

Semantic and Physical Properties of Peripheral Vision Are Used for Scene Categorization in Central Vision

Carole Peyrin, Alexia Roux-Sibilon, Audrey Trouilloud, Sarah Khazaz, Malena Joly, Cédric Pichat, Muriel Boucart, Alexandre Krainik, Louise Kauffmann

▶ **To cite this version:**

Carole Peyrin, Alexia Roux-Sibilon, Audrey Trouilloud, Sarah Khazaz, Malena Joly, et al.. Semantic and Physical Properties of Peripheral Vision Are Used for Scene Categorization in Central Vision. Journal of Cognitive Neuroscience, 2021, pp.1-15. 10.1162/jocn_a_01689 . hal-03139604

HAL Id: hal-03139604
https://hal.science/hal-03139604

Submitted on 12 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Semantic and Physical Properties of Peripheral Vision Are Used for Scene Categorization in Central Vision

Carole Peyrin[1], Alexia Roux-Sibilon[1], Audrey Trouilloud[1], Sarah Khazaz[1], Malena Joly[1], Cédric Pichat[1], Muriel Boucart[2], Alexandre Krainik[3], and Louise Kauffmann[1]

Abstract

■ Theories of visual recognition postulate that our ability to understand our visual environment at a glance is based on the extraction of the gist of the visual scene, a first global and rudimentary visual representation. Gist perception would be based on the rapid analysis of low spatial frequencies in the visual signal and would allow a coarse categorization of the scene. We aimed to study whether the low spatial resolution information available in peripheral vision could modulate the processing of visual information presented in central vision. We combined behavioral measures (Experiments 1 and 2) and fMRI measures (Experiment 2). Participants categorized a scene presented in central vision (artificial vs. natural categories) while ignoring another scene, either semantically congruent or incongruent, presented in peripheral vision. The two scenes could either share the same physical properties (similar amplitude spectrum and spatial configuration) or not. Categorization of the central scene was impaired by a semantically incongruent peripheral scene, in particular when the two scenes were physically similar. This semantic interference effect was associated with increased activation of the inferior frontal gyrus. When the two scenes were semantically congruent, the dissimilarity of their physical properties impaired the categorization of the central scene. This effect was associated with increased activation in occipito-temporal areas. In line with the hypothesis of predictive mechanisms involved in visual recognition, results suggest that semantic and physical properties of the information coming from peripheral vision would be automatically used to generate predictions that guide the processing of signal in central vision. ■

Uncorrected Proof

INTRODUCTION

We understand most of our visual environment at a glance, regardless of its complexity. This ability would be based on the rapid extraction of the gist from visual scenes, a first global and rudimentary visual representation (Oliva, 2005). At a conceptual level, the gist refers to the semantic content that can emerge quickly when a visual scene is perceived. At the perceptual level, it refers to the structural representation of the scene. In other words, the gist describes the structure, or spatial configuration, of the scene (i.e., the presence of different elements and their spatial relations). For example, the spatial configuration of a beach scene can be described by arranging three large horizontal bands, from top to bottom: one for the sky, one thinner for the sea, and one last for the sand. Access to a representation of the spatial configuration of the scene requires a coarse, large-scale spatial analysis but does not require the analysis of details contained in the different part of the scene. This information extracted at a large spatial scale is contained in the low spatial frequencies (LSF) of the visual signal, that is, the wide variations of signal in luminance. On the contrary, the details or the precise contours of the objects in the scene, which allow a finer analysis, are contained in the high spatial frequencies (HSF) of the visual signal. Thus, the extraction of the gist would be prominently based on the rapid analysis of the LSF of the visual signal, which would allow a coarse categorization of the scene and the objects that compose it.

This assumption is supported by neurophysiological recordings in primates that have demonstrated the selectivity of cells of the visual system to different spatial frequencies (Skottun et al., 1991; De Valois, Albrecht, & Thorell, 1982), as well as the faster propagation of LSF through the magnocellular pathway than HSF through the parvocellular pathway (Nowak & Bullier, 1997; Nowak, Munk, Girard, & Bullier, 1995). The anatomical and functional underpinnings of spatial frequency processing have therefore led to the hypothesis of a coarse-to-fine (CtF) analysis of visual information in scenes. The rapid processing of LSF would allow a first categorization that would then be refined, validated, or invalidated by the progressive accumulation of information in HSF. This predominant CtF scheme during scene perception was evidenced in several behavioral studies (Kauffmann, Chauvin, Guyader, & Peyrin, 2015; Musel, Chauvin, Guyader, Chokron, & Peyrin, 2012; Schyns & Oliva, 1994; Parker, Lishman, & Hughes, 1992).

Through a notorious experience using hybrid images, Schyns and Oliva (1994) have made an important contribution to the CtF hypothesis. Hybrid images are constructed by superimposing a filtered scene that preserves only the

[1]Université Grenoble Alpes, France, [2]Université Lille, France, [3]Université Hospital of Grenoble, France

© 2021 Massachusetts Institute of Technology

https://doi.org/10.1162/jocn_a_01689

LSF information on another filtered scene that preserves only HSF information. In this experiment, the two scenes belonged to two different semantic categories (e.g., a highway scene in LSF superimposed on a city scene in HSF). The hybrid image was briefly presented to participants (30 or 150 msec), followed by an unfiltered target scene. In half of the trials, the target scene was actually the unfiltered version of the LSF component or the HSF component of the hybrid. In the other half of the trials, the target scene was a new scene. Participants had to decide whether the unfiltered target scene was present in the hybrid image. When the hybrid image was presented for 30 msec, the participants were more likely to answer that the target scene was present in the hybrid when it corresponded to the LSF component of the hybrid than when it corresponded to the HSF component. When the hybrid image was presented for 150 msec, the reverse pattern was observed. The participants answered more often that the target scene was present in the hybrid when it corresponded with the HSF component than the LSF component. The results of this study indicate that the use of spatial frequencies evolves over time. The LSF would dominate the perception during the rapid processing of the scenes whereas the analysis of the fine information conveyed by HSF would be privileged when more time is allocated to the processing of the scene. Subsequently, other studies directly tested the hypothesis of the advantage of a CtF sequence of processing (Kauffmann, Chauvin, Guyader, et al., 2015; Musel et al., 2012). The experimental paradigm consisted in presenting sequences composed of six filtered images of the same scene (bandpass filtering) and assembled from LSF to HSF or from HSF to LSF, to impose a CtF or a fine-to-coarse sequence of analysis, respectively. Results showed that the categorization of these stimuli as indoor or outdoor scenes was faster for the CtF sequences, suggesting that the visual system benefits from rapidly accessing to LSF before HSF.

Critically, the first representation of the visual scene, based on LSF, would also trigger predictive mechanisms that would then guide a more detailed visual analysis. Consistent with predictive coding theories of visual processing (de Lange, Heilbron, & Kok, 2018; Friston, 2005; Rao & Ballard, 1999), the proactive model of visual object recognition proposed by Bar (2003, 2007) postulates that LSF contained in an object stimulus would be rapidly projected via the magnocellular pathway on the pFC and more particularly on the orbitofrontal cortex, which would generate predictions based on the coarse characteristics of the object. Predictions would be sent back into the inferotemporal cortex allowing to activate potential representations useful for the recognition of the object. Predictions based on the rapid processing of LSF would thereby facilitate the final process of object recognition when HSF arrive in the temporal cortex. Experimental arguments supporting this assumption were actually provided in several neuroimaging studies (Petras, ten Oever, Jacobs, & Goffaux, 2019; Kauffmann, Bourgin, Guyader, & Peyrin, 2015; Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Trapp & Bar, 2015; Kauffmann, Ramanoël, & Peyrin, 2014; Mu & Li, 2013; Peyrin et al., 2010; Kveraga, Boshyan, & Bar, 2007; Bar et al., 2006).

