Influence of physical features from peripheral vision on scene categorization in central vision

Audrey Trouilloud, Pauline Rossel, Cynthia Faurite, Alexia Roux-Sibilon, Louise Kauffmann, Carole Peyrin

This is the submitted version of the manuscript of an article published by Taylor & Francis in Visual Cognition on 14 june 2022, available online:
http://www.tandfonline.com/10.1080/13506285.2022.2087814.

Published source:
Audrey Trouilloud, Pauline Rossel, Cynthia Faurite, Alexia Roux-Sibilon, Louise Kauffmann & Carole Peyrin (2022). Influence of physical features from peripheral vision on scene categorization in central vision, Visual Cognition, 30:6, 425-442, DOI: 10.1080/13506285.2022.2087814

Influence of physical features from peripheral vision on scene categorization in central vision

Audrey Trouilloud[1], Pauline Rossel[1], Cynthia Faurite[1], Alexia Roux-Sibilon[2], Louise Kauffmann[1] and Carole Peyrin[1]

[1] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France
[2] Psychological Sciences Research Institute (IPSY), UC Louvain, Louvain-la-Neuve, Belgium

Short title: Scene perception in peripheral vision

***Corresponding Author:**
Carole Peyrin, PhD
Laboratoire de Psychologie et NeuroCognition (LPNC)
CNRS UMR 5105 - Université Grenoble Alpes
Bâtiment Michel Dubois - 1251 Av Centrale CS40700
38058 Grenoble Cedex 9 - France
carole.peyrin@univ-grenoble-alpes.fr

Abstract

The spatial resolution of the human visual field decreases considerably from the center to the periphery. However, several studies have highlighted the importance of peripheral vision for scene categorization. In Experiment 1, we investigated if peripheral vision could influence the scene categorization in central vision. We used photographs of indoor and outdoor scenes from which we extracted a central disk and a peripheral ring. Stimuli were composed of a central disk and a peripheral ring that could be either semantically congruent or incongruent. Participants had to categorize the central disk while ignoring the peripheral ring or the peripheral ring while ignoring the central disk. Results revealed a congruence effect of peripheral vision on central vision, as strong as the reverse. In Experiment 2, we investigated the nature of the physical signal in peripheral vision that influences the categorization in central vision. We used either intact, phase-preserved, or amplitude-preserved peripheral rings. Participants had to categorize the central disk while ignoring the peripheral ring. Results showed that only phase-preserved peripheral rings elicited a congruence effect as strong as the one observed with intact peripheral rings. Information contained in the phase spectrum (spatial configuration of the scene) may be critical in peripheral vision.

Keywords: Scene perception; Peripheral vision; Central vision; Congruence effect; Spatial configuration

Introduction

The human visual field is the visible portion of the environment for a given retinal image during eye fixation. Central vision is the small portion of the visual field where the eyes look directly. It is considered as the most important part of the visual field, certainly due to its very good spatial resolution, or visual acuity. Visual acuity is the highest in a small area of central vision, the fovea, which is estimated to cover approximately 1° of eccentricity (Curcio & Allen, 1990; Curcio et al., 1990; Larson & Loschky, 2009; Yamada, 1969). The remaining part of the visual field projects onto the parafovea, which extends to about 4-5° of retinal eccentricity, and onto the peripheral retina, which covers the entire remaining visual field (Larson & Loschky, 2009; Loschky, Nuthmann, Fortenbaugh, & Levi, 2017). Although our subjective visual experience seems rich and detailed (in part thanks to constant eye movements), vision outside the central region is considerably degraded. At the level of the retina, spatial resolution decreases in peripheral vision due to the density of photoreceptors decreasing with retinal eccentricity. This is further amplified at the cortical level, where central vision is overrepresented relative to peripheral vision (cortical magnification). Namely, more neurons of the primary visual cortex (V1) are allocated to the processing of information from the fovea than periphery (Daniel & Whitteridge, 1961). Besides mere resolution, peripheral vision is highly sensitive to visual clutter, a phenomenon known as visual crowding whereby objects that are otherwise identifiable at a given eccentricity become unrecognizable when surrounded by other objects (Pelli, 2008; Whitney & Levi, 2011). Therefore, perception in peripheral vision is overall rather coarse. On the contrary, due to its favored processing at both the retinal and cortical levels, and the fact that it is not affected by crowding (at least in normally sighted adults, see e.g., Greenwood et al., 2012), central vision provides the best spatial precision, allowing us to read, recognize faces and objects, and do a variety of other daily tasks.

Nevertheless, several studies in visual recognition research have demonstrated the crucial role of peripheral information for a rapid categorization of scenes (Boucart et al., 2013; Larson & Loschky, 2009; Loschky et al., 2019; Trouilloud et al., 2020). In particular, peripheral vision alone is sufficient to perform rapid scene categorization. This has been first demonstrated by Larson and Loschky (2009), who used an experimental paradigm known as the window-scotoma to examine the relative contributions of central and peripheral vision during rapid scene categorization. Participants of that study were presented with either the central part of a scene (window condition), the peripheral part (scotoma condition), or the whole scene (control condition). Radii of window and scotoma stimuli varied according to four different values (1°, 5°, 10.8°, or 13.6°). According to the authors, as the central vision extends to about 5° of visual angle, the 5° radius condition was used to compare the relative contribution of central and peripheral vision. Results showed that categorization performances were better in the 5° scotoma condition (i.e., peripheral vision) than in the 5° window condition (i.e., central vision). Moreover, the performance in the scotoma condition when only peripheral information was available did not differ from the maximum performance obtained in the control condition where the whole scene was visible. Coarse peripheral vision would therefore be more useful than high resolution central vision for quickly categorizing a scene. The 10.8° radius condition was created in order to have an equal viewable area in window and scotoma stimuli. In this condition, Larson and Loschky (2009) observed a central vision advantage when the surface was equalized between central and peripheral stimuli. However, when considering the cortical magnification factor, the window stimulus activated more neurons in the primary visual cortex than the scotoma stimulus. This is the ambiguity of studies comparing central vision and peripheral vision. Peripheral vision would have an advantage over central vision for scene categorization only because it covers a larger part of the visual field revealing more scene content, for an equivalent activation of the visual cortex. Indeed, Geuzebroek and van den Berg (2018) observed that when they presented the same amount of semantic content in central and peripheral vision (by manipulating the size of a scene), the advantage of peripheral vision disappeared. The process of scene categorization would therefore be independent of the eccentricity of scene presentation if the amount of semantic information remains identical in central and peripheral vision, which is not the case in real-life conditions. It should however be noted that some experiments revealed an advantage of central vision. For example, in a subsequent experiment using the same paradigm, Larson, Freeman, Ringer, and Loschky (2014) manipulated the presentation times to examine how the relative contributions of the central and peripheral visual field change over time (Larson, Freeman, Ringer, & Loschky, 2014). The results revealed an advantage in favor of central vision during the first 100 ms, after which the relative contribution of peripheral vision increases. To interpret these results, the authors proposed the zoom-out hypothesis in which attention is first allocated at the locus of fixation and expands outward over the span of a fixation.

The advantage of peripheral vision was also found by Trouilloud et al. (2020) who replicated the findings of Larson and Loschky (2009) using stimuli that consider the cortical magnification factor. More precisely, stimuli were built so that the image content presented in peripheral and central vision activated approximately the same amount of neurons in the primary visual cortex. However, in this study, the advantage of peripheral vision was observed even for very short presentation time (33 ms), which contradicts the zoom-out hypothesis. To go further, Trouilloud et al. (2020) conducted a second experiment to determine whether scene analysis followed a peripheral to central analysis rather than a central to peripheral analysis during rapid scene categorization. To this end, they created dynamic sequences of the scene which participants had to categorize as indoor or outdoor scenes. The sequences were composed of five versions of a scene revealing part of the visual information through a set of rings of different eccentricities presented sequentially, from the peripheral to the central part or from the central to the most peripheral part. The results revealed an advantage for the peripheral to central categorization sequences. Contrary to the zoom-out hypothesis, peripheral information could therefore allow a first categorization of the scene which would then be validated by the analysis of the central information.

According to predictive models of visual recognition (Bar, 2003; Bar, 2007; Trapp & Bar, 2015; Kauffmann et al., 2014; Kveraga et al., 2007; Peyrin et al., 2010), low resolution information is sufficient and useful to trigger cortical predictive mechanisms that would then guide a more detailed visual analysis. Low spatial frequency content of the retinal image would be rapidly projected to the cortex (especially to the orbitofrontal cortex), allowing to activate a coarse representation of contextual information which could then be used to generate predictions about the objects likely to be found. These predictions, or expectations, would thus help to refine the possible interpretations of the object and facilitate the processing of the detailed information contained in high spatial frequencies. Predictive mechanisms based on a rapid extraction of low spatial frequencies have been first demonstrated in central vision (Kauffmann, Bourgin, Guyader, & Peyrin, 2015; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kveraga et al., 2007). For example, Kauffmann, Bourgin, Guyader, and Peyrin(2015) used hybrid images presented in central vision created by superimposing a low-pass filtered scene with a high-pass filtered scene belonging to the same or different semantic category. Participants had to categorize the high-pass filtered scene of the hybrid image while ignoring the low-pass filtered scene. Authors observed a semantic congruence effect: the categorization was better with a congruent than an incongruent scene. Since low spatial frequencies are mostly available in peripheral vision, Roux-Sibilon et al. (2019) studied the hypothesis that information from peripheral vision may be used to generate predictive signals that guide the categorization of information in central vision. The authors tested the effect of a peripheral scene background on the categorization of an object in central vision using a paradigm evaluating a semantic congruence effect. The peripheral background was either congruent (e.g., a chair in an indoor scene) or incongruent with the object (e.g., a chair in an outdoor scene). Also, the peripheral background was presented either simultaneously to the object, 30 ms before, or 150 ms before. The authors observed that participants were faster to categorize the object when it was in a congruent than incongruent environment, suggesting an automatic (or unintentional) processing of peripheral information. Importantly, this congruence effect only occurred when the peripheral background was perceived slightly before the object (30 ms and 150 ms), and the effect of congruence increased linearly with the stimulus onset asynchrony (SOA) between the peripheral background and the central object, supporting the view that peripheral information is used to guide visual processes in central vision in a predictive manner, rather than central and peripheral information being processed concurrently and integrated in a feedforward manner. Therefore, the visual system would rely on a first coarse representation from the low spatial frequencies available in the whole visual field to rapidly generate predictions useful for the recognition of visual details. However, since low spatial frequencies are mostly available in peripheral vision, we believe a prioritization of the peripheral vision on central vision to trigger predictive mechanisms.

