another object in its scene (μ= 0.161, SD= 0.090, p = .015). The effect size, calculated using Cohen's d, indicates a small effect (d = 0.25) (Figure 6B).

5 Experiment 4

Finally, to assess the causal neural mechanisms underlying object-facilitated scene recognition, two pre-registered TMS studies on the same participants were conducted: a four-pulse TMS study (https://aspredicted.org/W76_DS2) and a chronometric TMS study (https://aspredicted.org/XRJ_5QW). The four-pulse TMS study was used for "functional localization" and will be referred to as "selection procedure". The selection procedure will not be discussed in this thesis.

In the chronometric TMS study, participants received TMS at three time windows (early: 60-100 ms; middle: 160-200 ms; or late: 260-300 ms after stimulus onset) over either LOC or OPA while performing a scene-recognition task. In this scene classification task, participants viewed either a very degraded scene with an object in it ("scene-with-object" condition) or an only slightly degraded scene without an object present ("scene-without-object" condition; see Figure 7). This design allows us to systematically investigate the interaction between scene- and object-selective regions over time in support of object-facilitated scene recognition. We expect LOC stimulation to result in worse scene recognition in the scene-with-object but not in the scene-without-object condition. Because object recognition peaks at 200 ms, we hypothesize LOC stimulation to lead to worse performance in the scene-with-object condition with stimulation at the middle time point specifically, but not at the early or late time point. On the other hand, we expect OPA stimulation to decrease accuracy more strongly in the scene-without-object condition than in the scene-with-object condition. In the scene-without-object condition, we hypothesize OPA stimulation to result in a stronger decrease in accuracy at the middle stimulation time point than with early or late stimulation onset. In the scene-with-object condition, we hypothesize OPA stimulation at the middle time point to result in worse performance than stimulation at the early time point, reflecting the processing of the ambiguous scene cues. Since we assume the object-facilitated scene recognition (i.e., the integration of the information about the scene with information about the object sent through feedback connections to the OPA by the LOC) to occur at around 300 ms after stimulus onset, we expect OPA stimulation in the scene-with-object condition at a late time point to result in even worse performance than stimulation at the middle time point (see Figure 8). For an overview of the two studies, see Figure 9.

5.1 Methods

5.1.1 Participants

A preregistered sample size of 48 healthy participants (31 females, 17 males; mean age 22.96 years; SD = 3.54 years) participated in both TMS studies. All participants had normal or corrected-to-normal

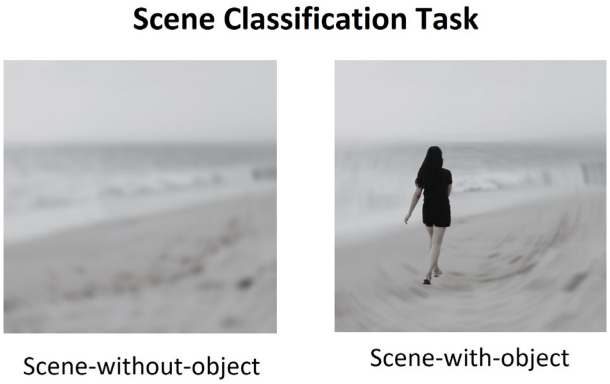


Figure 7: **Sample Stimuli of Experiment 4:** *Note.* Stimuli could either depict a desaturated and degraded scene without an object (“scene-without-object” condition) or a desaturated and degraded scene containing an object (”scene-with-object” condition). Stimulus conditions were matched on accuracy.

vision, were right-handed and gave informed consent. This sample size finds a medium interaction effect size (d = 0.5) with a power of $\beta = 0.9$. Participants with neurological or psychiatric disorders, CNS-acting medication, a (family) history of epilepsy or convulsions or seizures, brain surgery, cochlear or metallic implants in their head or neck area, cardiac pacemaker or intra-cardiac lines, a medication infusion device, a neuro-stimulator, or pregnancy were excluded from participation. Three participants withdrew from the experiment due to discomfort of the TMS stimulation during the selection procedure and one participant withdrew after completing the selection procedure. These four participants were subsequently removed from the analysis. One participant participated in the chronometric TMS experiment without prior participation in the selection procedure. Four outliers were removed from the analysis and replaced to match the preregistered number of participants. In the chronometric TMS experiment, 24 participants (19 females, 5 males; mean age 22.13 years; SD = 2.92 years) received OPA stimulation, and 24 participants received LOC stimulation (12 females; 12 males; mean = 23.92 years; SD = 3.88 years). The chronometric TMS experiment took place three to seven days after the selection procedure. This study was approved by the Radboud Ethics Committee (NL72752.091.20) and all participants involved in the study gave their informed consent before participating in both studies. Participants were monetarily compensated (15 euros per hour) or received course credits for their participation.

5.1.2 Stimuli

The stimulus set consisted of a subset of 64 indoor and outdoor scenes used in Experiment 1. In one condition, the object was presented in the ambiguous scene (“scene-with-object” condition, same as in Experiments 1-3), whereas the other condition showed a very slightly degraded scene (“scene-without-object” condition; see Figure 7). The stimulus conditions were validated and optimized in an online behavioural pilot experiment, using the same procedure as the TMS studies to ensure that the two conditions for each image did not differ in categorisation accuracy (Repeated Measures

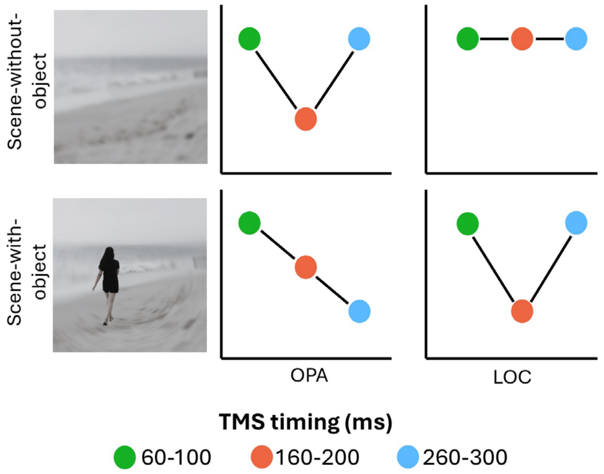


Figure 8: **Time Specific Hypothesis:** *Note.* This figure shows the time-specific predictions for the chronometric TMS experiment. We hypothesized that scene recognition without an object in the scene (top row) would be causally supported by the scene-selective OPA 160-200 ms after stimulus onset (middle time point). The object-selective LOC was not hypothesized to be involved in this process. We hypothesized scene recognition with an object in the scene (bottom row) to be causally supported by the OPA at 160-200 ms (middle time point) and crucially, the LOC also was also hypothesized to causally support object-facilitated scene recognition at that middle time point. Finally, object-facilitated scene recognition was hypothesized to occur in the OPA at 260-300 ms after stimulus onset (late time point).