Using hybrid images as in the works of Schyns and Oliva (1994) mixed to a semantic interference paradigm, Mu and Li (2013) in an EEG study, and then Kauffmann, Bourgin, et al. (2015) in an fMRI study investigated how LSF could directly influence the processing of HSF. In these studies, hybrid images were constructed from scenes belonging to two categories: the artificial category (urban or semi-urban landscapes such as downtown streets, highways, etc.) and the natural category (landscapes beach, mountain, forest, etc.). The two scenes composing the hybrid image could be either semantically congruent (e.g., the LSF of an artificial scene superimposed on the HSF of another artificial scene) or semantically incongruent (e.g., the LSF of an artificial scene superimposed on the HSF of a natural scene). The authors warned participants that the stimuli contained two superimposed scenes, and asked them to attend and categorize the HSF component as an artificial versus natural scene, while ignoring the LSF component. Behavioral results showed a semantic interference effect. Participants were slower and made more errors to categorize the HSF component scene of the hybrid image when it was semantically incongruent to the LSF component than when it was semantically congruent. The authors also tested whether this semantic interference effect could be modulated by the physical similarity between the two components of the hybrid image. Thus, in each condition of semantic congruence, the two components of the hybrid image were either physically similar (PhySim; spatial superimposition—pixel by pixel—of the visual information of the two scenes and similarity of their amplitude spectra) or physically dissimilar (PhyDis). Mu and Li (2013), but not Kauffmann, Bourgin, et al. (2015), observed that the semantic interference effect was even greater when the two scenes shared the same physical properties, suggesting that the influence of LSF on HSF involves processing of both semantic and physical properties.

On a neurobiological level, ERP results from Mu and Li (2013) showed that the semantic interference effect was associated with a negative frontal component (N1) 122 msec after the hybrid image onset, well before the occipito-parietal (P2) and occipital (P3) components, 247 and 344 msec after the hybrid image onset, respectively. The difference in amplitude between the congruent and incongruent conditions on the early frontal component only appeared when the two scenes were PhySim. fMRI results from Kauffmann, Bourgin, et al. (2015) further showed greater activation of the inferior frontal cortex (at the level of the orbitofrontal cortex) when the LSF component was semantically incongruent with the HSF component than when it was congruent. The results of these two studies showed that LSF information, even if not relevant to the

Uncorrected Proof

task, would be processed automatically and would hinder the categorization of the scene in HFS. Mu and Li (2013) showed that this influence would be even greater when the spatial arrangement of the physical properties of the two scenes, as well as their spectral properties, are similar. Physical characteristics in the LSF scene would be also rapidly and automatically processed to generate predictions, resulting in a greater interference on the HSF scene categorization.

Unfortunately, the vast majority of experiments conducted to support either the CtF hypothesis or the predictive mechanisms of visual recognition are based on studies using small stimuli only displayed in central vision, without considering that the visual resolution varies considerably across the visual field. The density of retinal ganglion cells sensitive to HSF is the highest in the fovea whereas the density of retinal ganglion cells sensitive to LSF increases with eccentricity (Curcio & Allen, 1990; Wässle, Grünert, Röhrenberck, & Boycott, 1990). Therefore, while our subjective visual experience seems rich and detailed (partly thanks to eye movements), the extraction of HFS is only possible in the central retina, whereas the LSF are mainly extracted at the level of the peripheral retina. Because the majority of the signal in LSF comes from the peripheral vision, we can expect the visual system to use the information available in peripheral vision to activate predictions on a visual input and then use them to guide the analysis of the details contained in HSF in central vision.

The aim of this study was to understand how the low spatial resolution of information available in peripheral vision could modulate the processing of visual information presented in central vision. We combined behavioral measures (Experiment 1 and Experiment 2) and fMRI measures (Experiment 2) to investigate (1) whether a scene displayed in peripheral vision interferes with the categorization of a scene in central vision, (2) the role of the physical similarities between the scenes in this effect, and (3) the brain regions associated with the interference effect. We adapted the stimuli and the experimental protocol of Kauffmann, Bourgin, et al. (2015) by presenting the two scenes that originally are composed of a hybrid image simultaneously on the horizontal axis, one in central vision and the other one in peripheral vision, in either the right or the left visual field. Here, the two scenes were not filtered. Based on the nonhomogeneous spatial resolution of the visual information across the visual field, we considered that the presentation of the scenes in peripheral vision acted as a natural low-pass filter. We also believe that the filtering procedure may affect some properties of the signal, particularly when it comes from peripheral vision. For example, crowding mechanisms (Whitney & Levi, 2011; Pelli, 2008) may not work the same on a low-pass filtered image as they would on an intact image. By not filtering the scenes, we aimed at better imitating the natural signal and to preserve as much as possible the functioning of the mechanisms specific to peripheral vision. The two scenes belonged to the same category (semantically congruent) or to different categories (semantically incongruent). In addition, central and peripheral scenes could be either PhySim (similar amplitude spectra and spatial configuration) or PhyDis. This manipulation allowed us to assess whether the predictions resulting from the analysis of the peripheral scene are purely related to a semantic content or if, on the contrary, lower level physical characteristics are preserved in the predictive signal to constrain the analysis of the central scene. The participants categorized the scene presented in central vision according to the artificial or natural scene categories, while ignoring the scene presented in peripheral vision.

In the theoretical framework previously described, we hypothesized that during the categorization of a scene in central vision, the rapid processing of LSF available in the peripheral scene would allow the emergence of a rudimentary representation of the scene, named the gist. This representation could thus rapidly activate predictions about visual inputs, which would then constrain the processing of detailed information available in central vision. If these predictions guide the processing of information present in central vision, we expected to observe an interference effect of the peripheral scene on the categorization of the central scene. In behavioral measures, if the semantic information extracted from the peripheral scene is used to guide the categorization of the central scene, this would result in a classical semantic interference effect. We should observe more errors and longer response times (RTs) when the peripheral scene does not belong to the same semantic category as the central target scene (incongruent condition) than when it does (congruent condition). Furthermore, if these predictions also carry information about the physical properties of the scene (such as information on the content of spatial frequencies and orientations or on the spatial configuration), the semantic interference effect should be modulated by the physical similarity between the two scenes. Indeed, when the two scenes are semantically incongruent, an erroneous prediction including physical properties of the peripheral scene should be easier to reject if the central scene does not share such physical properties than when it does. In other words, the categorization of the central scene in the semantically incongruent condition would be even more impaired when the central scene matches the predictions in terms of physical properties. We therefore expected an interaction between the semantic congruence and the physical similarity of the scenes, that is, a greater semantic interference of the scene in peripheral vision when the two scenes are PhySim.

Alternatively, physical properties of the peripheral scenes may not be considered to activate predictions and may only interfere with the categorization of the central scene at early stages of visual processing as a visual distractor. Indeed, the simple detection of differences in low-level visual characteristics between the center and the periphery could distract the participant for performing

Uncorrected Proof

the categorization task in central vision. In this case, categorization performance would be impaired when the peripheral scene is PhyDis than similar irrespective of the semantic congruence between the scenes. Here again, the semantic interference effect should be modulated by the physical similarity. However, in that case, the categorization of the central scene in the semantically incongruent condition would be even more impaired when the peripheral scene is PhyDis.

In fMRI measures, Kauffmann, Bourgin, et al. (2015) previously observed that the semantic interference effect increased concomitantly the activation of the inferior frontal cortex (at the level of the orbitofrontal cortex) and the occipito-temporal cortex (at the level of the fusiform and parahippocampal gyri). We thus expected a greater activation of the inferior frontal cortex and occipito-temporal cortex associated with the semantic interference effect. Based on ERP results from Mu and Li (2013), these activations should be strengthened by the physical similarity between the two scenes.