In a recent fMRI study, Peyrin et al. (2021) highlighted the importance of both semantic and physical information available in peripheral vision in the generation of predictions. In this study, a scene to ignore was presented in peripheral vision and a target scene to categorize appeared simultaneously in central vision. The scene pairs were either semantically congruent or incongruent and could share the same physical properties (similar amplitude spectrum and spatial configuration) or not. The results showed that participants made more errors and were slower to categorize the central scene when the peripheral scene was semantically incongruent than when it was semantically congruent. At the cortical level, this semantic influence was associated with a stronger activation of the orbitofrontal cortex, consistent with previous fMRI studies addressing the predictive brain hypothesis in visual categorization (Bar et al., 2006; Kauffmann, Bourgin, Guyader, & Peyrin, 2015; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Peyrin et al., 2010). Critically, this effect was in fact observed when the two scenes shared the same physical properties, highlighting the importance of physical information available in peripheral vision in the generation of predictions. However, the nature of the physical features that were used to generate predictions remains unknown in this study. In the Fourier domain, an image is described by both its amplitude spectrum and its phase spectrum. The amplitude spectrum provides the distribution of luminance contrast as a function of orientations and spatial frequencies. The phase spectrum provides information about the relative position of the objects in a scene, that is the spatial configuration of the scene. Given that the physically similar scenes used in Peyrin et al. (2021) were matched both in terms of amplitude spectrum and spatial configuration (i.e., pixel to pixel correlation), it could not be disambiguated whether it is the information in the amplitude spectrum, in the phase spectrum, or both that is used to generate predictions. A possible answer can be provided by a second experiment conducted by Roux-Sibilon et al. (2019) in which the authors directly addressed how low-level physical features of a peripheral scene background could influence the categorization of an object in central vision. In this experiment, the peripheral background was either intact or phase scrambled (i.e., spatial frequency and orientation distribution was preserved while the phase of the scene was altered). A congruence effect was only observed for intact scene background suggesting that the information contained in the phase spectrum, rather than in the amplitude spectrum alone, was automatically used in peripheral vision to generate predictions about the category of the object in central vision.

Finally, some studies have shown that the categorization of the scene background can be also influenced by the semantic congruence of an object in the scene (Davenport, 2007; Davenport & Potter, 2004; Joubert et al., 2008). Therefore, Lukavský (2019) addressed the reciprocal influence between central and peripheral vision during scene categorization using conflicting stimuli. The author presented stimuli composed of a central disk and a peripheral ring revealing the central or peripheral parts of a scene respectively. The disk and the ring could belong to the same scene (congruent) or to two scenes from different categories (incongruent). Participants had to categorize the central disk while ignoring the peripheral ring or to categorize the peripheral ring while ignoring the central disk. Stimuli were presented for 33 ms, followed by a blank screen, and then a mask (stimulus-to-mask onset asynchrony of 200 ms). Results revealed that participants made more errors when the central disk and the ring were incongruent than congruent, whatever the position in the visual field of the stimuli to categorize (i.e., the central disk or the peripheral ring). This result indicates that the information to be ignored was automatically processed and influenced the categorization of the information to attend. Importantly, this congruence effect was as strong for a central categorization with a peripheral distractor than for a peripheral categorization with a central distractor. This result indicates that scene categorization relies on both central and peripheral visual information. To ensure that the absence of interaction was not due to a too long presentation time between the stimulus and the mask (200 ms) that could favor the processing of both central and peripheral information despite a precedence of one information on the other, Lukavský (2019) conducted a second experiment with the same procedure but additionally manipulated the stimulus-to-mask onset asynchrony. Even at the shortest asynchrony (33 ms), the congruence effect was as strong for the central and the peripheral distractors. According to the author, this result contradicted the hypothesis of a prioritization of the central processing at very short presentation time as stated in the zoom-out hypothesis of Larson et al. (2014). In this context, we can assume that it also contradicts the hypothesis of a prioritization of the peripheral processing as suggested by the advantage for the peripheral to central categorization sequences in Trouilloud et al. (2020). Lukavský (2019) interpreted the results in terms of evidence accumulation (Mirolli et al., 2010) or random walk model (Gold & Schadlen, 2007) postulating that all the perceived stimuli provide evidence supporting more or less several possible hypotheses. Evidence is then gradually accumulated by the system until the decision threshold of one of the hypotheses is reached. Since central and peripheral vision influence each other, Lukavský (2019) suggested that peripheral and central information would not be processed in isolation. Evidence from different parts of the visual field, even the to-be-ignored parts, is gradually gathered during gist recognition. Information extracted in peripheral vision would influence the analysis of central content and *vice versa*. Conflicting information in the to-be-ignored part, central or peripheral, increases uncertainty and more time is required to reach the decision boundary. In this interpretation, a congruence effect does not rely on predictive mechanisms. It should however be noted that the radius of the central disk was fixed to 5.54° of visual angle, so that it contained foveal and parafoveal information, while the peripheral ring extended from 5.54° to 11.08° of eccentricity. Due to the cortical magnification, the central disk was over-represented at the cortical level relative to the peripheral ring. Therefore, we believe that a prioritization of the peripheral vision on the central vision to trigger predictive mechanisms is still plausible after considering this limit.

The aim of the present study was to test this hypothesis using a semantic interference paradigm inspired by Lukavský (2019) but considering the cortical magnification factor. To this end, we used the stimuli of Trouilloud et al. (2020) which allow us to stimulate different parts of the visual field while controlling that approximately the same cortical surface is activated in V1. Participants were presented simultaneously with a central disk and a peripheral ring revealing part of scenes that could either belong to the same category (semantically congruent) or to different categories (semantically incongruent). In Experiment 1, participants had to categorize the central disk while ignoring the peripheral ring or categorize the peripheral ring while ignoring the central disk in the same way as the Lukavský (2019) experiment. Based on previous works using conflicting semantic stimuli (Lukavský, 2019; Roux-Sibilon et al., 2019; Peyrin et al., 2021), we expected to observe a semantic congruence effect (i.e., impaired performance in the semantically incongruent condition relative to the semantically congruent condition) regardless of the position of the scene to be categorized (central vs. peripheral). Moreover, if predictions are mostly generated in peripheral vision, we expected a greater congruence effect of the peripheral ring on the categorization of the central disk than the reverse. In Experiment 2, we investigated the nature of the physical signal in peripheral vision that influences categorization in central vision. We used either an intact peripheral ring, or a peripheral ring whose amplitude spectrum was preserved (phase-scrambling), or a peripheral ring whose phase spectrum was preserved (amplitude-scrambling). In this experiment, participants had only to categorize the central disk while ignoring the peripheral ring. As for Experiment 1, we expected to observe a semantic congruence effect for the intact peripheral ring. Furthermore, if information contained in the phase spectrum (i.e., spatial configuration of the scene) is used to generate predictions, we expected a semantic congruence effect in this condition. Similarly, if the information contained in the amplitude spectrum (i.e., distribution of information across spatial frequencies and orientations) is used to generate predictions, we expected a semantic congruence effect in this condition.

Experiment 1

Method

Participants

Twenty-four undergraduate students of Psychology from University Grenoble Alpes (19 women, mean age ± standard deviation: 20.83 ± 2.47) participated in the experiment. All were right handed, with normal or corrected-to-normal vision. The study was approved by the ethics committee of the University Grenoble Alpes (CER-Grenoble Alpes, COMUE University Grenoble Alpes, IRB00010290). All participants involved in the study gave their informed written consent.

Stimuli

We used the stimuli from Trouilloud et al. (2020)'s experiment. The stimuli were constructed from 50 color photographs from the Pixabay website (https://pixabay.com/fr/), a photo sharing site under CC0 (Creative Commons Zero) license. Half of the photographs represented indoor scenes (e.g., living room, kitchen, bathroom), and the other half represented outdoor scenes (e.g., cityscape, mountain, beach). We choose an indoor vs outdoor categorization task since the superordinate-level precedes basic-level distinctions for rapid scene categorization (Rousselet, Joubert, & Fabre-Thorpe, 2005; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Loschky & Larson, 2010; Kadar & Ben-Shahar, 2012). In addition, all photographs contained an object semantically related to the scene category in central vision (e.g., a car for an outdoor scene, a kitchen utensil for an indoor scene).