ANOVA: n = 29; $_{\text{scene-with-object}}$ = 0.773; $_{\text{scene-without-object}}$ = 0.796; p = .109). This is an important prerequisite for the stimulation study to eliminate a possible main effect of stimulus condition. These two conditions allowed us to distinguish between object-facilitated scene recognition, involving both scene- and object-selective regions, and isolated scene recognition, involving just scene-selective regions. Stimuli were displayed on a BenQ Mobiuz 27” 120Hz computer screen with a fixed size of 10x10 cm.

5.1.3 Transcranial Magnetic Stimulation

TMS was applied via a Cool-B65 figure-of-8 coil with an outer diameter of 75 mm, which received input from a Magpro-X-100 magnetic stimulator (MagVentrue, Farum, Denmark). In two participants, a MC-B70 butterfly-shaped figure-of-8 coil with an outer diameter of 96 mm was used instead. Participants received double-pulse TMS at a stimulation intensity adjusted to 85% of the individual phosphene threshold of each participant. The double-pulse was given at 25 Hz for a total stimulation duration of 80 ms. The phosphene threshold was determined by increasing stimulation intensity during single-pulse stimulation of the early visual cortex until visual phosphenes were reported in 50% of the trials while participants fixated on a grey screen in a dimly lit room. The TMS coil was placed with the help of an infrared-based neuronavigation system (Localite, Bonn, Germany) using an individually adapted standard brain model over the left LOC and left OPA. In four participants, the TMS coil was placed with the help of the infrared-based neuronavigation system Visor2 (Antneuro, Hengelo,

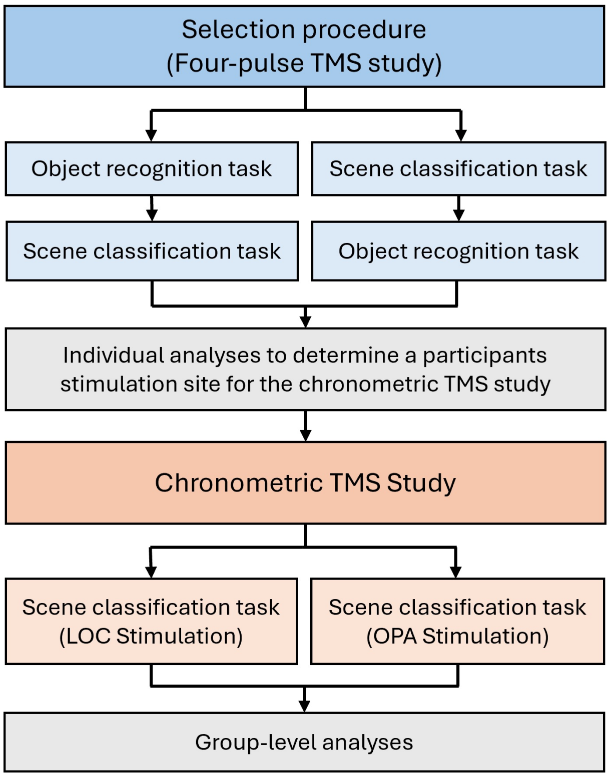


Figure 9: **Schematic overview of the two TMS studies:** *Note.* In the selection procedure, half of the participants first completed an object recognition task and then a scene classification task, and half of the participants first completed a scene classification task and then an object recognition task. Individual analyses were conducted to assign participants to a stimulation site condition in the chronometric TMS study. Participants showing reduced accuracy during LOC stimulation compared to OPA and vertex stimulation in the object recognition task were assigned to receive LOC stimulation and participants showing reduced accuracy during OPA stimulation when compared to LOC and vertex stimulation in the scene classification task were assigned to receive OPA stimulation in the chronometric TMS study. In the chronometric TMS study, taking place three to seven days after the selection procedure, participants performed the same scene classification task as in the selection procedure. Finally, group level analyses were conducted.

The Netherlands). The stimulation coordinates for LOC and OPA were identified through Talairach coordinates set in the Localite or Visor2 neuronavigation system. The coordinates used for the left LOC were -45, -74, 0 and for the left OPA, the coordinates were -34, -77, 0. Note that the coordinates used were the left-hemispheric homologue coordinates of the LOC (Pitcher et al., 2009) and OPA (Julian et al., 2016) defined in the right hemisphere. These right-hemispheric coordinates were also used by Wischnewski and Peelen (2021a; 2021b). The stimulation of the left hemisphere is based on the results of Brandman and Peelen (2019), who found higher activation in object-facilitated scene recognition in the left OPA as compared to the right OPA. The vertex was individually determined at half of the distance between inion and nasion.

5.1.4 Procedure

The chronometric TMS study took place three to seven days after the selection procedure and participants performed a scene classification task. Before beginning the task, participants gave informed consent, received instructions and the individual phosphene threshold was determined. Each trial of the scene-classification task started with a fixation point at the centre of the screen (500 ms), followed by the stimulus presentation (50 ms) and a blank screen for 500 ms. Then, participants had a response period of 2 s and after these two seconds had passed or a response was recorded, a variable inter-trial interval of 2-5 seconds started (see Figure 10). This relatively long interval was chosen to prevent the coil from overheating during the trials and to avoid TMS effects building up over trials and is commonly adopted in online TMS experiments (Gandolfo and Downing, 2019; Pitcher et al., 2008;2007; Urgesi et al., 2004; Gandolfo et al., 2024). Participants responded using F for indoors and J for outdoors. During the task, participants received double-pulse TMS at 2 5Hz (i.e., 80 ms total duration of stimulation) at an individually adjusted intensity set to 85% of the phosphene threshold. TMS was applied over either LOC or OPA at one of three different time points: at 60 ms after stimulus onset, 160 ms after stimulus onset, or at 260 ms after stimulus onset.