EXPERIMENT 1

Methods

Participants

Twenty-six right-handed participants with normal or corrected-to-normal visual acuity were included in this experiment (24 women; mean age = 20, *SD* = 3 years). They were all psychology students at University Grenoble Alpes and received course credits for their participation. They provided informed written consent before participating in the study, in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki, 2013).

Stimuli

Stimuli were 160 black-and-white photographs of scenes (256-level gray scales, 256 × 256 pixels) from the Labelme database (Oliva & Torralba, 2001), previously used in works of Kauffmann et al. (2017) and Kauffmann, Chauvin, Pichat, et al. (2015). Scenes belonged to two distinct semantic categories: 80 man-made scenes (buildings, streets, highways) and 80 natural scenes (beach, open countryside, forests). The mean luminance and standard deviation (*SD*) of each scene were fixed at 117 (for pixel values comprised between 0 and 255) and 64, respectively. These values correspond to the mean luminance and the mean *SD* of the 160 scenes.

Scenes were selected to form 80 pairs: 40 pairs of scenes semantically congruent (congruent condition, 20 pairs of man-made scenes and 20 pairs of natural scenes) and 40 pairs of scenes semantically incongruent (incongruent condition; all composed of a man-made scene and a natural scene). Moreover, for each semantic congruence condition (congruent and incongruent), half of pairs was made of PhySim scenes and the other half was made of PhyDis scenes. Based on the works of Kauffmann et al. (2017) and Kauffmann, Chauvin, Pichat, et al. (2015), the physical similarity between the two scenes was assessed on two dimensions: (1) the similarity between their amplitude spectrum in the Fourier space, based on the correlation between pixel intensity values of the distribution of amplitude over spatial frequencies and orientations of the two scenes and (2) the similarity between their spatial configurations, based on the correlation between pixel intensity values of the two scenes, pixel per pixel. Two scenes were considered as PhySim if both correlation coefficients were superior to 0.6 and as PhyDis if the correlation coefficient based on the amplitude spectrum was inferior to 0.6 and the correlation coefficient based on pixel intensity was inferior to 0.01. Thus, two PhySim scenes shared similar statistics in terms of spatial frequencies and dominant orientations, and they shared a subjectively similar spatial configuration (spatial superimposition pixel by pixel; Figure 1A).

Procedure

Pairs of scenes were displayed using E-prime software (E-prime Psychology Software Tools Inc.) against a gray background (pixel value of 117 on a 256-level grayscale matching the scenes mean luminance value) on a 30-in. monitor with a resolution of 3560 × 1600 pixels and a refresh rate of 60 Hz. Participants sat in a darkened room, with their head stabilized with a chinrest at 55 cm from the screen. At this distance, each scene subtended 6 × 6 degrees of visual angle. Participants performed two experimental blocks: one block with only PhySim pairs and the other one with only PhyDis pairs. Within each block, pairs of scenes were selected randomly within the two semantic congruence conditions. Each trial began with a central fixation dot for 500 msec to attract the gaze direction to the center of the screen, immediately followed by a pair of scenes for 100 msec to avoid saccadic eye movement usually initiated within 100–150 msec (Fischer & Weber, 1993). One scene of the pair was displayed at the center of the screen, and the other one was randomly displayed in either the left visual field (half trials) or the right visual field (half trials), along the horizontal meridian (Figure 1B). For each pair, one scene was displayed 2 times peripherally, once in the left visual field, once in the right visual field (whereas the other scene was displayed 2 times in the central visual field), and 2 times in the central visual field (whereas the other scene was displayed 2 times peripherally, once in the left visual field, once in the right visual field). Thus, each scene of our stimulus base was seen 4 times throughout the experiment. The eccentricity of lateralized scenes in peripheral vision was set at 3.75° of visual angle from their inner edge. The peripheral image center was thus lateralized at 6.75° of visual angle. Thereafter, a gray background was displayed for 1900 msec during which participant could answer.

Uncorrected Proof

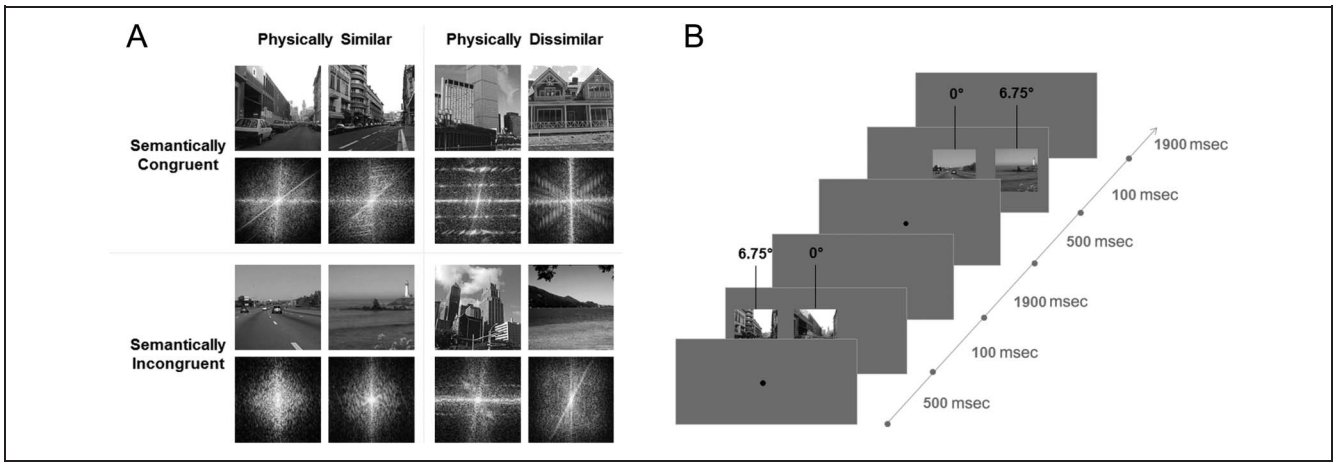


Figure 1. (A) Example pairs of scenes with their amplitude spectrum used in Experiments 1 and 2: The two scenes of could be either semantically congruent (e.g., two man-made scenes) and PhySim, semantically incongruent (a man-made scene and a natural scene) and PhySim, semantically congruent (e.g., two man-made scenes) and PhyDis, or semantically incongruent (a man-made scene and a natural scene) and PhyDis. (B) Trial schematic: One scene of the pair was displayed at the center of the screen, and the other one was randomly displayed either in the left visual field or the right visual field, along the horizontal meridian. The peripheral image center was lateralized at 6.75° of visual angle.

Participants had to ignore the peripheral scene and to categorize the central scene as a man-made or a natural scene by pressing the corresponding key with the forefinger and the middle finger of their dominant hand. They were instructed to respond as accurately and fast as possible. Response keys were counterbalanced across participants. For each trial, response accuracy and RTs (in msec) were recorded. There were 320 trials in total, 80 in each experimental condition (congruent-PhySim, congruent-PhyDis, incongruent-PhySim, incongruent-PhyDis). The experiment lasted about 20 min. Before the experiment, participants underwent a training with pairs of scenes that differed from those used in the experiment to get familiarized with the stimuli, response keys, and task.

Results

RTs for each participant and each experimental condition were trimmed to reduce the effect of extreme values, by removing RTs exceeding ± 2.5 *SD*s from the mean. This resulted in removing an average of 2.68% of the trials. A logarithmic transformation was applied on RTs' data to bring the distribution of statistical models residuals closer to normality.