Image manipulation was performed using MATLAB (Mathworks Inc., Sherborn, MA, USA). For each scene, we built a square image of 1500 × 1500 pixels centered on an object representative of its category. To cover both central and peripheral vision, the angular size of stimuli was fixed at ~30° of visual angle. The object fitted into a 100 × 100 pixel square. For indoor scenes, the object could be a lamp, a book, a candle, a radio, a clock, a vase, a household appliance, a pair of glasses, dishes or kitchen utensils. For outdoor scenes, the object could be a house, a tree, a plant, a car, a building, a boat, a mountain or a bridge. All images were converted to 256 gray levels by averaging the luminance values from the three color channels at each pixel. All stimuli were normalized so that each individual scene used was attributed an average luminance of 0.5 and a RMS contrast (Root Mean Square) of 0.2 for pixel intensity values between 0 and 1 (i.e., an average luminance of 128 and an RMS contrast of 51 on a scale of 256 grey levels). The RMS contrast value of 0.2 was chosen to correspond to the mean RMS of the whole set of images. As some of the original images had asymmetric luminance histograms, pixel values falling outside the range 0-1 (or 0-255) for these images were clipped during this normalization process. This however resulted in very small variations in mean luminance (mean = 0.50, standard deviation = 0.002 for pixel intensity values between 0 and 1, or mean = 127, standard deviation = 0.39 on a 256 gray level scale) and RMS contrast (mean = 0.20, standard deviation = 0.004 for pixel intensity values between 0 and 1, or mean = 51, standard deviation = 1.1 on a 256 gray level scale) of the resulting images. Thus, 50 scenes of size 1500 × 1500 pixels (25 indoor and 25 outdoor scenes) were created.

Then, for each image, Trouilloud et al. (2020) built five stimuli consisting of a central disk and four peripheral rings (Fig. 1b) revealing different parts of the scene. The eccentricity and size of each stimuli was defined using an equation empirically derived from retinotopic measurement (Wu, Yan, Zhang, Jin, & Guo, 2012; for a similar procedure, see Geuzebroek & van den Berg, 2018):

$$y = \frac{\left(\int_{r\ inner}^{r\ outer} \left(553.99\ -\ 123.98\, In(x)\right)dx\right)}{7.5}$$

In this equation, y is the integrated V1 surface area in square millimeters, representing the activated cortical surface, and *r inner* and *r outer* are the inner and outer radii in degrees, respectively. This progression was chosen to better consider the properties of the visual system. Indeed, information from central and peripheral visual fields undergoes a deformation on the visual cortex: The central information is over-represented at cortical level, relative to peripheral information (i.e., cortical magnification; Daniel & Whitteridge, 1961). In order to consider these transformations at cortical level, the surface of the image revealed by the rings increased with their eccentricity. Based on this equation, Trouilloud et al. (2020) created five stimuli of different eccentricities (a central disk and four peripheral rings) but with the same surface of activation on V1 (20%, i.e. ~136 mm^3). For the central stimuli, the scene was revealed through a small disk of only 1.66° radius of visual angle (i.e. 83 pixels). For the first peripheral ring, adjacent to the central disk, the inner edge was fixed at 1.66° visual angle (83 pixels) and the outer edge was fixed at 4.10° visual angle (205 pixels). For the second peripheral ring, adjacent to the first one, the inner edge was fixed at 4.10° visual angle (205 pixels) and the outer edge was fixed at 7.10° visual angle (355 pixels). For the third peripheral ring, adjacent to the second one, the inner edge was fixed at 7.10° of visual angle (355 pixels) and the outer edge was fixed at 10.70° of visual angle (535 pixels). Finally, for the fourth peripheral ring, adjacent to the third one, the inner edge was fixed at 10.70° visual angle (535 pixels) and the outer edge was fixed at 15° visual angle (750 pixels).

In the present experiment, we only used the central disk and the third peripheral ring from Trouillloud et al. (2020; Figure 1). We specifically used the third peripheral ring for two reasons. First, it was the stimulus with the best categorization performance in peripheral vision (for average error rates and average correct response times). Second, Trouilloud et al (2020) showed that there is a large area of peripheral vision for which categorization is optimal (i.e., categorization of peripheral ring 1, 2, and 3 did not differ from the categorization of the whole scene). Beyond this area, recognition is less efficient. The central disk and the third peripheral ring stimuli were presented simultaneously. The angular size of the pattern was thus fixed at 21.4° of visual angle. The central disk and the peripheral ring could be either extracted from the same photograph and thus semantically congruent (e.g., a central indoor scene and the same peripheral indoor scene) or semantically incongruent (e.g., a central indoor scene and a peripheral outdoor scene). Stimuli were presented on a gray background of 0.5 average luminance.

Stimuli can be downloaded from https://osf.io/q3rtw/.

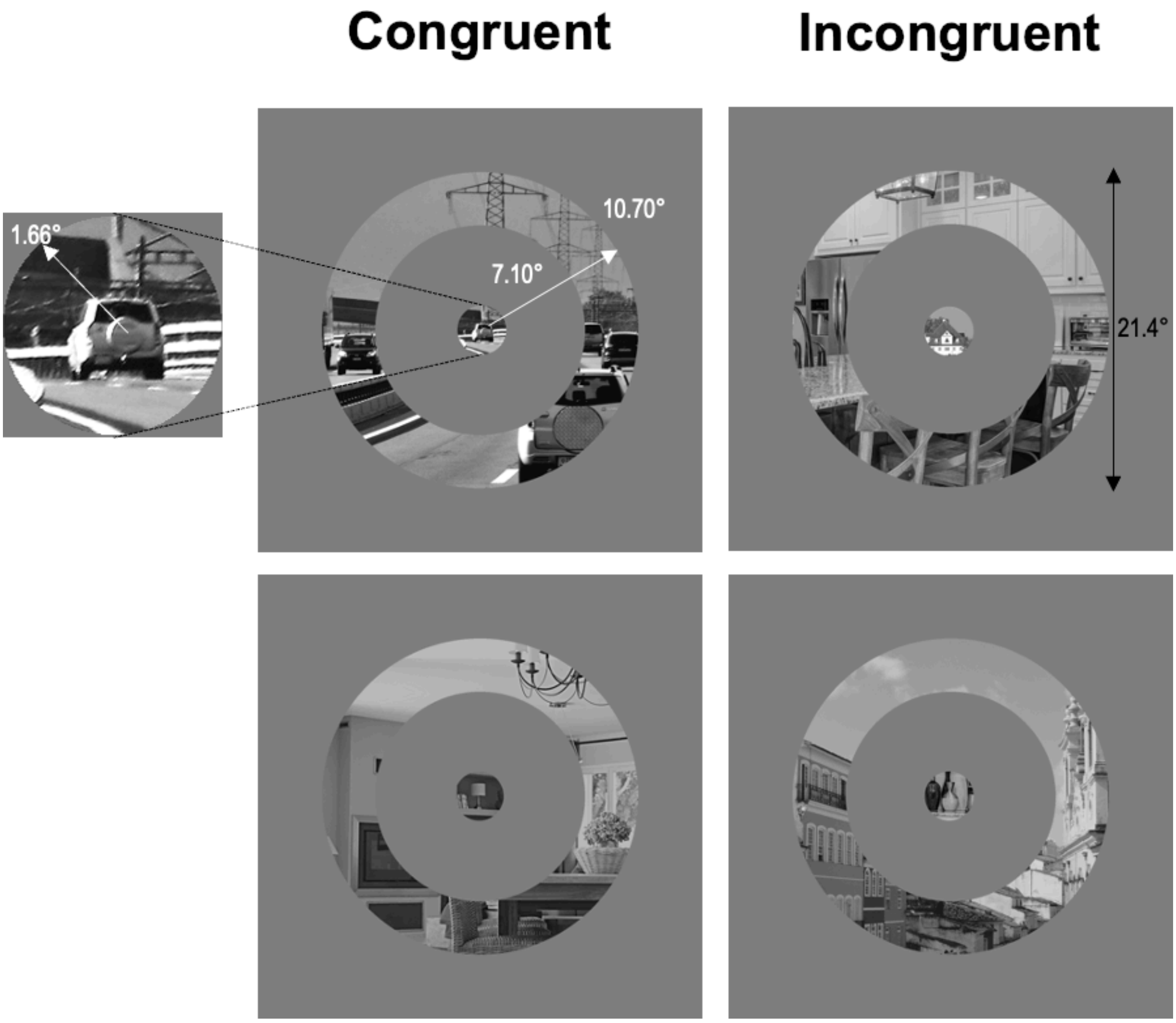

Figure 1. Example of stimuli used in Experiment 1. The central disk and the peripheral ring stimuli were presented simultaneously. They could be either semantically congruent (e.g., a central outdoor scene and the same peripheral outdoor scene) or semantically incongruent (e.g., a central outdoor scene and a peripheral indoor scene).