5.1.5 Design

The chronometric TMS study used the scene classification task in a 2x2x3 mixed design. The dependent variables were accuracy, reaction time and LISAS. The between-groups factor was stimulation site (LOC, OPA) and the within-groups factors included stimulus condition (scene-with-object, scene-without-object) and stimulation onset (early, middle, late). Depending on the selection procedure, participants received either LOC or OPA stimulation during the experiment. The stimulation strength was set to 85% of the individually determined phosphene threshold. Two TMS pulses could be applied at 60 ms and 100 ms after stimulus onset (early), 160 ms and 200 ms after stimulus onset (middle), or at 260 ms and 300 ms after stimulus onset (late). The scene classification task was changed such that every stimulus was repeated twice for each stimulation onset, resulting in 384 trials. The trials were presented in random order. The task was divided into 12 blocks of 32 trials, each lasting approximately 3 minutes, with short breaks in between of approximately 1 minutes. Thus, completing the task took about 60 minutes. The total duration of the experiment, including preparation and phosphene threshold determination, was approximately 90 minutes.

5.1.6 Analyses

Three separate 2x2x3 repeated measures ANOVAs were conducted, on the three dependent variables of interest —accuracy, reaction times in accurate trials and LISAS with stimulus condition (scene-with-object, scene-without-object), stimulation site (LOC, OPA), and stimulation onset (early, middle, late) as factors. Outliers were removed prior to the analyses with a pre-registered criterion of being 2.5 standard deviations slower or less accurate than the overall mean. In accordance with our hypotheses, the following pre-registered planned pairwise t-tests were computed. For OPA stimula-

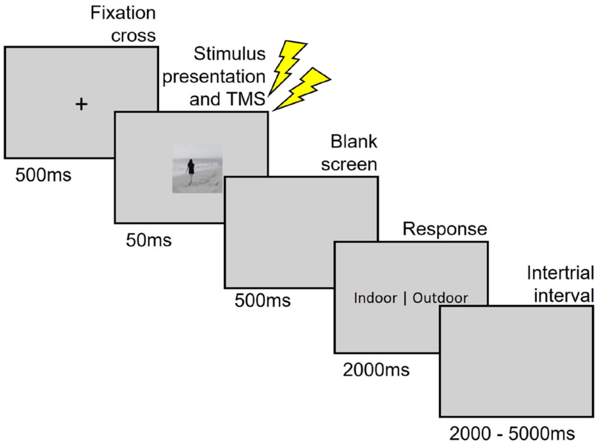


Figure 10: **Trial Procedure of Experiment 4:** *Note.* Schematic trial of the scene classification task. After stimulus onset, participants received double pulse TMS over LOC or OPA at one of three time points (60 ms after stimulus onset, 160 ms after stimulus onset, 260 ms after stimulus onset).

tion, performance in the scene-without-object condition is compared between early TMS onset and middle TMS onset and between middle TMS onset and late TMS onset. In the scene-with-object condition, performance between early TMS onset and late TMS onset is compared. For LOC stimulation, performance in the scene-with-object condition is compared between early TMS onset and middle TMS onset and between middle TMS onset and late TMS onset. Performance in the scene-without-object condition is compared between all stimulation onsets.

5.2 Results

There was no significant three-way interaction between stimulus condition, stimulation site and stimulation onset on accuracy (F(2,92) = 0.780, p = .780, partial $\eta 2$ = .017). Importantly, there was no main effect of stimulus condition (F(1,46) = 1.417, p = .240, partial $\eta 2$ = .030), indicating that the stimulus conditions were well equated in terms of difficulty. A significant two-way interaction between stimulation site and stimulus condition was found (F(1,46) = 7.471, p = .009, partial $\eta 2$ = .140) and a simple interaction analysis showed that, compared to the scene-without-object condition, participants were less accurate in the scene-with-object condition during LOC stimulation ($_{\text{scene-without-object}} = 0.918$, $SD_{\text{scene-without-object}} = 0.041$, $_{\text{scene-with-object}} = 0.889$, $SD_{\text{scene-with-object}} = 0.065$, t = -2.774, p = .008). The effect size, as measured by Cohen's d, was d = .53, indicating a medium effect. During OPA stimulation, there were no significant differences between stimulus conditions ($_{\text{scene-without-object}} = 0.915$, $SD_{\text{scene-without-object}} = 0.033$, $_{\text{scene-with-object}} = 0.926$, $SD_{\text{scene-with-object}} = 0.043$, t = 1.091, p = .281), and the effect size (d = 0.21) indicates a small effect (Figure 11A). Within the scene-with-object condition, accuracy was significantly lower after LOC stimulation than after OPA stimulation (μLOC = 0.889, SDLOC = 0.065, μOPA = 0.926, SDOPA = 0.043, t = 2.317, p = .025) (Figure 11B). The effect size indicates a medium effect (d = .70). These findings confirm that the LOC is causally involved in object-facilitated scene recognition. However, we cannot make inferences about when and how the LOC is involved due to the lack of time-specific effects. There were no significant interactions between stimulus condition, stimulation site and stimulation onset on reaction time, but a significant main effect of condition was found (F(1,46)=12.205, p = .001, partial $\eta 2$ = .210). A simple main effect analysis showed that reaction times for the scene-without-object condition (µ = 0.646, SD = 0.095) were significantly lower than for the scene-with-object condition (µ = 0.660, SE = 0.100, t = 3.494, p = .001). The effect size, measured by Cohen's d, was d = 0.41, indicating a small effect. A significant two-way interaction between stimulation site and stimulus condition on LISAS was found (F(1,46)=4.799, p = .034, partial $\eta 2$ = .094). A simple interaction analysis showed that, compared to the scene-without-object condition, participants had higher LISAS in the scene-with-object condition during LOC stimulation ($_{\text{scene-without-object}}$ = 0.720, $SD_{\text{scene-without-object}}$ = 0.133, $_{\text{scene-with-object}}$= 0.748, $SD_{\text{scene-with-object}}$ = 0.123, t = 3.532, p < .001, d = 0.57) but not during OPA stimulation ($_{\text{scene-without-object}}$ = 0.686, $SD_{\text{scene-without-object}}$ = 0.119, $_{\text{scene-with-object}}$ = 0.689, $SD_{\text{scene-with-object}}$ = 0.126, t = 0.434, p = .666, d = 0.07), indicating that the previously reported effects depending on stimulus condition occurring during LOC stimulation but not during OPA stimulation were not the results of speed-accuracy trade-offs. With LOC stimulation, participants performed worse in the scene-with-object condition. In accordance with our hypotheses, the following pre-registered planned pairwise t-tests were computed. These pairwise t-tests showed that classification accuracy during LOC stimulation in the scene-without-object condition did not differ between early and middle (t = 0.571, p > .999, d = 0.10), middle and late (t = 0.721, p > .999, d = 0.11) or early and late stimulation onset (t = 0.109, p > .999, d = 0.01). Contrary to our hypotheses, no differences between early and middle (t = 0.130, p > .999, d = 0.03) and middle and late stimulation onset (t = 1.882, p = .218, d = 0.36) were found during LOC stimulation in the scene-with-object condition. Further, planned pairwise t-tests did not reveal significant differences in accuracy in the scene-without-object condition during OPA stimulation between early and middle (t = 0.322, p > .999, d = 0.03) or middle and late stimulation onset (t = 0.345, p > .999, d = 0.03). Finally, contrary to our hypotheses, OPA stimulation in the scene-with-object condition showed no differences in accuracy between early and middle (t = 0.002, p > .999, d = 0.04), middle and late (t = 0.005, p > .999, d = 0.10) or early and late stimulation onset (t = 0.878, p > .999, d = 0.13).