We analyzed accuracy using mixed-effects logistic regression models, and RTs with mixed-effects linear regression models, using lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R. Mixed models take the whole data set as input (i.e., accuracy for each trial of each participant) and estimate simultaneously the effects at the population level (fixed effects) and the variability of these effects between participants (random effects). Estimation of the random effects provides a precise estimate of the fixed effects. Models aiming at testing our main hypotheses included as independent variables either semantic congruence only (for the test of the main semantic interference effect), physical similarity only (for the test of the main effect of the physical similarity), or semantic congruence and physical similarity (for the test of the interaction effect). In addition, to ensure that there was no asymmetry in the effect of the peripheral scene according to its location in the visual field, we ran more complex interaction models including the visual field (left and right visual fields) in addition to the semantic congruence and physical similarity. We included in each model a varying intercept by participant as a random factor. Because random effects per se are not interesting in the context of our research question and are only included in the model to allow for better parameter estimates, we only report fixed effects. In lme4 syntax, the model is written as follows: DV ~ 1 + *effect of interest* + (1 + *effect of interest* | Participant), with DV as the dependent variable, either accuracy or RTs, 1 as the intercept, and *effect of interest* as the fixed-effects term that depends on the tested effect (Congruence, Physical Similarity, Congruence × Physical Similarity, Congruence × Physical Similarity × Visual Field).

For accuracy, the statistical significance of the logistic models was tested with a Wald test and the size of the interference effect was reported through the odds ratio (*OR*), computed as *odds(Congruent)/odds(Incongruent)* where *odds(Congruent)* is the exponential of the *intercept* of the model and *odds(Incongruent)* is the exponential of *intercept* + β. For RTs, the linear models were fitted using a restricted maximum likelihood estimation of variance component and statistical significance was tested by deriving degrees of freedom using the Satterthwaite approximation. Effect sizes are reported through the percent change (%change), computed as $(exp(\beta) - 1) \times 100$. It should be noted that we analyzed log(RTs), but raw RTs (in msec) are actually displayed on the graphs. Statistical significance was set at an alpha level of .05.

Uncorrected Proof

Mean error rate (mER) and mean correct reaction times in milliseconds (mRT), with *SD*, for each experimental condition (Semantic Congruence × Physical Similarity × Visual Field of Presentation) are reported in Table 1A. Accuracy was very high over the different experimental conditions, with distributions of error rates suggesting a ceiling effect (Figure 2A). We tested whether the semantic congruence between the peripheral and the central scene influenced accuracy. This analysis showed a significant main effect of semantic congruence (i.e., a semantic interference effect) on accuracy. The mER was significantly higher when two scenes were semantically incongruent than congruent (congruent: $M = 3.75$, $SD = 4.57\%$; incongruent: $M = 4.74$, $SD = 4.67\%$; $\beta = -0.26$, $OR = 1.29$, $z = -2.32$, $p = .020$; Figure 2A). We then tested the main effect of physical similarity between the scenes that was not significant (PhySim = 5.96 ± 4.49%, PhyDis = 3.54 ± 2.80%, $\beta = 0.56$, $z = 3.98$, $p < .001$). There was no interaction between the semantic congruence and the physical similarity ($\beta = 0.20$, $z = 0.89$, $p = .374$). It should be noted that the interaction model including the visual field was not significant (Semantic Congruence × Physical Similarity × Visual Field; $\beta = 0.03$, $z = 0.07$, $p = .941$).

Then, we tested whether the semantic congruence between the peripheral and the central scenes influenced RTs. This analysis showed a significant main effect of semantic congruence. The categorization of the central scene was longer when the peripheral scene was semantically incongruent than congruent (congruent = 627 ± 97 msec, incongruent = 641 ± 96 msec; $\beta = 0.02$, %change = 2.29%, $t(7693) = 4.72$, $p < .001$). The main effect of physical similarity was also significant. The categorization of the central scene was longer when the peripheral scene was PhySim than dissimilar (PhySim = 655 ± 105 msec, PhyDis = 611 ± 92 msec; $\beta = 0.07$, %change = 6.77%, $t(7693) = 13.83$, $p < .001$). Importantly, the interaction model then showed that the semantic interference effect was modulated by the physical similarity between scenes ($\beta = 0.03$, $t(7691) = 2.97$, $p = .003$; Figure 2A). More precisely, an analysis of the simple effects showed that there was an interference effect when the two scenes were PhySim (Congruent-PhySim = 645 ± 104 msec; incongruent-PhySim = 668 ± 108 msec; $\beta = 0.04$, %change = 3.78%, $t(3826) = 5.52$, $p < .001$), but not when they were PhyDis (congruent-PhyDis = 610 ± 98 msec; incongruent-PhyDis = 613 ± 87 msec; $\beta = 0.009$, $t(3840) = 1.38$, $p = .168$). Moreover, the simple effect of physical similarity was significant both when the two scenes were semantically congruent ($\beta = 0.051$, $t(3851) = 7.73$, $p < .001$) and semantically incongruent ($\beta = 0.080$, $t(3815) = 11.87$, $p < .001$). For both conditions, the categorization of the central scene was longer when the peripheral scene was PhySim than dissimilar. However, the significant interaction suggests that the physical similarity effect is more important in the incongruent than congruent condition. The modulation of the interference effect by the physical similarity did not depend on the visual field (Semantic Congruence × Physical Similarity × Visual Field; $\beta = 0.008$, $t(7692) = 0.39$, $p = .694$).

EXPERIMENT 2

Methods

Participants

Fifteen right-handed participants (who were not included in Experiment 1) with normal or corrected-to-normal

Uncorrected Proof

Table 1. mER and mRT, with *SD*s, for Semantically Congruent and Incongruent pairs of Scenes, Either PhySim or Dissimilar (PhyDis) and Displayed Either in Left Visual Field (LVF) or Right Visual Field (RVF) in (A) Experiment 1 and (B) Experiment 2

	Congruent				*Incongruent*			
	PhySim		*PhyDis*		*PhySim*		*PhyDis*	
	LVF	*RVF*	*LVF*	*RVF*	*LVF*	*RVF*	*LVF*	*RVF*
(A) Experiment 1								
mER	3.87	3.46	3.95	3.67	5.14	5.05	4.43	4.34
SD	5.27	5.15	4.61	5.55	6.13	5.29	4.49	5.27
mRT	617	604	645	644	616	612	667	668
SD	107	92	104	109	91	90	108	114
(B) Experiment 2								
mER	4.67	5.56	3.83	4.33	6.83	7.00	3.11	2.93
SD	4.58	5.59	4.52	4.58	6.58	5.01	5.11	3.37
mRT	621	605	618	604	643	636	622	605
SD	84	79	107	86	103	91	89	76