Procedure

Stimuli were displayed using the Psychtoolbox (Brainard, 1997; Kleiner, 2007; Pelli, 1997) implemented in MATLAB R2019 (MathWorks, Natick, MA, USA) on a 30′ monitor (DELL ULTRASHARP), with a resolution of 2560 × 1600 pixels and a refreshing rate of 60 Hz. The participants' heads were placed on a chinrest at 70 cm from the screen in order to respect the stimuli angular size. All participants performed two sessions. In the first session, participants had to categorize the central disk (the target) and ignore the peripheral ring (the distractor). In the second session, participants had to categorize the peripheral ring (the target) and ignore the central disk (the distractor). The two experimental sessions were counterbalanced between each participant. For each session, a trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed by both the central disk and peripheral ring for 100 ms on a 0.5 luminance background and then, by a gray screen (0.5 luminance) of 1900 ms during which participants could respond. Participants were asked to categorize the target (either the central disk or the peripheral ring) as an indoor or an outdoor scene. Participants were instructed to answer as correctly and as quickly as possible and to fixate on the center of the screen for the duration of the trial. They were instructed to press the corresponding response key with the middle finger and the forefinger of the right hand. Response keys were counterbalanced across participants. Each session included 160 trials (40 congruent stimuli from the Indoor category, 40 congruent stimuli from the Outdoor category and 80 incongruent stimuli) and lasted 45 min, with breaks every 80 trials. For each trial, response accuracy and response time (in ms) were recorded. Before each experimental session, participants performed a training session (25 trials for each session) using stimuli that were not included in the main experiment.

In order to ensure that participants fixated on the center of the screen for the duration of a trial, we recorded eye movements throughout the experiment. Eye movements were recorded using an Eyelink 1000 eye-tracker (SR Research) with a sampling rate of 1,000 Hz and a nominal spatial resolution of 0.01° of visual angle. For each participant, only the right eye was recorded using the "pupil-corneal reflection" mode. The Eyelink software automatically detected saccades with the following thresholds: speed >30°/s, acceleration >8,000°/s2, and saccadic displacement >0.15°. Fixations were detected when the pupil was visible and no saccade was in progress. Blinks were detected during partial or total occlusion of the pupil. Each session was preceded by a calibration procedure in which participants had to orient their gaze toward nine separate dots appearing sequentially in a 3 × 3 grid that occupied the entire display. A drift correction was performed every 10 trials. A new calibration was done in the middle of the experiment.

Data Analysis

Data were analyzed using R (R Core Team, 2019) and *lme4* package (Bates, Mächler, et al., 2015) to perform the analyses on the error rates (ER) and correct response times (RT). To analyze the ER, we set up a mixed effect logistic (the dependent variable being a binary variable) regression model. The participant's response was coded 1 when incorrect and 0 when correct. To analyze RT, we set up a linear mixed effects model of the Congruence (congruent vs. incongruent) and the Target position (central vs. peripheral) in a 2 × 2 within-subjects design. Congruence, Target position and their interaction term were entered into each model as fixed effects (i.e., as the effects of the variables of interest). Intercepts for subjects, as well as subject-wise random slopes for the Congruence effect, Target position and their interaction were specified as random effects (i.e., as the effects of the variables to which we want to generalize our results). Based on previous works using conflicting semantic stimuli (Lukavský, 2019; Roux-Sibilon et al., 2019; Peyrin et al., 2021), we expected to observe a semantic congruence effect regardless of the position of the Target (central vs. peripheral). Moreover, if information in peripheral vision is used to generate a predictive signal that guides visual categorization in central vision, the congruence effect should be greater for the categorization of the central target than the peripheral target. To test these hypotheses, we calculated the main effects of Congruence, Target position as well as their interaction for ER and RT.

We used the method proposed by Bates, Kliegl et al. (2015) to construct parsimonious mixed models preventing convergence problems. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained with a Wald test for ER analyses and by Satterthwaite approximation for RT analyses with the *lmerTest* package (Kuznetsova et al., 2017). The significance threshold was set at .05 for ER and RT analyses. Effect sizes were estimated using Cohen's d_z (Lakens, 2013).

Eye movements were used to control that the gaze position on each trial was maintained within 1.7° of eccentricity from the center of the screen (i.e. the radius of the central disk stimulus). For each participant and each experimental session, we removed trials in which saccades were initiated with an amplitude beyond 1.7° around the fixation cross. This resulted in removing 0.78% of the trials. Data from one participant who shifted his gaze beyond 1.7° of eccentricity from the center of the screen in a majority of trials were also discarded from the analyses. The statistical analyses were therefore conducted on 23 participants (18 women, mean age: 20.77 ± 2.43).

Data are available in the Open Science Framework repository: https://osf.io/q3rtw/.

Results

Results are shown in Figure 2. The analyses performed on ER revealed a significant main effect of the Congruence, $\beta = 0.67$, $z = 5.49$, $p < .001$. Participants made more errors when the two scenes were semantically incongruent (mean ± standard error: 10.10 ± 6.28%) than congruent (5.22 ± 4.64%). This indicated that when the two scenes were semantically incongruent, the scene to ignore was processed at the semantic level and influenced the categorization of the target (i.e., semantic congruence effect). We also observed a significant main effect of the Target position, $\beta = 0.35$, $z = 2.41$, $p = 0.016$. Specifically, participants made fewer errors for categorizing the central disk (6.09 ± 4.99%) than the peripheral ring (9.24 ± 6.04%). However, the interaction between the Congruence and the Target position was not significant, $\beta = 0.10$, $z = 0.45$, $p = 0.656$, suggesting that the semantic congruence effect was as important for categorizing a target in central vision with a distractor in peripheral vision as for categorizing a target in peripheral vision with a distractor in central vision.

The analyses performed on RT revealed a significant main effect of the Congruence, $\beta = 35.25$, $t(6670.97) = 8.08$, $p < 0.001$, $d_z = 1.685$. Participants were slower when the two scenes were semantically incongruent (683 ± 44 ms) than congruent (648 ± 45 ms), suggesting a semantic congruence effect on response times. We also observed a significant main effect of the Target position, $\beta = 59.78$, $t(21.88) = 3.98$, $p < 0.001$, $d_z = 0.830$. Participants were faster for categorizing the central disk (637 ± 39 ms) than the peripheral ring (695 ± 49 ms). As for the ER analysis, the interaction between the Congruence and the Target position was not significant, $\beta = -5.25$, $t(22.06) = -0.368$, $p = 0.717$, $d_z = 0.077$, suggesting again that the semantic congruence effect was as important for categorizing a target in central vision with a distractor in peripheral vision as for categorizing a target in peripheral vision with a distractor in central vision.

To go further, we also analyzed the effect of the Category of the scene to categorize on ER and RT a posteriori. The analyses performed on ER revealed no significant main effect of the Category, $\beta = 0.14$, $z = 0.98$, $p = .329$, or interaction involving the Category (Category*Target position, $\beta = 0.05$, $z = 0.16$, $p = .871$; Category*Congruence, $\beta = -0.05$, $z = -0.20$, $p = .840$; Category*Target position*Congruence, $\beta = -0.82$, $z = -1.56$, $p = .118$). The analyses performed on RT revealed only a significant interaction between the Category and the Target position, $\beta = -0.04$, $t(21.73) = -2.52$, $p = .019$, $d_z = 0.525$ (main effect of the Category, $\beta = -0.01$, $t(22.02) = -1.01$, $p = .325$, $d_z = 0.210$; Category*Congruence, $\beta < -0.001$, $t(6628)$

= - 0.07, p = .948, d_z = 0.014; Category*Target position*Congruence, β = - 0.02, t(6627) = - 1.01, p = .313, d_z = 0.210). More specifically, in peripheral vision, participants were faster to categorize an outdoor than an indoor scene irrespective of the content of the central scene, β = - 0.03, t(21.93) = - 2.44, p = .023, d_z = 0.509, while there was no significant difference between the categorization of outdoor and indoor scenes in central vision, β = 0.01, t(21.77) = 1.20, p = .241, d_z = 0.251.

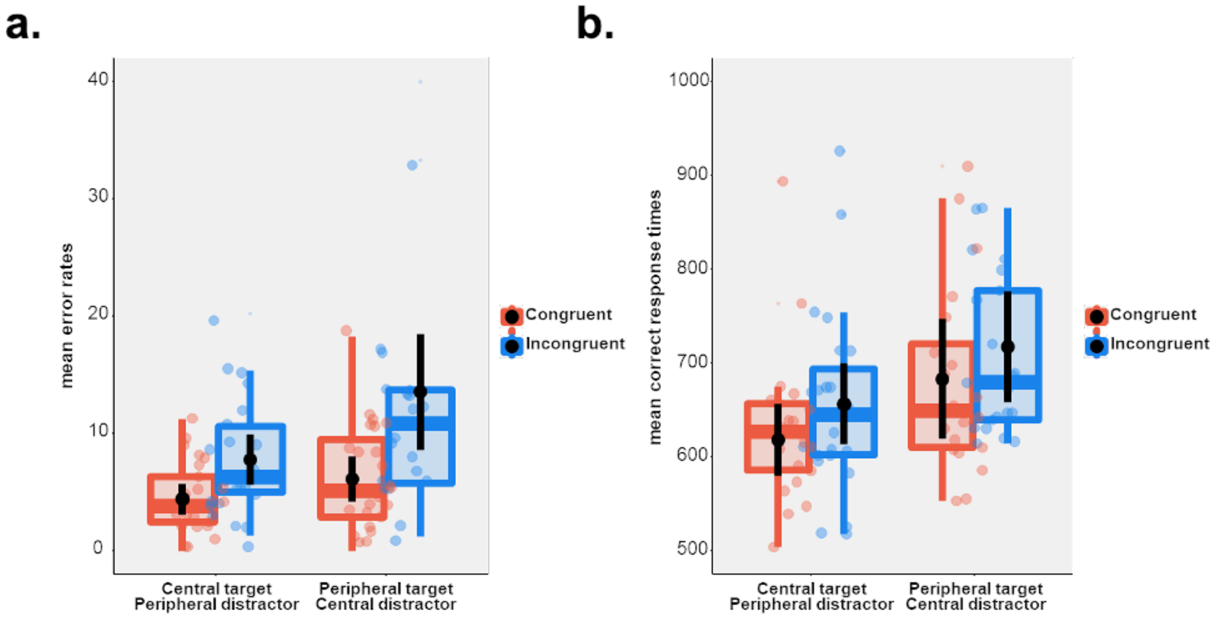

Figure 2. Box plots of (a) mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the central target and the peripheral target according to their congruence (congruent, incongruent). A box represents the median and quartiles and the whiskers represent the minimum and maximum sample without the outliers. Black dots and error bars indicate mean and 95% confidence intervals, respectively. Dots correspond to individual observations.