6 General Discussion

In the current study, we examined the facilitatory effect of objects placed in scenes on scene recognition. The key findings were that objects facilitate scene recognition both behaviourally and neurally through involvement of the LOC. Scene recognition was significantly improved when an object was present in the ambiguous scene as compared to the scene without the object or the object on a neutral background. The facilitatory effect of the object in the scene completely diminished once the scene loses its structure and decreased when the object was in a random position in the scene, as compared to its original position. Finally, brain stimulation results show a causal involvement of the LOC in the object-facilitatory effect on scene recognition. These results show both behavioral and causal evidence for interactions between object and scene processing, with object processing supporting scene

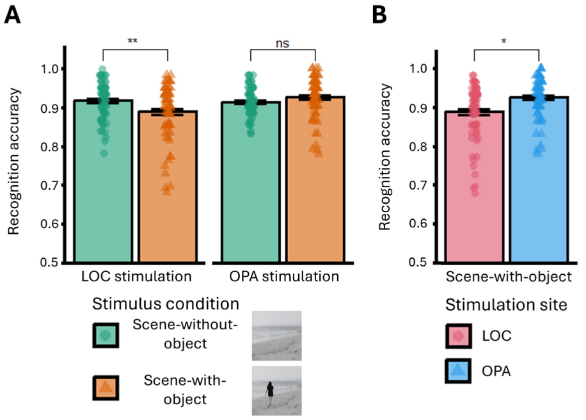


Figure 11: **Results of TMS Experiment:** *Note.* Data are presented as mean recognition accuracy, error bars indicate SEM, points represent individual participants. $**p \leq .01$; $*p \leq .05$ (A) Accuracy in the scene-with-object condition was significantly decreased during LOC stimulation, but not during OPA stimulation. (B) This is a comparison between LOC and OPA stimulation in the scene-with-object condition, shown in a separate graph for clarity. In the scene-with-object condition, accuracy was significantly lower during LOC stimulation than during OPA stimulation.

recognition through involvement of the LOC.

6.1 Behavioural Experiments

In Experiment 1, participants were significantly better at correctly recognizing the scene when an object was present in the scene. Crucially, participants were least accurate when only the object was shown on a neutral background, indicating that the object itself was not sufficient for accurate scene classification. This demonstrates that scenes not only enhance object recognition (Brandman and Peelen, 2019), but that this effect also occurs the other way around in that objects facilitate ambiguous scene recognition. These findings are in accordance with previous studies showing a facilitatory effect of objects on scene recognition behaviourally (Davenport and Potter, 2004) and on a neural scale, where an object enhanced scene representation in the left scene selective OPA (Brandman and Peelen, 2019). While previous studies found a congruency effect for scene recognition with object placed in scenes (Davenport and Potter, 2004; Davenport, 2007), these studies used both intact scenes and objects indicative of a specific scene. On the other hand, our study used ambiguous scenes as well as objects that are equally likely to be found indoors and outdoors and are thus not indicative of scene category just by object category alone. Importantly, we still found an object-facilitatory effect on scene classification. This suggests that if the scene contains little information, scene classification relies more strongly on the object even if the object category itself provides little information about the scene. Thus, other properties of the object like object position, size or how it is embedded in the scene must drive object-facilitated scene recognition. Previous research showed that spatial aspects of a scene are crucial for rapid scene recognition in humans (Greene and Oliva, 2009) and DNNs (Weng et al., 2016). In line with this, the current findings could indicate that the object properties are used to disambiguate the spatial scene aspects.

In Experiment 2, we assessed how scene classification changes when the scene retains some low-level properties, but scene structure is removed through phase scrambling. Participants were less accurate when the object was shown in the phase scrambled scene than when it was shown in the ambiguous scene, and crucially, accuracy in the scrambled-scene-with-object condition was not significantly different from when the object was placed on a neutral background. Thus, object information is not merely combined with low-level scene properties to facilitate high-level scene recognition. Instead, this finding highlights the importance of scene structure over low-level properties for object-facilitated scene recognition and shows that the object-facilitatory effect can no longer occur once scene structure is removed. Hence, this suggests that the object aids in building scene structure and that it is this construction of scene structure based on the object that facilitates scene recognition. Using jumbling paradigms, previous literature demonstrated a sensitivity of scene-selective areas for spatial scene structure that, importantly, did not purely depend on low-level features (Kaiser et al., 2019). On the other hand, Oliva and Schyns, 2000 demonstrated colour to be a useful cue for scene recognition even with strong blur, and Yao and Einhäuser, 2008 provided evidence of colours to aid natural scene recognition in the late stages of scene recognition. In relation to our current findings, this indicates that the phase-scrambling disrupts the early stages of scene processing, such that colour can no longer be a useful indicator, and therefore emphasizes the importance of scene structure over colour perception for the distinction of indoor and outdoor scenes. Thus, the facilitatory effect of the object is based on constructing the scene and its content around the object. Another possible explanation for the lack of object facilitatory effect when the object is placed in a scrambled scene is that the scrambled scene is no longer recognized as a scene. Schindler and Bartels, 2016 showed that low-level features alone are not enough for evoking responses in scene-selective areas, but that high-level stimulus interpretation is necessary as well. Since phase scrambling inhibits the recognition of high-level image properties, scene-selective areas could be no longer activated in response to the scrambled scene. Hence, the spatial aspects of the scene can no longer be disambiguated by object predictions in scene-selective areas.