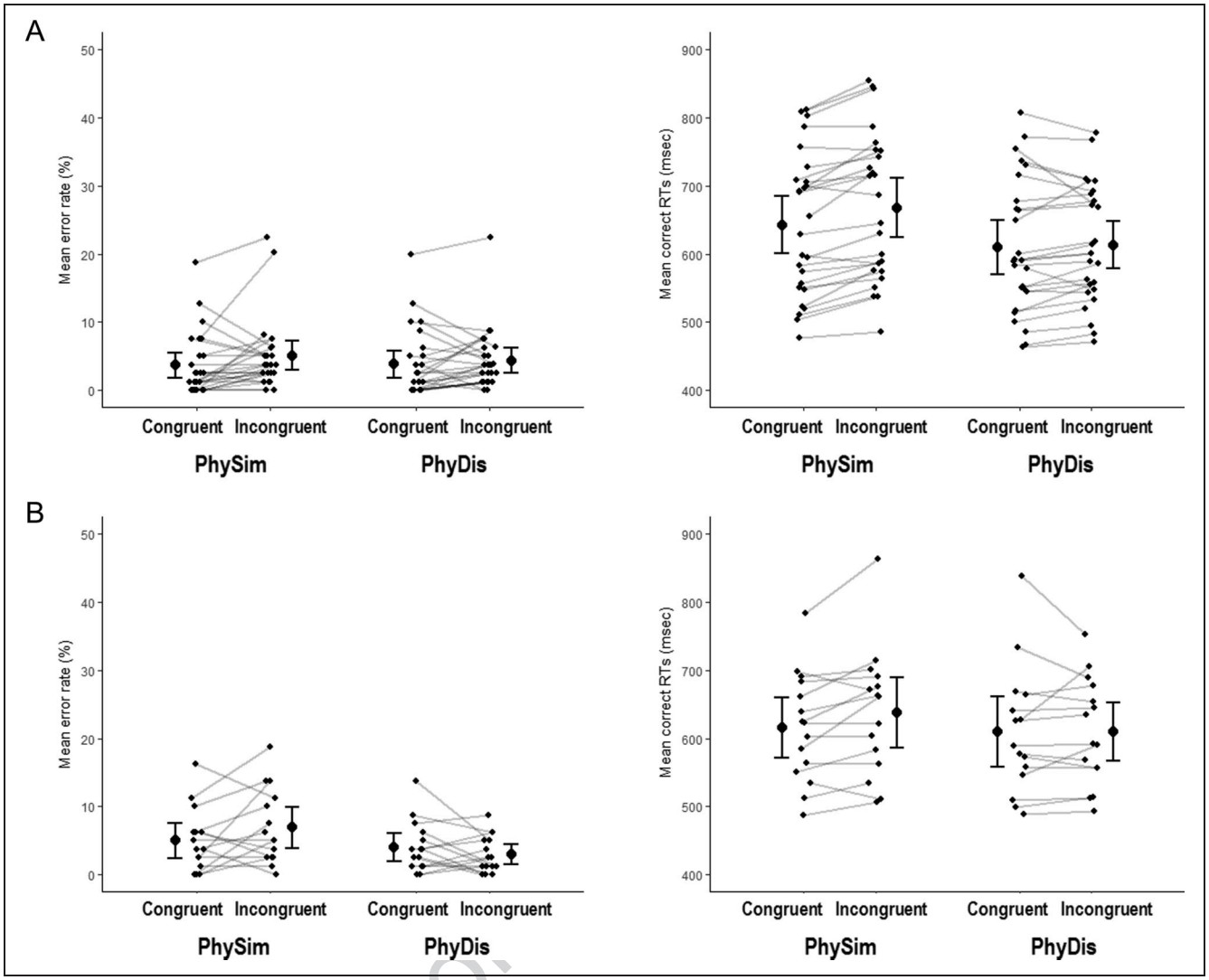


Figure 2. mERs (in %) and mean correct RTs (in msec) in (A) Experiment 1 and (B) Experiment 2 for the categorization of the central scene in semantically congruent and incongruent pairs of scenes, either PhySim or dissimilar (PhyDis). The small black dots are the averages of each participant (slightly jittered horizontally for better visualization), and the bigger black dots with error bars indicate means at the group level and 95% confidence intervals.

Uncorrected Proof

visual acuity and no history of neurological disorders (eight women; $M = 23$, $SD = 3$ years) participated in the experiment. They provided informed written consent before participating in the study, which was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) and approved by the local ethics committee (CPP ID/RCB : 2016-A00637-44).

Stimuli and Procedure

Stimuli were the same as in Experiment 1 and displayed using E-prime software. They were back-projected onto a translucent screen positioned at the rear of the MRI magnet. Participants viewed the screen through a mirror fixed on the head coil. In the scanner, the participants had to categorize the central scene (as being "natural" or "artificial") while ignoring the peripheral scene, as in Experiment 1. Each trial began with a central fixation dot on a gray background for 500 msec to control the gaze direction to the center of the screen, immediately followed by a pair of scenes for 100 msec. One scene on the pair was displayed at the center of the screen, and the other one was randomly displayed in either the left visual field (half trials) or the right visual field (half trials), along the horizontal meridian. As in Experiment 1, the eccentricity of lateralized scenes was set at 3.75° of visual angle from their inner edge. The image center was thus lateralized at 6.75° of visual angle. Thereafter, a gray background was displayed for 1900 msec during which the participants could answer. They were instructed to respond as accurately and quickly as possible by pressing the keys of a response box disposed inside the scanner with their right hand. Response keys were counterbalanced across participants. At each trial, accuracy and RT were recorded. Before the test session, participants underwent a training session

outside the scanner with pairs of scenes that differed from those used in the experiment to get familiarized with the stimuli and task.

The experiment included two functional scans designed with a pseudorandomized event-related paradigm. Each functional run was made of 160 events (40 by experimental condition: congruent-PhySim, congruent-PhyDis, incongruent-PhySim, incongruent-PhyDis) and of 26 rest events (including six at the end of the scan, and during which a fixation dot was displayed on a gray background). The order of conditions and rest events was pseudorandomized based on an optimization algorithm (Friston, Zarahn, Josephs, Henson, & Dale, 1999). The different pairs of scenes were displayed after two different pseudorandomized orders, one for each functional run. With each functional run, 186 functional volumes were acquired. Each of them lasted 7 min and 45 sec.

fMRI Acquisition and Analysis

The experiment was performed using a whole-body 3 T Philips scanner (Achieva 3.0T TX Philips, Philips Medical Systems) with a 32-channel head coil at the Grenoble MRI facility IRMaGe in France. For each functional scan, a manufacturer-provided gradient-echo/ T2*-weighted EPI method was used. Forty-two adjacent axial slices parallel to the bicommissural plane were acquired in sequential mode from the bottom to the top, including the cerebellum. Slice thickness was 3 mm. The in-plane voxel size was 3 × 3 × 3 mm (240 × 240 × 126 mm field of view acquired with a 80 × 80 pixel data matrix; reconstructed with zero filling to 80 × 80 pixels). The main sequence parameters were as follows: repetition time = 2.5 sec, echo time =30 msec, flip angle = 82°. Before each functional run, six "dummies scans" were acquired to allow for signal equilibration. After the two functional runs, a T1-weighted high-resolution three-dimensional anatomical volume was acquired, by using a 3-D T1 TFE sequence (field of view = 256 × 224 × 175; resolution: 1.333 × 1.750 × 1.375 mm; acquisition matrix: 192 × 115 × 128 pixels; reconstruction matrix: 288 × 288 × 128 pixels).

Functional data of each participant was then analyzed using the general linear model (Friston et al., 1995) for event-related designs. At the subject level, four conditions of interest (congruent-PhySim, congruent-PhyDis, incongruent-PhySim, and incongruent-PhyDis) were modeled as four regressors convolved with a canonical hemodynamic response function. We also entered the movement parameters derived from realignment corrections (three translations and three rotations) into the design matrix as additional factors of no interest to account for head motion-related variance. Analyses were performed at the individual subject level to examine the contrasts between conditions of interest. These contrast images were then entered into second-level random effect analyses to test for within-group effects (one-sample *t* tests). The significance of activations was assessed with a statistical threshold of 0.05, FWE rate corrected at the cluster level, and a minimum cluster extent of 10 voxels.

Results

Behavioral Results

Behavioral data analyses were conducted as in Experiment 1. mER and mRT, with *SD*s, for each experimental condition (Semantic Congruence × Physical Similarity × Visual Field of Presentation) are reported in Table 1B. In this experiment, 2.89% of trials were trimmed before the analysis of RTs. Here again, the low rate of errors suggested a ceiling effect (Figure 2A). There was no main effect of the semantic congruence between the peripheral and the central scenes on accuracy (congruent = 4.54 ± 3.97%, incongruent = 4.96 ± 4.00%; $\beta = -0.09$, $z = -0.69$, $p = .488$; Figure 2B). However, the main effect of the physical similarity between the scenes was significant. The mER was significantly higher when the two scenes were PhySim than dissimilar (PhySim = 4.38 ± 4.57%, PhyDis = 4.10 ± 4.46%, $\beta = 0.07$, $z = 0.66$, $p = .512$). Importantly, the physical similarity interacted with the semantic congruence ($\beta = -0.68$, $z = -2.40$, $p = .017$). The analysis of the simple effects showed that there was an interference effect when the two scenes were PhySim (Congruent-PhySim: 5.00 ± 4.63%, incongruent-PhySim: 6.92 ± 5.47%; $\beta = -0.35$, $OR = 1.43$, $z = -2.03$, $p = .042$), but not when they were PhyDis (congruent-PhyDis: 4.08 ± 3.73%, incongruent-PhyDis: 3.00 ± 2.71%; $\beta = 0.32$, $z = 1.47$, $p = .143$). In addition, when the two scenes were semantically incongruent, the mER was significantly higher when the two scenes were PhySim than dissimilar ($\beta = -0.90$, $OR = 2.46$, $z = -4.48$, $p < .001$). There was no main effect of the physical similarity when the two scenes were congruent ($\beta = -0.22$, $z = -1.11$, $p = .266$). The interaction model including the visual field was not significant (Semantic Congruence × Physical Similarity × Visual Field; $\beta = 0.03$, $z = 0.05$, $p = .958$).