Discussion

The aim of this first experiment was to examine whether peripheral information could actually influence the categorization of scenes in central vision and *vice versa*. To this end, we used a paradigm in which a central disk and a peripheral ring depicting two scenes semantically congruent or incongruent were presented simultaneously. The task consisted in categorizing the scene contained in the central disk while ignoring the peripheral ring (i.e, peripheral distractor) or in categorizing the peripheral ring while ignoring the central disk (i.e., central distractor). Irrespective of the congruence manipulation, results revealed that the categorization of scenes in peripheral vision was faster for outdoor than indoor scenes (irrespective of the semantic content in central vision): This result was also observed in Trouilloud et al. (2020) in which participants had to categorize as indoor or outdoor different parts of a scene revealed through circular rings at different eccentricities. Participants were faster for categorizing outdoor than indoor scenes for the more peripheral stimulus. Outdoor scenes are usually well characterized by their global spatial properties while indoor scenes are better characterized by their objects (Quattoni & Torralba, 2009; even if indoor scenes could also be characterized by global spatial properties). The low spatial resolution of information available in peripheral vision could therefore benefit the categorization of outdoor scenes.

More importantly, results showed a semantic congruence effect. Namely, participants were faster and made fewer errors to categorize congruent than incongruent scenes. When the distractor does not belong to the same category as the target, this conflicting information disturbs the categorization of the target scene, so that the visual system is more likely to make more categorization errors and needs more time to categorize the scene. Although information from the distractor is not relevant to the task, it is automatically processed and influences the categorization of the target scene. It should be noted however that our experiment did not allow us to specify at which stage of the categorization process this influence occurs (perceptual, semantic, attentional, or decisional).

Based on our hypothesis that information in peripheral vision is used to generate a predictive signal that guides visual categorization in central vision, we expected the semantic congruence effect of the distractor to be larger for the categorization of the central disk than the peripheral ring. However, this congruence effect of peripheral information on central categorization was as important as the congruence effect of central information on peripheral categorization, in line with the Lukavský (2019)'s findings. This result will be addressed in the general discussion.

However, it should be noted that the task required the participants to keep focusing their gaze at the center of the screen throughout the experimental trial (from the presentation of the fixation point to their response), which may have added difficulty in ignoring this central information when we asked participants to categorize peripheral stimuli while ignoring central information. Besides, even if participants had no difficulties shifting their attention between the two experimental sessions (central disk target vs. peripheral ring target), they had better performances (both on error rates and response times) when categorizing the central disk with a peripheral distractor (irrespective of its congruence) than when categorizing the peripheral ring with a central distractor. This suggests that it is difficult to ignore the high-resolution information from central vision. Nevertheless, keeping the gaze fixed in the center of the screen is inherent to conditions for studying vision.

Overall, the results of this first experiment are consistent with previous findings of Lukavský (2019) showing mutual influence between central and peripheral vision during rapid scene categorization. High-resolution information from central vision cannot not be ignored and therefore influences peripheral information categorization. However, the influence of peripheral information on categorization in central vision appears to be as important as central influence despite the low quality inherent to peripheral vision. Although irrelevant for the task, peripheral information was also automatically processed and influenced the categorization of the central scene. To go further, we aimed at determining the nature of the physical features used in peripheral information to influence categorization in central vision. To this end, we conducted a second experiment (Experiment 2) using the same paradigm but we used either an intact peripheral ring, or a peripheral ring whose amplitude spectrum was preserved (phase-scrambling), or a peripheral ring whose phase spectrum was preserved (amplitude-scrambling). In this way, we were able to independently assess the contribution of the phase signal and that of the amplitude signal in the semantic congruence effect of peripheral vision.

Experiment 2

Methods

Participants

Twenty-two undergraduate students of Psychology from University Grenoble Alpes (21 women, mean age: 20.45 ± 1.71) participated in the experiment. The sample size was chosen based on a power analysis with estimated effect size of $d_z = 1.296$ (effect size of the semantic congruence effect for a central categorization in the RT analysis of Experiment 1) to achieve power of 0.99 at an alpha level of 0.05. All were right-handed, with normal or corrected-to-normal vision. The study was performed within the same ethical framework as Experiment 1.

Stimuli and procedure

The central disk and peripheral ring stimuli were built from the same stimuli as those used in Experiment 1. In this experiment, we manipulated the features available in the peripheral ring stimuli which could be either Intact, Phase-preserved (amplitude-scrambled) or Amplitude-preserved (phase-scrambled). Phase-preserved and Amplitude-preserved stimuli were created as follows. First, we calculated the Fourier transform of the scene to obtain its amplitude spectrum and its phase spectrum. We also created an image of the same size with white noise and computed its Fourier transform in order to obtain an amplitude spectrum and a phase spectrum of a white noise. To create the Phase-preserved stimuli, we then performed an inverse Fourier transform using the phase spectrum of the scene and the amplitude spectrum of the white noise. To create the Amplitude-preserved stimuli, the inverse Fourier transform was done using the amplitude spectrum of the scene and the phase spectrum of the white noise. All images brought back into the spatial domain were then equalized in terms of mean luminance (0.5) and RMS contrast (0.2) as in Experiment 1. We then built a third peripheral ring by following the procedure described in Experiment 1.

As in Experiment 1, the central disk and the peripheral ring stimuli were presented simultaneously (Figure 3). They could be either semantically congruent (e.g., a central indoor scene and the same peripheral indoor scene) or semantically incongruent (e.g., a central indoor scene and a peripheral outdoor scene). The information contained in the peripheral ring stimulus could be either Intact, Phase-preserved or Amplitude-preserved.

Stimuli were displayed using the Psychtoolbox implemented in MATLAB R2019 (MathWorks, Natick, MA, USA) using the same 30′ monitor as in Experiment 1 with a viewing distance of 70 cm in order to respect the stimuli angular size. All participants had to categorize the central scene and ignore the peripheral scene. A trial began with a central black fixation point presented for 500 ms on a 0.5 luminance background, followed by the stimulus for 100 ms on a 0.5 luminance background and then, by a gray screen (0.5 luminance) for 1900 ms during which participants could respond. Participants' task was similar to Experiment 1 (i.e., indoor vs. outdoor categorization). They were instructed to answer as correctly and as quickly as possible at the onset of stimuli and to fixate on the center of the screen for the duration of the experiment. Central fixation was monitored using the same Eyelink 1000 eye-tracker (SR Research) as in Experiment 1. They were instructed to press the corresponding response key with the middle finger and the forefinger of the right hand. Response keys were counterbalanced across participants. The experiment included 240 trials (80 intact peripheral rings including 40 congruent and 40 incongruent stimuli; 80 phase-preserved peripheral rings including 40 congruent and 40 incongruent stimuli; 80 amplitude-preserved peripheral rings including 40 congruent and 40 incongruent stimuli) and lasted 30 minutes, with a break in the middle of the experiment. Trials were fully randomized. For each trial, response accuracy and response time (in ms) were recorded. Before the experiment, participants performed a training session (25 trials) using stimuli that were not included in the main experiment.

Stimuli can be downloaded from https://osf.io/q3rtw/.

Data Analysis

Data were analyzed using R (R Core Team, 2019) and *lme4* package (Bates, Mächler, et al., 2015) to perform the analyses on the error rates (ER) and correct response times (RT). To analyze the ER, we set up a mixed effects logistic regression model. The participant's response was coded 1 when incorrect and 0 when correct. To analyze RT, we set up a linear mixed effects model of the Congruence (congruent vs. incongruent) and the Physical features of peripheral information (intact vs. amplitude-preserved vs. phase-preserved) in a 2 × 3 within-subjects design. Congruence, Physical features and their interaction term were entered into each model as fixed effects. Intercepts for subjects, as well as subject-wise random slopes for the effect of Congruence, Physical features and their interaction were specified as random effects. As in Experiment 1, we expected to observe a semantic congruence effect for the intact scene. Specifically, we aimed to determine which type of peripheral information contributed to influence categorization in central vision. Basically, the contribution of the physical features contained in the amplitude spectrum should result in a semantic congruence effect for peripheral rings whose amplitude spectrum was preserved. Similarly, the contribution of the physical features contained in the phase spectrum should result in a semantic congruence effect for peripheral rings whose phase spectrum was preserved. If both physical information are used in peripheral vision, we should observe congruence effects in both conditions. To test these hypotheses, we calculated whether the congruence effects were significant in each condition of Physical features for ER and RT.

We used the method proposed by Bates, Kliegl, et al. (2015) to construct parsimonious mixed models preventing convergence problems. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality. P-values were obtained with a Wald test for ER analyses and by Satterthwaite approximation for RT analyses with the *lmerTest* package (Kuznetsova et al., 2017). The significance threshold was set at .017 after Bonferroni correction for three non-orthogonal tests (.05/3 = .017) for ER and RT analyses. Effect sizes were estimated using Cohen's d_z (Lakens, 2013). Trials in which saccades were initiated from more than 1.7° around the fixation cross (i.e., radius of the central disk) were discarded from the analyses. This resulted in removing 0.08% of the trials.