To determine whether the object-facilitatory effect depends on the object building scene structure, we conducted Experiment 3. In this experiment, we compared object-facilitated scene recognition with the object at its original position in the scene against object-facilitated scene recognition when the object was at a random position in the scene. As in Experiment 1, scene classification was better when an object was present in the scene than in any other condition. Notably, when an object was present in the scene, accuracy was significantly lower when the object was not in its original position, and the object-facilitatory effect was decreased. However, when the object was shown on a neutral background, position changes did not have any effect on scene classification accuracy. This suggests that a change in position of the object disrupts the disambiguation of scene layout through the object. In other words, changing object position no longer allows for a good estimation of the distances to scene surfaces that are crucial for rapid scene recognition (Kravitz et al., 2011). Together with the results of Davenport and Potter, 2004) and Davenport, 2007), who showed that semantic violations

of object categories in a scene lead to worse scene recognition, these results suggest that the rules and regularities of our visual environment, described by Võ, 2021) as scene grammar, are not only used in object recognition, but also in scene recognition. This is further supported by the finding that object position does not change scene recognition accuracy when the object was presented on a neutral background. Since there is no scene, no scene layout is built, and changing the position of the object does not violate the syntactic and semantic regularities used to build up scene grammar.
Troiani et al., 2014 hypothesized that scene-preferring regions, such as the OPA or PPA, also process spatial properties that are applicable to both scenes and objects. In line with this hypothesis, the reduction in scene recognition accuracy with the object at a different location can be explained by the landmark suitability of the object being severely diminished once removed from its original position. The results of this experiment further highlight that it is not object category per se that facilitates scene recognition, but how the object is integrated into the scene. Future research should investigate which aspects of object position specifically are important. Here, we assessed position change with a binary measure (original vs. non-original position), but systematic assessment of position changes of the same distance on either axis or manually changing object positions to locations in the scene that can either naturally occur or are impossible to occur naturally (e.g., move a chair to a new position on the lawn vs moving it to a position that suggests it is floating) will allow for a precise investigation of the importance of object position in disambiguating scene layout. These behavioral findings not only support previous findings that objects facilitate scene recognition, but they also provide novel evidence for this effect not to be a purely semantic effect, but that the integration of the object in the scene is crucial for its disambiguation. Further, they highlight the importance of scene structure remaining in the scene for this facilitatory effect to occur.

6.2 Chronometric Brain Stimulation Experiment

Using TMS to assess the causal neural mechanisms of object-facilitated scene recognition, we found that scene classification performance during LOC stimulation decreased when an object was present in the scene compared to when a scene without an object was shown. This involvement of the LOC only during the object-facilitated scene recognition but not intact scene recognition provides causal evidence for its involvement in object-facilitated scene recognition and is congruent with previous finding that showed an involvement of the LOC in pure object recognition (Dilks et al., 2013; Wischnewski and Peelen, 2021a) and research that suggested the LOC to support scene recognition through combining the information of multiple "signature objects" in a natural scene (MacEvoy and Epstein, 2011). In addition to their proposed role of the LOC, our results show that even a single object that is not associated with the scene category is sufficient for the LOC to facilitate scene recognition. Being sensitive to object position, size, and viewpoint (Eger et al., 2008; Sayres and Grill-Spector, 2008; Andresen et al., 2009), this result suggests that rather than object (category) recognition facilitating scene recognition per se, the recognition of the object provides information about the distance of the object in the scene, given its real world size together with the size of the object, as presented on-screen, on the retina. This information is then forwarded to the scene-selective areas to facilitate scene recognition. In Experiment 1, we showed that the object itself was not sufficient for accurate scene recognition and in Experiment 3, we highlighted the importance of object position for the facilitatory effect. Taken together, the behavioural and brain stimulation results indicate that object-facilitated scene recognition it is not driven the recognition of object category per se, but by spatial properties of the object that are extracted once the object is recognized. However, since no time-specific effects were found, we cannot say how the LOC is involved in object-facilitated scene recognition and therefore neither support scene-first (Schyns and Oliva, 1994;Hochstein and Ahissar, 2002), object-first (Crouzet et al., 2012) nor parallel processing accounts (Peelen et al., 2024). The null result of OPA stimulation is not in accordance with previous studies that showed an involvement of the OPA in scene recognition, both with neuroimaging (R. A. Epstein and Baker, 2019) and brain stimulation (Dilks et al., 2013, Wischnewski and Peelen, 2021a). This could be due to a variety of factors. On the one hand, we stimulated the left hemisphere, whereas previously, the right OPA was stimulated (Dilks et al., 2013;Wischnewski and Peelen, 2021a; Gandolfo and Downing, 2019). The stimulation of the left OPA was based on fMRI results that showed higher activation for object-facilitated scene recognition in the left OPA as compared to the right OPA (Brandman and Peelen, 2019). Another explanation for this null effect could be that we stimulated the wrong area. Despite previous research (Wischnewski and Peelen, 2021b) also using Talairach coordinates and the Localite neuronavigation system, they stimulated the right hemisphere, whereas we used the homologue coordinates of coordinates defined in the right hemisphere to stimulate the left hemisphere. Thus, it may be that interhemispheric differences in the position of the target areas lead to inaccurate stimulation sites. On top of that, interindividual differences in brain anatomy may have resulted in even bigger inaccuracies of stimulation site. E.g., brain-scalp distance has been shown to influence stimulation intensity at the target region and larger CSF thickness has been shown to result in less intense and more diffuse stimulation (Lee et al., 2018). Using individual fMRI coordinates and localisation procedures, as used in Dilks et al., 2013 or Gandolfo and Downing, 2019, would have allowed for better localisation. However, we did use functional localization in the selection procedure to counteract this issue. Previous literature on brain stimulation has reported large individual differences in effects of brain stimulation. During the selection procedure, participants received a fixed stimulation intensity at 60% of maximal output strength. This fixed stimulation intensity does not account for interindividual differences, possibly leading to decreased effects of stimulation. Factors such as age, hemisphere, and TMS machine all influence susceptibility to brain stimulation (Corp et al., 2021; Lee et al., 2018). While an initial stage of cortical hypoexcitability to brain stimulation below the age of 20 has been shown with increasing excitability until the age of 25 to 35, hypoexcitability has been shown to be stronger, the younger participants are (Corp et al., 2021). Falling between the age of 18 and 33, it cannot be excluded that some of our participants still are in the initial stage of hypoexcitability and would have required a higher stimulation intensity to show more pronounced stimulation effects, allowing for better functional localization. The left hemisphere has also been shown to have a reduced resting motor threshold (Corp et al., 2021), although this effect disappeared when including only right-handed participants. Since we only tested right-handed participants, it is unlikely that the change of hemisphere itself is cause to a lack of effect. In their large-scale analysis, Corp et al., 2021 have also shown the TMS machine itself to have an influence on stimulation effects. While we used a different TMS machine and coil from Dilks et al., 2013, who first showed a double dissociation between OPA and LOC in scene- and object-recognition tasks, we used the same TMS machine for all participants and the same coil for most of our participants as Wischnewski and Peelen, 2021a, who later replicated the findings of Dilks et al., 2013 with a larger sample size. Removing those participants with a different coil and neuronavigator from our data set did not alter significance levels, and therefore, machine and coil are unlikely to cause of the lack of effect either. While the previously mentioned aspects could have decreased the validity of the selection procedure, they would be controlled for during the chronometric TMS study assessing time-specific effects of stimulation site. Hence, the driving factor for the lack of effect likely is the use of homologue coordinates of right-hemispheric defined Talairach coordinates for stimulating left OPA and LOC. It is recommended for future research to use individual localisation procedures and base the neuronavigation off of individual fMRI coordinates. Further, previous scene categorization tasks used for online TMS assessed scene orientation (Gandolfo et al., 2024), scene category (Dilks et al., 2013) or were related to navigation (Julian et al., 2016). Thus, one reason that may give rise to the lack of effect of OPA stimulation could be the task itself, as the task in our study required participants to make indoor-outdoor judgements. The OPA has been proposed to be biased towards the lower half of the visual field, reflecting navigational affordances of scenes (Bonner and Epstein, 2017, Bonner and Epstein, 2018), whereas the PPA is more biased towards the upper half of the visual field (Silson et al., 2015) including mostly landmarks and immovable objects (R. A. Epstein and Baker, 2019). This suggests that the PPA may be involved more in indoor vs outdoor judgements than the OPA. Similarly, the PPA has been hypothesized to summarize visual input according to its spatial properties (MacEvoy and Epstein, 2011; Park et al., 2011) and was shown to integrate the object the most in its representation (Aminoff and Durham, 2023). With object position playing an influential role in the facilitatory effect on scene recognition, it may be that the feedback between object- and scene-selective regions is stronger between the LOC and the PPA than between the LOC and the OPA, as hypothesized. Then, information about object position from the LOC can be used by the PPA to disambiguate the scene layout and therefore facilitate the recognition of the scene. Future brain stimulation research could consider stimulating the PPA when assessing object-on-scene interactions. While this is currently not possible using TMS, it could be achieved using Transcranial Ultrasound Stimulation (TUS), given that this technology continues its promising development. Rafique et al., 2015 have shown interhemispheric connections between the left and right LOC, where inhibition of the LOC in one hemisphere resulted in worse object processing in the LOC of the other hemisphere. However, they did not find this interhemispheric dependence for left and right OPA. In relation to our findings this may suggest that, after stimulation of the LOC, object processing is inhibited wholly, whereas after OPA stimulation, the right OPA is able to substitute for the left OPA in scene recognition.