Then, we tested whether the semantic similarity between the peripheral and the central scene influenced RTs. This analysis showed a significant main effect of semantic congruence on RTs. The categorization of the central scene was longer when the peripheral scene was semantically incongruent than congruent (congruent: 612 ± 85 msec, incongruent: 623 ± 84 msec; $\beta = 0.02$, %change = 1.80%, $t(4419) = 2.65$, $p = .008$; Figure 2B). The main effect of physical similarity was also significant. Categorization of the central scene was longer when the peripheral scene was PhySim than dissimilar (PhySim = 627 ± 85 msec, PhyDis = 609 ± 84 msec; $\beta = 0.03$, % change = 2.65%, $t(4419) = 3.89$, $p < .001$). Unlike in Experiment 1, there was no interaction between the semantic congruence and the physical similarity ($\beta = 0.03$, $t(4417) = 1.93$, $p = .054$). Because the effect was significant in Experiment 1 that was more powered

Uncorrected Proof

(26 participants vs. 15 here), we still tested the simple effects in a post hoc analysis, at a Bonferroni-corrected significance threshold of .0125 ($\alpha = .05/4$). This analysis first showed that, similarly to Experiment 1, there was a semantic interference effect when the two scenes were PhySim (congruent-PhySim: 615 ± 79 msec, incongruent-PhySim: 639 ± 93 msec; $\beta = 0.03$, %change = 3.22%, $t(2176) = 3.26$, $p = .001$), but not when they were PhyDis (congruent-PhyDis: 611 ± 93 msec, incongruent-PhyDis: 611 ± 80 msec; $\beta = 0.005$, $t(2227) = -0.01$, $p = .582$). When the scenes were semantically incongruent, the effect of the physical similarity was significant ($\beta = 0.04$, %change = 4.04%, $t(2188) = 4.18$, $p < .001$), RTs being longer when the two scenes were PhySim than dissimilar. When the scenes were semantically congruent, the effect of the physical similarity was not significant ($\beta = 0.01$, $t(2215) = 1.42$, $p = .155$). Finally, the interaction model including the visual field was not significant (Semantic Congruence × Physical Similarity × Visual Field; $\beta = -0.03$, $t(4418) = -0.99$, $p = .324$).

Whole-Brain fMRI Results

The analysis of accuracy and RT showed no effect of the visual field. This factor was therefore not further considered in the analysis of fMRI data. The main objective of these analyses was to identify brain regions associated with the semantic interference effect by contrasting activations elicited by the incongruent condition to activations elicited by the congruent condition ([incongruent > congruent] contrast; Figure 3A). When considering both physical similarity conditions, the result of this contrast did not reveal any significant activation. By convention and for exploratory purposes, we also tested the reverse [congruent > incongruent] contrast. This contrast showed an activation of the right lingual gyrus (Montreal Neurological Institute coordinates of the peak: 12x, −73y, −10z, BA 18, $k = 143$, $t = 7.09$; Figure 3B). We also tested the effect of the physical similarity between the central and peripheral scenes with the [PhySim > PhyDis] and [PhyDis > PhySim] contrasts. Neither of the two contrasts revealed any significant activation. We next assessed the interaction between the semantic congruence and the physical similarity between the two scenes. A significant interaction was observed in the right temporal cortex at the level of the middle temporal gyrus (54x, −52y, 2z, BA 37/21, $k = 39$, $t = 5.98$).

As behavioral results of Experiments 1 and 2 highlighted a semantic interference effect when the two scenes were PhySim, we thus tested the effect of semantic congruence with the [incongruent > congruent] and [congruent > incongruent] contrasts for each condition of physical similarity independently. When two scenes were PhySim (PhySim condition), the [incongruent-PhySim > congruent-PhySim] contrast revealed a bilateral activation within the frontal cortex, reflecting the interference effect observed in behavior. This activation was located in the inferior frontal gyrus (pars orbitalis and pars triangularis) and extended in both hemispheres to the anterior insular cortex (right hemisphere: 33x, 11y, 2z, BA 47/13, $k = 131$, $t = 7.59$; left hemisphere: −33x, 23y, 2z, BA 47/13, $k = 58$, $t = 4.11$; Figure 3B). On the other hand, the reverse [congruent-PhySim > incongruent-PhySim] contrast did not reveal any significant activation. When two scenes were PhyDis (PhyDis condition), the [incongruent-PhyDis > congruent-PhyDis] contrast did not reveal any significant cluster, reflecting the lack of interference effect observed in behavior. However, the reverse contrast ([congruent-PhyDis > incongruent-PhyDis]) shows a more widespread activation in the visual regions of the right hemisphere. In

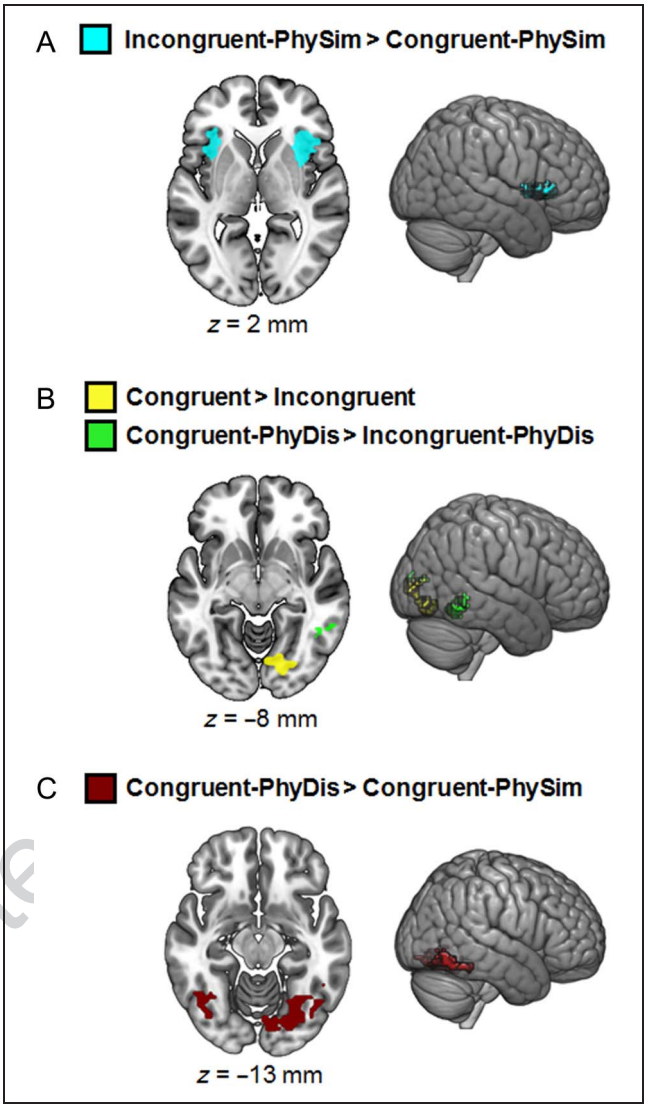


Figure 3. (A) Activations elicited by the semantic interference effect ([incongruent > congruent] contrast). (B) Activations elicited by the reverse contrast ([congruent > incongruent]) as well as by the same contrast for the PhyDis condition ([congruent-PhyDis > incongruent-PhyDis]). (C) Activations elicited by the physical dissimilarity ([congruent-PhyDis > congruent-PhySim] contrast). All activations are reported at a statistical threshold of $p < .05$ FWE-corrected at the cluster level. The z value (mm) represents the slice level with respect to the bicommissural plane.