Data are available in the Open Science Framework repository: https://osf.io/q3rtw/.

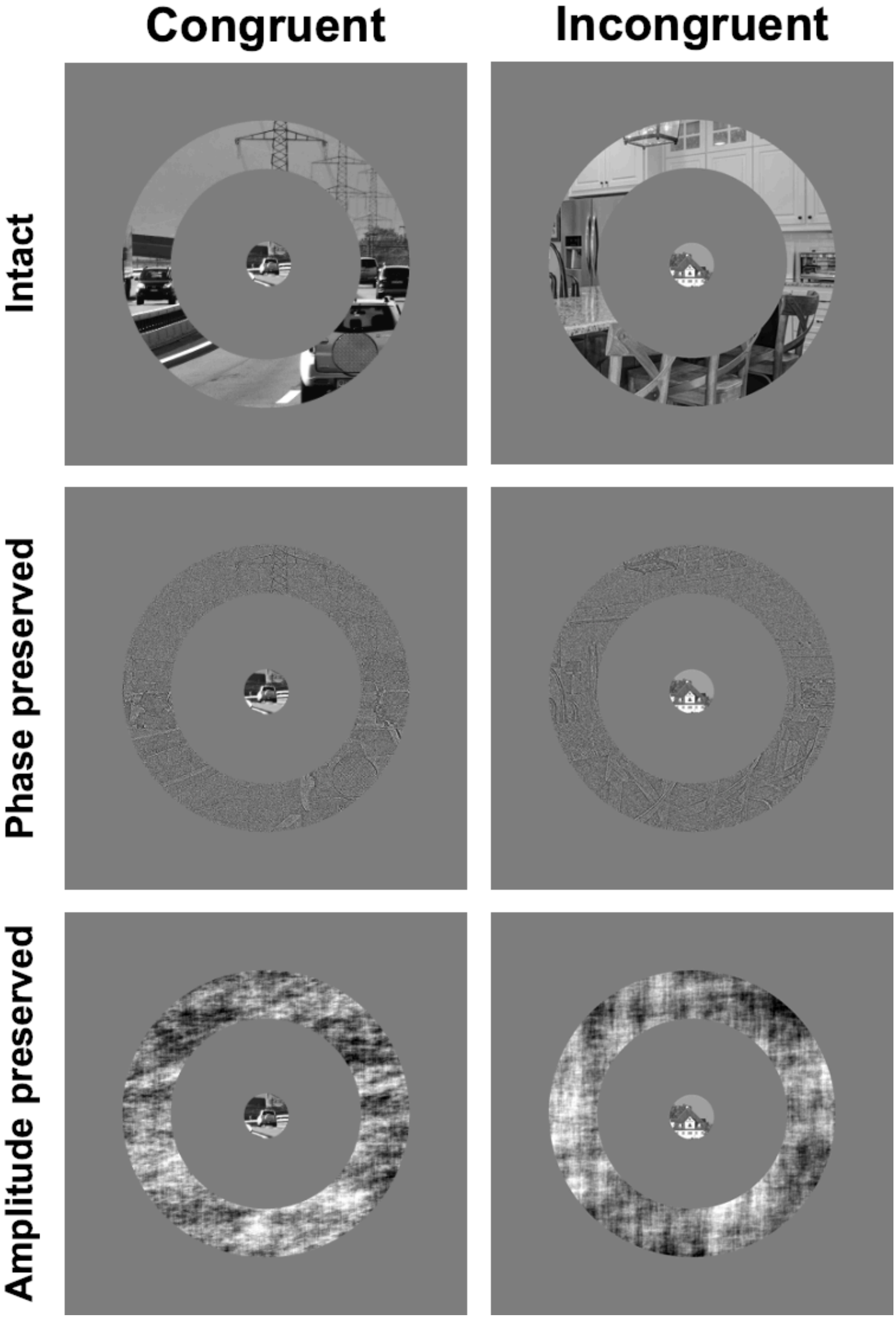

Figure 3. Example of stimuli used in Experiment 2. The central disk and the peripheral ring stimuli were presented simultaneously. They could be either semantically congruent (e.g., a central outdoor scene and the same peripheral outdoor scene) or semantically incongruent (e.g., a central outdoor scene and a peripheral indoor scene). The information contained in the peripheral ring stimulus could be either intact, phase-preserved (i.e., spatial configuration of the scene was preserved) or amplitude-preserved (i.e., distribution of information across spatial frequencies and orientations was preserved).

Results

Results are shown in Figure 4. The analyses performed on ER revealed neither a significant congruence effect for the intact scene, $\beta = 0.70$, $z = 2.38$, $p = .0171$, nor a significant congruence effect for the phase-preserved scene, $\beta = 0.09$, $z = 0.27$, $p = .784$, nor a significant congruence effect for the amplitude-preserved scene, $\beta = -0.25$, $z = -1$, $p = .317$.

The analyses performed on RT revealed a significant congruence effect for the intact scene, $\beta = 30.73$, $t(4981.13) = 4.14$, $p < 0.001$, $d_z = 0.883$, with participants being slower when the two scenes were semantically incongruent (665 ± 35 ms) than congruent (633 ± 34 ms). We also observed a significant congruence effect for the phase-preserved scene, $\beta = 19.05$, $t(4981.03) = 2.60$, $p = .009$, $d_z = 0.554$, with participants being slower when the two scenes were semantically incongruent (643 ± 37 ms) than congruent (623 ± 33 ms). However, the congruence effect was not significant for the amplitude-preserved scene, $\beta = 14.18$, $t(4981.04) = 1.93$, $p = .054$, $d_z = 0.411$.

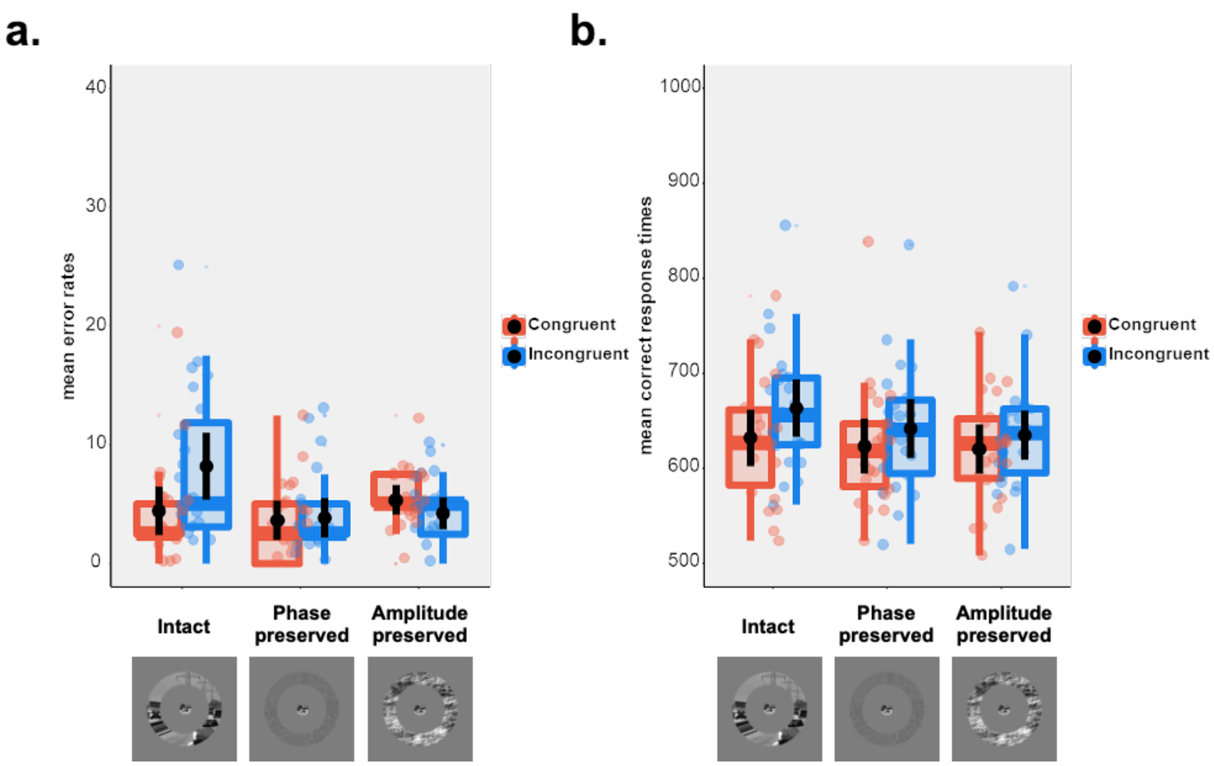

Figure 4. Box plots of (a) mean error rates in percentage and (b) mean correct response times in milliseconds during the categorization of the central disk with a peripheral ring containing an intact scene, an amplitude-preserved scene and a phase-preserved scene according to the congruence between the central disk and the peripheral ring (congruent, incongruent). A box represents the median and quartiles and the whiskers represent the minimum and maximum sample without the outliers. Black dots and error bars indicate mean and 95% confidence intervals, respectively. Dots are individual observations.