6.3 Future Directions

The object-facilitated scene recognition identified here opens avenues for additional future behavioural research. For example, future research could compare the influences of low-level image properties, such as object size or position, to high-level properties, such as object category or object-scene congruence. Further, a future study could create a stimulus set consisting of ambiguous scenes with both congruent and incongruent objects within. These objects should not be placed in the scene through photo editing, but photographs of the same scene with the different objects should be used, thereby controlling for confounds that arise with placing objects in scenes artificially, such as differences in lighting or saturation. Future research could also investigate the importance of spatial relations in object-facilitated scene recognition by manipulating the perceived distance between object and observer and other spatial relations by randomly varying object size, similar to how Experiment 3 varied object position. Another line of research could progressively blur the facilitatory object in the scene to determine how ambiguous objects can facilitate the recognition of ambiguous scenes and assess the importance of object properties like texture and shading. Finally, to investigate whether the absence of the object-facilitatory effect in Experiment 2 is based on the scrambled scene no longer being recognized as a scene due to the lack of scene structure, future research could ask the participants to categorize the scene in a larger set of scene categories for a more sensitive measure of scene recognition.

7 Conclusion

The current research adds to the body of research that aims to explain how humans are so proficient at rapidly perceiving the visual environment. We show that humans capitalize on the visual information that is available to us in any given moment and combine it to effortlessly navigate through our environment and perform daily life tasks, such as determining whether one is currently situated indoors or outdoors. More specifically, we provide convincing behavioural and neural evidence for objects to facilitate scene processing. This object-facilitatory effect does not solely rely on object category but depends on object position, since the object is used to construct the structure of the scene. On a neural scale, the LOC is causally involved in object-facilitated scene recognition. Together, the current results provide novel behavioral and neural evidence for the hypothesis that objects facilitate scene recognition through involvement of the LOC.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron*, *21*(2), 373–383. https://doi.org/10.1016/S0896-6273(00)80546-2

Aminoff, E. M., & Durham, T. (2023). Scene-selective brain regions respond to embedded objects of a scene. *Cerebral Cortex*, *33*(9), 5066–5074. https://doi.org/10.1093/cercor/bhac399

Andresen, D. R., Vinberg, J., & Grill-Spector, K. (2009). The representation of object viewpoint in human visual cortex. *NeuroImage*, *45*(2), 522–536. https://doi.org/10.1016/j.neuroimage.2008.11.009

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Bastin, J., Vidal, J. R., Bouvier, S., Perrone-Bertolotti, M., Bénis, D., Kahane, P., David, O., Lachaux, J. P., & Epstein, R. A. (2013). Temporal components in the parahippocampal place area revealed by human intracerebral recordings. *Journal of Neuroscience*, *33*(24), 10123–10131. https://doi.org/10.1523/JNEUROSCI.4646-12.2013

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177. https://doi.org/10.1016/0010-0285(82)90007-X

Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences*, *114*(18), 4793–4798. https://doi.org/10.1073/pnas.1618228114

Bonner, M. F., & Epstein, R. A. (2018). Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLoS Computational Biology*, *14*(4), e1006111. https://doi.org/10.1371/journal.pcbi.1006111

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, *86*, 35–42. https://doi.org/10.1016/j.visres.2013.04.006

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556–566. https://doi.org/10.1037/0096-1523.15.3.556

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fmri and meg decoding. *Journal of Neuroscience*, *37*(32), 7700–7710. https://doi.org/10.1523/JNEUROSCI.0582-17.2017