Uncorrected

this specific case, we observed activations at the junction between the fusiform gyrus and the inferior temporal gyrus (48x, −55y, −13z, BA 19/37, k = 52, t = 5.26), and in the cuneus (12x, −91y, 11z, BA 47/13, k = 42, t = 4.80).

We also tested the effect of physical similarity between the central and peripheral scenes with the [PhySim > PhyDis] and [PhyDis > PhySim] contrasts for each condition of semantic congruence independently. When the scenes were semantically congruent, the [congruent-PhyDis > congruent-PhySim] contrast revealed bilateral activations within the occipito-temporal cortex at the level of the lingual and fusiform gyri (peak coordinate in the left hemisphere: −24x, −61y, −19z, BA 19/37, k = 259, t = 4.41; in the right hemisphere: 39x, −70y, −16z, BA 19/37, k = 89, t = 3.97; Figure 3C). The reverse [congruent-PhySim > congruent-PhyDis] contrast did not reveal any significant activation. Finally, when the scenes were semantically incongruent, neither [incongruent-PhySim > incongruent-PhyDis] and [incongruent-PhySim > incongruent-PhyDis] contrasts reveal any significant cluster.

DISCUSSION

Experiment 1 showed that the categorization of a scene in central vision was influenced by the presence of a scene in peripheral vision. Participants made more errors to categorize the central scene when the peripheral scene was semantically incongruent than congruent. This result suggests that peripheral information is processed automatically at a semantic level and integrated into the process of recognition in central vision. Consistently with Mu and Li (2013) who used hybrid images, we observed that this peripheral interference effect on RTs was strengthened when the two scenes shared similar physical properties (i.e., similar amplitude spectra and spatial configurations). In the present experiment, the task was easier than the one used with hybrid images, possibly favoring the observation of an effect of physical similarity not observed in the works of Kauffmann, Bourgin, et al. (2015). Importantly, this result suggests that a peripheral scene PhyDis did not act as a visual distractor, but that physical properties of the peripheral scene are also integrated into the process of recognition in central vision.

In Experiment 2, we adapted Experiment 1 to the fMRI technique to couple behavioral measures with neurobiological measures. We first observed the same behavioral interference effect of Experiment 1 on correct RTs. The interference effect was less consistent between the two experiments for the proportion of errors, which is possibly because of the sample size, smaller in Experiment 2 than in Experiment 1. However, in Experiment 2, we observed that participants made more errors and had longer RTs to categorize a central scene when an incongruent peripheral scene was PhySim than dissimilar. This result supports again the use of physical properties of the peripheral scene for categorizing the central scene. fMRI results showed that the semantic interference effect was associated with the bilateral activation of the inferior frontal gyrus, extending to the anterior insula. In other words, this region was more active when the central and peripheral scenes were incongruent than congruent. This activation was only observed when the two scenes had similar physical properties, again emphasizing the role of the physical properties of stimuli in the semantic interference effect of the peripheral scene. An unexpected result concerns the occipito-temporal cortex activation observed when the two scenes were semantically congruent but did not share similar physical properties.

Peripheral Vision Influence and Physical Similarity Effect

Several studies have highlighted the importance of peripheral vision during the categorization of scenes (Trouilloud et al., 2020; Loschky, Szaffarczyk, Beugnet, Young, & Boucart, 2019; Lukavsky, 2019; Geuzebroek & van den Berg, 2018; Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013; Larson & Loschky, 2009). The categorization of scenes remains possible even in far peripheral vision (70° retinal eccentricity; Boucart et al., 2013). Using a window-scotoma paradigm in which participants saw either the central part of a scene (window stimulus) or the peripheral part by hiding its center (scotoma stimulus), Larson and Loschky (2009) showed that the participants' categorization performance was higher in the scotoma/peripheral than the window/central condition. Using similar stimuli, Trouilloud et al. (2020) revealed that scene sequences revealing the peripheral part of a scene before the central part were categorized more rapidly than the reverse sequences. The low resolution of peripheral vision would therefore be more useful than the high resolution of central vision to categorize a scene very quickly. Lukavsky (2019) presented scotoma and window stimuli simultaneously and asked participants to attend to and categorize one of the stimuli while ignoring the other one. Similarly to our study, stimuli were either semantically congruent or incongruent. Participants' performances decreased when the two scenes were incongruent whatever the stimuli to categorize (i.e., the scotoma or window one), suggesting that the information available in peripheral vision could not be ignored and was integrated to the processing of central information, and vice-versa. Roux-Sibilon et al. (2019) demonstrated that the information available in peripheral vision facilitates the processing of foveated objects in a visual scene in a predictive way, by presenting the information in peripheral vision slightly before the one in central vision. Participants performed a categorization task of an object displayed in central vision while a semantically congruent or incongruent scene background was displayed in peripheral vision and had to be ignored. Results showed that the congruence effect was stronger when the peripheral scene was displayed before the object's onset. Participants' performance was higher to categorize objects in a congruent scene background than

Uncorrected Proof

in an incongruent one. In line with the hypothesis of predictive mechanisms involved in visual recognition, these results suggest that when the scene background was sufficiently processed before the object onset, participants automatically used this information in the peripheral visual field (although it was irrelevant to the task) and generated predictions about the visual input in central vision. Similarly, in the present experiment, we hypothesized that the rapid processing of LSF in the peripheral scene would allow the emergence of the gist of the scene used to activate predictions about visual inputs, which would then constrain the processing of the central scene. We thus expected an interference effect of the peripheral scene on the categorization of the central scene formalized by an effect of semantic congruence between the target scene in central vision and the distracting scene in peripheral vision (better performance when the peripheral scene was congruent than incongruent).

Overall, the patterns of behavioral results of the two experiments of this study are consistent with these previous findings and allow us to draw the same conclusions. First, and consistent with our hypothesis, we observed an interference effect of peripheral vision on categorization in central vision. In addition, in both experiments, we observed a ceiling effect that occurred when accuracy data were considered alone (the average proportion of errors was around 0.05%), suggesting that the categorization task was easy to perform, making it difficult to detect the expected effects. Regarding RTs, the semantic interference effect was observed in both experiments. This effect suggests that even if the peripheral scene presented is useless for performing the task, its processing is automatic and integrated to the categorization process of the scene in central vision. Results of this study can be explained by a predictive mechanism in which the information coming from peripheral vision would be automatically used to generate predictions integrated into the processing of signal in central vision. In the context of predictive coding models of visual perception (de Lange et al., 2018; Bar, 2003, 2007; Friston, 2005), an irrelevant prediction compared to the central scene to categorize would lead to a prediction error that would delay the process. In everyday life, this mechanism would have an interest for the visual system. Prediction error could be used to update prior knowledge about the visual input.

In both experiments, we observed that the interference effect of peripheral vision was dependent on the physical similarity between the two scenes. Indeed, we observed a significant semantic interference effect when the two scenes were PhySim, but not when they were PhyDis. This result suggests that the physical characteristics of the peripheral stimulus were also used and modulated the processing of the signal in central vision. In addition, the interaction effect suggests that these low-level properties would be used in priority to generate predictions, because the PhyDis peripheral scene did not influence the categorization in central vision. However, we strongly believe that the integration of high-level information in PhyDis peripheral scenes remains plausible. Indeed, we can hypothesize that predictions containing the highest level of information conveyed by these peripheral scenes are still useful. However, the influence of these predictions would be minimal and therefore the absence of interference effect in the condition PhyDis may be either related to a problem of experimental design or statistical power. It is also possible that the processing of the PhyDis information between the two scenes facilitated the rejection of erroneous semantic predictions about the central scene, resulting in a reduced interference effect.