We performed post-hoc mixed models analyses on RT to examine whether the congruence effect was actually larger in one of the conditions. We calculated as fixed effect the interaction between the Congruence (congruent vs. incongruent) and the Physical features (intact vs. amplitude-preserved or phase-preserved vs. amplitude-preserved or intact vs. phase-preserved) on the RT. However, these post-hoc analyses failed to demonstrate neither a stronger congruence effect for intact than phase-preserved scenes, $\beta = 11.90$, $t(3311.09) = 1.13$, $p = .257$, $d_z = 0.241$, nor a stronger congruence effect for intact than amplitude-preserved scenes, $\beta = 16.43$, $t(3\ 293.21) = 1.59$, $p = .113$, $d_z = 0.339$, nor a stronger congruence effect for phase-preserved than amplitude-preserved scenes, $\beta = 4.90$, $t(3337.08) = 0.47$, $p = .637$, $d_z = 0.100$.

General discussion

Past studies suggest that coarse information from peripheral vision plays a critical role during scene categorization (Larson & Loschky, 2009; Loschky et al., 2019; Roux-Sibilon et al., 2019; Trouilloud et al., 2020) and would be useful to generate predictions which would then influence and guide the categorization of detailed information from central vision (Bar, 2003; Bar, 2007; Peyrin et al., 2021; Roux-Sibilon et al., 2019). The aim of the present study was therefore to determine (1) whether peripheral information influences the categorization of scenes in central vision and (2) which type of physical features from peripheral vision actually guides categorization in central vision. To this end, we used a paradigm in which we presented simultaneously a central disk and a peripheral ring, both containing scenes belonging to the same category (congruent condition) or not (incongruent condition).

In Experiment 1, participants had to categorize as an indoor or an outdoor scene either the scene contained in the central disk while ignoring the peripheral ring or the scene contained in the peripheral ring while ignoring the central disk. Participants had to fixate the center of the screen. Results showed a semantic congruence effect indicating that the information to be ignored was however processed and influenced the categorization of the information to attend. As in the study of Lukavský (2019), this effect was present for both central and peripheral categorization despite our control of the cortical magnification factor so that the central disk and the peripheral ring activated the same amount of neurons in the primary visual cortex.

Importantly for our research issue, results demonstrated that despite its low quality, information available in peripheral vision influences and guides the categorization of the information that the participants fixated in central vision. This result could be first interpreted in the context of the predictive models of visual recognition (Bar, 2003, 2007; Trapp & Bar, 2015; Kauffmann et al., 2014; Kveraga et al., 2007; Peyrin et al., 2010). According to these models, a coarse information is sufficient and useful to trigger cortical predictive mechanisms that would then guide a more detailed visual analysis. Roux-Sibilon et al. (2019) demonstrated that peripheral vision may be used to generate predictions that guide the processing of central information. Similarly, in our experiment, when the peripheral scene to ignore belongs to the same category as the central scene, it may be used to generate valid predictions about the category of the central scene which would then facilitate its categorization. On the other hand, when the peripheral scene belongs to a different category than the central scene, its rapid processing would result in the generation of invalid predictions, impairing the central categorization and resulting in lower categorization performances. It should however be noted that in our experiment, peripheral stimuli were presented close to the limits of central vision. Indeed, the angular size of the scotoma in the peripheral ring is 14.20° in diameter, the visual information was displayed just beyond the parafovea. It is therefore possible that the peripheral congruence effect could be weaker in the far periphery than close periphery used in the present study. Future studies might be relevant to investigate this effect at larger eccentricities using natural real-world scenes.

In more ecological viewing conditions, the rapid processing of peripheral information may also be used to generate predictions about the visual information present at the next fixation. Herwig and Schneider (2014) showed that peripheral pre-saccadic input is used to generate predictions about the appearance of the foveal post-saccadic input. In their study, participants were presented with two peripheral objects (a triangular or circular sinusoidal grating) and had to initiate a saccade toward one of them so that both objects were saccaded to in an equal amount of trials. During the execution of the saccade, the spatial frequency of the peripheral object could remain intact (normal condition) or change (swapped condition) depending on the object that was targeted so that participants implicitly learnt associations between the peripheral and foveal appearance of each object. In a subsequent "test" phase, one of the two peripheral objects was presented and participants had to initiate a saccade toward it. The object disappeared during the saccade and was replaced by a new object with a random spatial frequency. Participants had to adjust the spatial frequency of the object (which was now at fovea) in order to match the expected spatial frequency of the object they saw peripherally before initiating their saccade. Results revealed participants' judgments were influenced by the acquisition phase. Peripheral objects with the same spatial frequency were expected to have a different spatial frequency when foveated depending on the association learnt during the acquisition phase. This therefore indicated that associations between pre- and post-saccadic appearance of visual stimuli are used to predict object features across saccades. Several studies have also shown that peripheral information is used before a change in gaze location, allowing to initiate a saccadic movement towards a relevant area of the scene (Cajar et al., 2020; Nuthmann, 2014). Indeed, Nuthmann (2014) examined the importance of central and peripheral vision during object search in a scene by restricting the visual field to its central part using gaze-contingent spotlights of different eccentricities (i.e., masking either peripheral or both peripheral and parafoveal vision) or its peripheral parts using blindspots of different eccentricities (i.e., masking either foveal or both foveal and parafoveal vision). Results showed that when participants searched the object with only high resolution information from foveal vision, their accuracy and search times were significantly impaired relative to the control condition where all scene information was available. However, when peripheral information was available and foveal information was hidden, their search performances were as good as when all the image content was displayed. Interestingly, the authors decomposed the scan pattern of participants to identify which part of the visual field was useful to (1) localize and (2) identify the object. In the blindspots condition (i.e., masking central parts), participants were not impaired to localize the object but were slower to identify it. In the spotlights condition (i.e., masking peripheral parts), the opposite results were found. Participants were not impaired to identify the object but they needed more time to first localize it. These findings therefore support the critical utility of peripheral vision relative to central vision to generate saccades towards a relevant part of the scene and localize objects of interest. Overall, these findings highlight the crucial role of peripheral vision in updating the gaze location to localize an object and predict its features before its identification thanks to the high resolution of central vision. Despite its low quality, peripheral vision therefore appears essential to extract enough information to quickly categorize objects or scenes and guide saccadic eye movements to bring areas of interest into central vision, allowing a more detailed processing.

In this context, we formulated the hypothesis that the semantic congruence effect of peripheral vision would be stronger than the semantic congruence effect of central vision. We observed semantic congruence effects of both peripheral and central vision suggesting that peripheral and central vision are actually processed interactively (Davenport & Potter, 2004), but we failed to observe a significant interaction between the congruence and the spatial position of the information to be ignored. This result can be interpreted in the context of the evidence accumulation model postulated by Lukavsky (2019) in which information from the to-be-ignored parts is also included in the scene recognition process. Conflicting information in the to-be-ignored part, central or peripheral, increases uncertainty and more time is required to reach the decision boundary in both experimental conditions. Given this interpretation, the congruence effects observed in the present experiment do not rely on predictive mechanisms. It should be noted, however, that our experimental instruction induced attentional constraint to the task in favor of central vision that would not be compatible with natural vision habits. As we previously pointed out, our task required participants to fixate the center of the screen to avoid saccadic movements throughout the experiment (for both central and peripheral categorization). While our attention is distributed over the entire visual field in normal viewing conditions, participants here were submitted to focused attention in central vision (even when this part of the visual field had to be ignored). Brand and Johnson (2018) investigated whether scene categorization required distributed attention over the entire visual field or whether focused attention in central vision was sufficient. They presented a large scene and they superimposed two rectangles, one at the center of the screen and the other surrounding the periphery of the scene. The task consisted in categorizing the scene as natural or man-made and then in indicating the orientation of either the central rectangle (i.e., focused attention) or the peripheral one (i.e., distributed attention). The results revealed better scene categorization performances in the distributed than focused attention condition, suggesting that distributed attention facilitated scene categorization at the superordinate level. In other words, attention would be naturally distributed on the whole visual field when perceiving a scene, and information from the whole visual field would be used to categorize it at a superordinate level (as in our experiment). It is thus possible that our task involves a central attentional bias which increases the weight of the semantic congruence effect from the center on the peripheral categorization and conversely, underestimates the weight of the influence from the periphery on the central categorization. In addition, central information is intrinsically the one with the highest spatial resolution and previous studies revealed that it is more difficult to ignore HSF than LSF in central vision (Kauffmann et al., 2017). Besides, irrespective of the congruence effect, we observed that participants were faster and categorized the central disk more accurately than the peripheral ring. This result contradicts those of Trouilloud et al. (2020) using exactly the same stimuli, but presented in isolation (i.e. central disk alone, peripheral ring alone) and showing that the categorization of the peripheral ring was faster than the categorization of the central disk. Even if the two experiments differ objectively on the visual display (one versus two images), the discrepancies further question an attentional bias. Future studies therefore need to ensure that attention is distributed over the whole visual field in order to ensure that (1) the absence of significant interaction between the congruence and the spatial position of the information to be ignored, and (2) the prioritization of central vision observed in the present experiment is not explained by methodological problems.