Brandman, T., & Peelen, M. V. (2019). Signposts in the fog: Objects facilitate scene representations in left scene-selective cortex. *Journal of Cognitive Neuroscience*, *31*(3), 390–400. https://doi.org/10.1162/jocn_a_01258

Brandman, T., & Peelen, M. V. (2023). Objects sharpen visual scene representations: Evidence from meg decoding. *Cerebral Cortex*, *33*(16), 9524–9531. https://doi.org/10.1093/cercor/bhad222

Carlson, T., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1–1. https://doi.org/10.1167/13.10.1

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. https://doi.org/10.1038/13217

Choo, H., & Walther, D. B. (2016). Contour junctions underlie neural representations of scene categories in high-level human visual cortex. *NeuroImage*, *135*, 32–44. https://doi.org/10.1016/j.neuroimage.2016.04.021

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Corp, D. T., Bereznicki, H. G., Clark, G. M., Youssef, G. J., Fried, P. J., Jannati, A., Davies, C. B., Gomes-Osman, J., Kirkovski, M., Albein-Urios, N., Fitzgerald, P. B., Koch, G., Di Lazzaro, V., Pascual-Leone, A., & Enticott, P. G. (2021). Large-scale analysis of interindividual variability in single and paired-pulse tms data. *Clinical Neurophysiology*, *132*(10), 2639–2653. https://doi.org/10.1016/j.clinph.2021.06.014

Crouzet, S. M., Joubert, O. R., Thorpe, S. J., & Fabre-Thorpe, M. (2012). Animal detection precedes access to scene category. *PLoS ONE*, *7*(12), e51471. https://doi.org/10.1371/journal.pone.0051471

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393–401. https://doi.org/10.3758/BF03193280

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *Journal of Neuroscience*, *31*(31), 11305–11312. https://doi.org/10.1523/JNEUROSCI.1935-11.2011

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, *33*(4), 1331–1336. https://doi.org/10.1523/JNEUROSCI.4081-12.2013

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Eger, E., Kell, C. A., & Kleinschmidt, A. (2008). Graded size sensitivity of object-exemplar–evoked activity patterns within human loc subregions. *Journal of Neurophysiology*, *100*(4), 2038–2047.

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

Epstein, R. A., & Baker, C. I. (2019). Scene perception in the human brain. *Annual Review of Vision Science*, *5*, 373–397. https://doi.org/10.1146/annurev-vision-091718-014809

Epstein, R. A., & Morgan, L. K. (2012). Neural responses to visual scenes reveals inconsistencies between fmri adaptation and multivoxel pattern analysis. *Neuropsychologia*, *50*(4), 530–543. https://doi.org/10.1016/j.neuropsychologia.2011.09.042

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10–10. https://doi.org/10.1167/7.1.10

Furtak, M., Mudrik, L., & Bola, M. (2022). The forest, the trees, or both? hierarchy and interactions between gist and object processing during perception of real-world scenes. *Cognition*, *221*, 104983. https://doi.org/10.1016/j.cognition.2021.104983

Ganaden, R. E., Mullin, C. R., & Steeves, J. K. (2013). Transcranial magnetic stimulation to the transverse occipital sulcus affects scene but not object processing. *Journal of Cognitive Neuroscience*, *25*(6), 961–968. https://doi.org/10.1162/jocn_a_00372

Gandolfo, M., Abassi, E., Balgova, E., Downing, P. E., Papeo, L., & Koldewyn, K. (2024). Converging evidence that left extrastriate body area supports visual sensitivity to social interactions. *Current Biology*, *34*(2), 343–351. https://doi.org/10.1016/j.cub.2023.12.009

Gandolfo, M., & Downing, P. E. (2019). Causal evidence for expression of perceptual expectations in category-selective extrastriate regions. *Current Biology*, *29*(15), 2496–2500. https://doi.org/10.1016/j.cub.2019.06.024

Greene, M. R., & Oliva, A. (2009). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. https://doi.org/10.1016/j.cogpsych.2008.06.001

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*(2), 159–166. https://doi.org/10.1016/S0959-4388(03)00040-0

Grill-Spector, K., Kushnir, T., Edelman, S., Itzchak, Y., & Malach, R. (1998). Cue-invariant activation in object-related areas of the human occipital lobe. *Neuron*, *21*(1), 191–202. https://doi.org/10.1016/S0896-6273(00)80526-7

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*(8), 536–548. https://doi.org/10.1038/nrn3747

Groen, I. I., Ghebreab, S., Lamme, V. A., & Scholte, H. S. (2016). The time course of natural scene perception with reduced attention. *Journal of Neurophysiology*, *115*(2), 931–946. https://doi.org/10.1152/jn.00896.2015

Groen, I. I., Ghebreab, S., Prins, H., Lamme, V. A., & Scholte, H. S. (2013). From image statistics to scene gist: Evoked neural activity reveals transition from low-level natural image structure to scene category. *Journal of Neuroscience*, *33*(48), 18814–18824. https://doi.org/10.1523/JNEUROSCI.3128-13.2013

Harel, A., Groen, I. I., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted eeg investigation. *eNeuro*, *3*(5). https://doi.org/10.1523/ENEURO.0139-16.2016

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*(1), 243–271. https://doi.org/10.1146/annurev.psych.50.1.243

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/S0896-6273(02)01091-7

Ishai, A., Ungerleider, L. G., Martin, A., & Haxby, J. V. (2000). The representation of objects in the human occipital and temporal cortex. *Journal of Cognitive Neuroscience*, *12*(Suppl 2), 35–51. https://doi.org/10.1162/089892900564055

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297. https://doi.org/10.1016/j.visres.2007.09.013

Julian, J. B., Ryan, J., Hamilton, R. H., & Epstein, R. A. (2016). The occipital place area is causally involved in representing environmental boundaries during navigation. *Current Biology*, *26*(8), 1104–1109. https://doi.org/10.1016/j.cub.2016.02.066

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, *23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Kanwisher, N., McDermott, J., & Chun, M. M. (2002). The fusiform face area: A module in human extrastriate cortex specialized for face perception [DOI: 10.1523/JNEUROSCI.17-11-04302.1997]. *Journal of Neuroscience*.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*(5534), 1506–1509. https://doi.org/10.1126/science.1061133

Kravitz, D. J., Peng, C. S., & Baker, C. I. (2011). Real-world scene representations in high-level visual cortex: It's the spaces more than the places. *Journal of Neuroscience*, *31*(20), 7322–7333. https://doi.org/10.1523/JNEUROSCI.4588-10.2011