To go further, we aimed at determining the nature of physical features from peripheral vision which influence the categorization of scenes in central vision. To this end, we performed a second experiment (Experiment 2) in which participants were only requested to categorize the central disk while ignoring the peripheral ring and we manipulated the physical features of the peripheral ring. The scene contained in the peripheral ring was therefore either intact (i.e., containing semantic information), only its amplitude spectrum was preserved (amplitude-preserved, i.e. containing only the distribution of information across spatial frequencies and orientations), or only its phase spectrum was preserved (phase-preserved, i.e. containing only information about the spatial structure of the scene). If information from the phase spectrum influences categorization in central vision, we expected a congruence effect in the phase-preserved condition. If information from the amplitude spectrum influences categorization, we expected a semantic congruence effect in the amplitude-preserved condition. As in Experiment 1, we observed the semantic congruence effect on the response times when the peripheral scene was intact. Response times were longer when the peripheral scene was incongruent than congruent with the central scene, suggesting again that the peripheral input, even irrelevant, could not be ignored and was integrated to the categorization processes in central information. Unlike Experiment 1, we cannot however conclude on error rates, the congruence effect being not significant after the Bonferroni correction. It should however be noted that the very low error rates in Experiment 2 (below 8%) suggest a ceiling effect. The categorization of a central disk may be too easy to detect the expected effect on error rates. Importantly, the semantic congruence effect of peripheral rings was also observed when the phase information of the peripheral ring was preserved, but not when only the amplitude information was preserved. The incongruent information contained in the phase spectrum, unlike amplitude spectrum information, delayed the response as much as the information contained in the intact scene. In the same way, Roux-Sibilon et al. (2019) observed that the semantic congruence effect of a scene background on the categorization of an object in central vision disappeared when the amplitude spectrum of the scene background was preserved, but the spatial configuration was altered (phase scrambling of scene images). Altogether, these results suggest that semantic and spatially-arranged features contained in the phase spectrum of a scene in peripheral vision are primordial to guide the categorization of information in central vision. It should be noted that the absence of congruence effect for amplitude-preserved peripheral scenes does not necessarily imply that amplitude information is not used during scene categorization. Indeed, many studies have shown that although this information alone may not be sufficient (Loschky & Larson, 2008; Loschky et al., 2007) statistical regularities from the amplitude spectrum of scenes from different categories can be exploited for rapid and coarse scene categorization (Field, 1987; Guyader et al., 2004; Oliva & Torralba, 2001; Torralba & Oliva, 2003). However, our results suggest that such information may be less relevant than phase information to guide the categorization in central vision.

To conclude, these two experiments show mutual influence between central and peripheral information during rapid scene categorization. Despite the poor quality of the peripheral content, the influence of peripheral information on central categorization is as important as the influence of detailed information from central vision which could not be ignored. Although irrelevant for the task, peripheral vision would be automatically processed and would influence the categorization of central information, arguing in favor of the predictive role of peripheral vision to guide the analysis of central vision content. More importantly, these predictions would be based on the spatial configuration of the scene contained in the phase spectrum. Unfortunately, the present study did not allow us to clearly demonstrate the role of peripheral vision in a proactive model. As previously performed by Roux-Sibilon et al. (2019), future studies could manipulate the time during which peripheral information is processed before central information onset in order to enhance predictive processes. An increasing congruence effect with SOA would support a predictive model rather than the simultaneous processing of central and peripheral information.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by NeuroCoG IDEX UGA in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

References

Bar, M. (2003). A Cortical Mechanism for Triggering Top-Down Facilitation in Visual Object Recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600-609. https://doi.org/10.1162/089892903321662976

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280-289. https://doi.org/10.1016/j.tics.2007.05.005

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449-454. https://doi.org/10.1073/pnas.0507062103

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

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

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*(4), 433-436.
Brand, J., & Johnson, A. P. (2018). The effects of distributed and focused attention on rapid scene categorization. *Visual Cognition*, *26*(6), 450-462. https://doi.org/10.1080/13506285.2018.1485808
Cajar, A., Engbert, R., & Laubrock, J. (2020). How spatial frequencies and color drive object search in real-world scenes: A new eye-movement corpus. *Journal of Vision*, *20*(7), 8. https://doi.org/10.1167/jov.20.7.8
Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of comparative Neurology*, *300*(1), 5-25.
Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *Journal of comparative neurology*, *292*(4), 497-523.
Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology*, *159*(2), 203-221.
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
Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, *4*(12), 2379-2394.
Geuzebroek, A. C., & van den Berg, A. V. (2018). Eccentricity scale independence for scene perception in the first tens of milliseconds. *Journal of Vision*, *18*(9), 9. https://doi.org/10.1167/18.9.9
Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535-574. https://doi.org/10.1146/annurev.neuro.29.051605.113038
Greenwood, J. A., Tailor, V. K., Sloper, J. J., Simmers, A. J., Bex, P. J., & Dakin, S. C. (2012). Visual acuity, crowding and stereo-vision are linked in children with and without amblyopia. Investigative Ophthalmology & Visual Science, 53(12), 7655–7665.
Guyader, N., Chauvin, A., Peyrin, C., Hérault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, *327*(4), 313-318.
Herwig, A., & Schneider, W. X. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. *Journal of Experimental Psychology: General*, *143*(5), 1903-1922. https://doi.org/10.1037/a0036781
Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *Journal of Vision*, *8*(13), 11-11. https://doi.org/10.1167/8.13.11
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

Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. Journal of vision, 12(13):16. https://doi.org/10.1167/12.13.16

Kauffmann, L., Bourgin, J., Guyader, N., & Peyrin, C. (2015). The Neural Bases of the Semantic Interference of Spatial Frequency-based Information in Scenes. *Journal of Cognitive Neuroscience*, *27*(12), 2394-2405. https://doi.org/10.1162/jocn_a_00861

Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain and Cognition*, *99*, 46-56. https://doi.org/10.1016/j.bandc.2015.07.004

Kauffmann, L., RamanoÃ«l, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*. https://doi.org/10.3389/fnint.2014.00037

Kauffmann, L., Roux-Sibilon, A., Beffara, B., Mermillod, M., Guyader, N., & Peyrin, C. (2017). How does information from low and high spatial frequencies interact during scene categorization? *Visual Cognition*, *25*(9-10), 853-867. https://doi.org/10.1080/13506285.2017.1347590

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception, 36*(14), 1–16.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of statistical software*, *82*(1), 1-26.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*(48), 13232-13240.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Larson, A. M., Freeman, T. E., Ringer, R. V., & Loschky, L. C. (2014). The spatiotemporal dynamics of scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 471-487. https://doi.org/10.1037/a0034986

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*(10), 6-6. https://doi.org/10.1167/9.10.6

Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision*, *8*(1), 4-4.

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made before basic-level distinctions in scene gist processing. *Visual Cognition*, *18*(4), 513-536.

Loschky, L. C., Nuthmann, A., Fortenbaugh, F. C., & Levi, D. M. (2017). Scene perception from central to peripheral vision. *Journal of Vision*, *17*(1), 6. https://doi.org/10.1167/17.1.6

Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 33(6), 1431.

Loschky, L. C., Szaffarczyk, S., Beugnet, C., Young, M. E., & Boucart, M. (2019). The contributions of central and peripheral vision to scene- gist recognition with a 1808 visual field. *Journal of Vision*, 21.

Lukavský, J. (2019). Scene categorization in the presence of a distractor. *Journal of Vision*, *19*(2), 6. https://doi.org/10.1167/19.2.6

Mirolli, M., Ferrauto, T., & Nolfi, S. (2010). Categorisation through evidence accumulation in an active vision system. *Connection Science*, *22*(4), 331-354. https://doi.org/10.1080/09540091.2010.505976

Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 342.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, *42*(3), 145-175.

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. Current opinion in neurobiology, 18(4), 445-451.

Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437-442.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., & Vuilleumier, P. (2010). The Neural Substrates and Timing of Top–Down Processes during Coarse-to-Fine Categorization of Visual Scenes: A Combined fMRI and ERP Study. *Journal of Cognitive Neuroscience*, *22*(12), 2768-2780. https://doi.org/10.1162/jocn.2010.21424

Peyrin, C., Roux-Sibilon, A., Trouilloud, A., Khazaz, S., Joly, M., Pichat, C., Boucart, M., Krainik, A., & Kauffmann, L. (2021). Semantic and Physical Properties of Peripheral Vision Are Used for Scene Categorization in Central Vision. *Journal of Cognitive Neuroscience*, *33*(5), 799-813. https://doi.org/10.1162/jocn_a_01689

Quattoni, A., & Torralba, A. (2009). Recognizing indoor scenes. IEEE Conference on Computer Vision and Pattern Recognition, 413-420.

Rousselet, G., Joubert, O., & Fabre-Thorpe, M. (2005). How long to get to the “gist” of real-world natural scenes? *Visual Cognition*, *12*(6), 852-877.

Roux-Sibilon, A., Trouilloud, A., Kauffmann, L., Guyader, N., Mermillod, M., & Peyrin, C. (2019). Influence of peripheral vision on object categorization in central vision. *Journal of Vision*, *19*(14), 7. https://doi.org/10.1167/19.14.7

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*(3), 391-412. https://doi.org/10.1088/0954-898X/14/3/302

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition: Predictions and competition. *Annals of the New York Academy of Sciences*, *1339*(1), 190-198. https://doi.org/10.1111/nyas.12680

Trouilloud, A., Kauffmann, L., Roux-Sibilon, A., Rossel, P., Boucart, M., Mermillod, M., & Peyrin, C. (2020). Rapid scene categorization: From coarse peripheral vision to fine central vision. *Vision Research*, *170*, 60-72. https://doi.org/10.1016/j.visres.2020.02.008

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. Trends in cognitive sciences, 15(4), 160-168.

Wu, J., Yan, T., Zhang, Z., Jin, F., & Guo, Q. (2012). Retinotopic mapping of the peripheral visual field to human visual cortex by functional magnetic resonance imaging. *Human brain mapping*, *33*(7), 1727-1740.

Yamada, E. (1969). Some structural features of the fovea centralis in the human retina. *Archives of ophthalmology*, *82*(2), 151-159.