Lauer, T., Cornelissen, T. H. W., Draschkow, D., Willenbockel, V., & Võ, M. L. H. (2018). The role of scene summary statistics in object recognition. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-32991-1

Lee, E. G., Rastogi, P., Hadimani, R. L., Jiles, D. C., & Camprodon, J. A. (2018). Impact of non-brain anatomy and coil orientation on inter- and intra-subject variability in tms at midline. *Clinical Neurophysiology*, *129*(9), 1873–1883. https://doi.org/10.1016/j.clinph.2018.04.749

Leroy, A., Faure, S., & Spotorno, S. (2020). Reciprocal semantic predictions drive categorization of scene contexts and objects even when they are separate. *Scientific Reports*, *10*(1), 8447. https://doi.org/10.1038/s41598-020-65158-y

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, timing, timing: Fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*, *62*(2), 281–290. https://doi.org/10.1016/j.neuron.2009.02.025

Lupyan, G. (2015). Object knowledge changes visual appearance: Semantic effects on color afterimages. *Acta Psychologica*, *161*, 117–130. https://doi.org/10.1016/j.actpsy.2015.08.006

Lupyan, G. (2017). Objective effects of knowledge on visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(4), 794–806. https://doi.org/10.1037/xhp0000343

Macdonald, S. N., & Culham, J. C. (2015). Do human brain areas involved in visuomotor actions show a preference for real tools over visually similar non-tools? *Neuropsychologia*, *77*, 35–41. https://doi.org/10.1016/j.neuropsychologia.2015.08.004

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

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., Ledden, P. J., Rosen, B. R., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, *92*(18), 8135–8139. https://doi.org/10.1073/pnas.92.18.8135

Mendola, J. D., Dale, A. M., Fischl, B., Liu, A. K., & Tootell, R. B. (1999). The representation of illusory and real contours in human cortical visual areas revealed by functional magnetic resonance imaging. *Journal of Neuroscience*, *19*(19), 8560–8572. https://doi.org/10.1523/JNEUROSCI.19-19-08560.1999

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107. https://doi.org/10.1006/cogp.1997.0667

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

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: Complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience*, *31*(4), 1333–1340. https://doi.org/10.1523/JNEUROSCI.3885-10.2011

Peelen, M. V., Berlot, E., & de Lange, F. P. (2024). Predictive processing of scenes and objects. *Nature Reviews Psychology*, *3*(1), 13–26. https://doi.org/10.1038/s44159-023-00254-0

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. *Trends in Cognitive Sciences*, *18*(5), 242–250. https://doi.org/10.1016/j.tics.2014.02.004

Persichetti, A. S., & Dilks, D. D. (2018). Dissociable neural systems for recognizing places and navigating through them. *Journal of Neuroscience*, *38*(48), 10295–10304. https://doi.org/10.1523/JNEUROSCI.1200-18.2018

Proklova, D., Kaiser, D., & Peelen, M. V. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate–inanimate distinction. *Journal of Cognitive Neuroscience*, *28*(5), 680–692. https://doi.org/10.1162/jocn_a_00924

Rafique, S. A., Solomon-Harris, L. M., & Steeves, J. K. (2015). Tms to object cortex affects both object and scene remote networks while tms to scene cortex only affects scene networks. *Neuropsychologia*, *79*, 86–96. https://doi.org/10.1016/j.neuropsychologia.2015.10.027

Rossel, P., Peyrin, C., Roux-Sibilon, A., & Kauffmann, L. (2022). It makes sense, so i see it better! contextual information about the visual environment increases its perceived sharpness. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(4), 331. https://doi.org/10.1037/xhp0000993

Sayres, R., & Grill-Spector, K. (2008). Relating retinotopic and object-selective responses in human lateral occipital cortex. *Journal of Neurophysiology*, *100*(1), 249–267. https://doi.org/10.1152/jn.01383.2007

Schindler, A., & Bartels, A. (2016). Visual high-level regions respond to high-level stimulus content in the absence of low-level confounds. *NeuroImage*, *132*, 520–525. https://doi.org/10.1016/j.neuroimage.2016.03.011

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–200. https://doi.org/10.1111/j.1467-9280.1994.tb00500.x

Silson, E. H., Chan, A. W. Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience*, *35*(34), 11921–11935. https://doi.org/10.1523/JNEUROSCI.0137-15.2015

Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human ventral temporal cortex. *eLife*, *8*, e47142. https://doi.org/10.7554/eLife.47142

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bülthoff, H. H. (2001). Detection of animals in natural images using far peripheral vision. *European Journal of Neuroscience*, *14*(5), 869–876. https://doi.org/10.1046/j.0953-816x.2001.01717.x

Tran, T. H. C., Rambaud, C., Despretz, P., & Boucart, M. (2010). Scene perception in age-related macular degeneration. *Investigative Ophthalmology & Visual Science*, *51*(12), 6868–6874. https://doi.org/10.1167/iovs.10-5517

Troiani, V., Stigliani, A., Smith, M. E., & Epstein, R. A. (2014). Multiple object properties drive scene-selective regions. *Cerebral Cortex*, *24*(4), 883–897. https://doi.org/10.1093/cercor/bhs364

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Vandierendonck, A. (2018). Further tests of the utility of integrated speed-accuracy measures in task switching. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.6

Võ, M. L. H. (2021). The meaning and structure of scenes. *Vision Research*, *181*, 10–20. https://doi.org/10.1016/j.visres.2020.11.003

Weng, C., Wang, H., Yuan, J., & Jiang, X. (2016). Discovering class-specific spatial layouts for scene recognition. *IEEE Signal Processing Letters*, *24*(8), 1143–1147. https://doi.org/10.1109/LSP.2016.2641020

Wischnewski, M., & Peelen, M. V. (2021a). Causal evidence for a double dissociation between object- and scene-selective regions of visual cortex: A preregistered tms replication study. *Journal of Neuroscience*, *41*(4), 751–756. https://doi.org/10.1523/JNEUROSCI.2162-20.2020

Wischnewski, M., & Peelen, M. V. (2021b). Causal neural mechanisms of context-based object recognition. *eLife*, *10*, e69736. https://doi.org/10.7554/eLife.69736

Yao, A. Y., & Einhäuser, W. (2008). Color aids late but not early stages of rapid natural scene recognition. *Journal of Vision*, *8*(16), 12–12. https://doi.org/10.1167/8.16.12