How to explain that the effect of semantic interference was amplified by the physical similarity of the peripheral scene? Our interpretation is that, in the context of such an artificial/natural categorization task, physical information (i.e., low-level statistics and overall spatial configuration) extracted in peripheral vision could rapidly be used to generate semantic predictions about the visual input in central vision (i.e., its category, either natural or artificial). Indeed, previous studies have, for example, shown that the mere distribution of spatial frequencies across dominant orientations available in the amplitude spectrum of scenes, or coarse spatial features such as the degree of openness or expansion of scenes, can be sufficient to distinct scene categories such as natural and man-made environments (Torralba & Oliva, 2003; Oliva & Torralba, 2001). When the peripheral and central scenes are semantically incongruent, the extraction of physical information in the peripheral scene would lead to erroneous semantic predictions. If the central scene is however consistent with these predictions in terms of physical properties (i.e., PhySim condition), its processing would result in increasing the uncertainty about the actual irrelevance of the predictions and, thus, the conflict between (1) top–down predictions based on peripheral vision and (2) bottom–up processing of the central scene. Therefore, the effect of physical similarity observed in both experiments suggests that the gist representation resulting from the analysis of a low spatial resolution information available in peripheral vision, and influencing recognition in central vision, would not be of a purely semantic nature. On the contrary, the representation would retain some low-level properties of the stimulus. The two dimensions that we manipulated to establish the physical similarity were (1) the amplitude spectral properties of the scenes (i.e., distribution of amplitude over spatial frequencies and orientations) and (2) their spatial configuration (i.e., spatial superimposition, pixel by pixel, of the luminance information). Unfortunately, the nature of our stimuli did not allow us to disentangle between these two dimensions, which would be useful to know more precisely which low-level properties of the stimulus are the most important for gist-based visual predictions.

Although our experimental paradigm allows us to precisely control some parameters of physical similarity, it is not very ecological. For example, we did not take into

Uncorrected Proof

account the phenomenon of cortical magnification (Daniel & Whitteridge, 1961). The projection of information from central and peripheral visions on the primary visual cortex undergoes a deformation so that the central vision is overrepresented at the cortical level in comparison to peripheral vision. Consequently, to stimulate the same number of cells of the primary visual cortex, the angular size of a scene presented in peripheral vision must be larger than the one in a scene presented in central vision. As it was not the case in this study (both scenes sized 6° of visual angle), it is therefore possible that we have underestimated the size of the interference effect and that the influence of peripheral vision would have been more important with a peripheral scene covering a larger part of the peripheral visual field.

Neural Bases of the Peripheral Vision Influence

In Experiment 2, we used fMRI to investigate the cerebral regions involved in the interference effect of peripheral vision. Previous neuroimaging studies (Kauffmann, Chauvin, Pichat, et al., 2015; Kauffmann et al., 2014; Peyrin et al., 2010; Kveraga et al., 2007; Bar et al., 2006) suggest that predictions are generated through the rapid processing of LSF in the inferior frontal cortex and especially in its orbitofrontal part. Predictions would then be transmitted to the occipito-temporal cortex where they would be compared to the ascending signal resulting mainly from the processing of HSF. The ascending signal that does not correspond to predictions would then be transmitted to the inferior frontal cortex to update the predictions. The larger the prediction error, the more the exchanges between these two regions should be important, which would result in an increase in brain activity. For example, in the works of Kauffmann, Bourgin, et al. (2015), participants had to categorize the HSF scene of a hybrid image presented in central vision while ignoring the LSF scene, either congruent or not with the HSF scene. The activity induced by the semantically congruent condition was subtracted from the one induced by the semantically incongruent condition. The semantic interference effect was associated with a bilateral activation of the inferior frontal gyrus (at the level of the orbitofrontal cortex), as well as the fusiform and parahippocampal gyri. Very similarly, in our fMRI experiment, even if the central scene was not filtered in HSF, participants could categorize a central scene of high spatial resolution while ignoring a peripheral scene of low spatial resolution (i.e., LSF content), either congruent or not with the central scene. When contrasting the congruent condition to the incongruent condition ([incongruent > congruent] contrast), we also observed a bilateral activation of the inferior frontal gyrus. This activation was only observed when the two scenes were PhySim, a result consistent with our behavioral results showing a semantic interference effect only in case of physical similarity.

The activation of the inferior frontal gyrus was consistent with those observed in previous studies aiming at studying the involvement of this region in the generation of LSF-based predictions about the visual input (Kauffmann, Bourgin, et al., 2015; Bar et al., 2006). Coordinates of the peak of activation (33x, 11y, 2z in the right hemisphere and −33x, 23y, 2z in the left hemisphere) were close to the one observed in Kauffmann, Bourgin, et al. (2015) in the right hemisphere (two peaks: 29x, 25y, −36z and 37x, 32y, 2z). The similarity of activation between these two studies suggests that there would be common prediction mechanisms based on the rapid analysis of LSF, whether these come from the central visual field (Kauffmann, Bourgin, et al., 2015) or from the peripheral visual field (Experiment 2). Our activation of the inferior frontal gyrus was also close to the one observed by Bar et al. (2006): −36x, 23y, −14z. Unfortunately, the low temporal resolution of fMRI did not allow to investigate the time point of the categorization process at which the inferior frontal gyrus is involved. In our fMRI experiment, the hypothetical role of the inferior frontal gyrus in the generation of predictions is based on a set of coherent research results, which make it possible to consider this frontal region as playing an important role in visual recognition. In this context, our interpretation is that the low spatial resolution information extracted in peripheral vision could be rapidly conveyed to the inferior frontal gyrus to generate predictions about the category of the peripheral scene (high-level abstract information). Predictions were either relevant (semantically congruent condition) or irrelevant (semantically incongruent condition) for the categorization of the central scene. In the case of an irrelevant prediction, the predictive signal compared to the ascending information in the central scene would lead to a conflict or prediction error. The resolution of this conflict (i.e., disentangling between alternative responses) and/or the updating of predictions based on the processing of ascending information from the central scene would thus result in increased activation of the inferior frontal gyrus. This interpretation would be also consistent with the known role of inferior frontal cortex in monitoring response conflict resolution (e.g., Nee, Wager, & Jonides, 2007). Using hybrid images, Kauffmann, Bourgin, et al. (2015) observed that the semantic interference effect increased concomitantly the activation of the occipito-temporal cortex (fusiform and parahippocampal gyri). An additional analysis of functional connectivity showed that the interference effect increased the functional connectivity between the fusiform gyrus and the inferior frontal gyrus (at the level of the orbitofrontal cortex). Authors postulated that the predictive signal would be compared to the ascending HSF information in the occipito-temporal cortex, leading to a prediction error and to greater exchanges between the occipito-temporal and the orbitofrontal cortex. In the present experiment, we failed to observe significant activation of the occipito-temporal cortex. However, the use of a more liberal statistical threshold (p <.005, uncorrected)

Uncorrected Proof

reveals an activation of the left fusiform gyrus when the two scenes shared the same physical properties ($-36x$, $-52y$, $-16z$, AB 37, $k = 22$, $t = 3.97$). The absence of significant activation in these regions may be because of the ease of our behavioral task, in comparison to studies using hybrid images (Kauffmann, Bourgin, et al., 2015).

Interestingly, we observed that the semantic interference effect was strengthened when the two scenes shared similar physical properties. This result suggests that the physical properties of the peripheral scene are integrated in the recognition process of the central scene. More precisely, in the theoretical framework proposed above, predictions would be also based on the processing of physical properties in the peripheral scene (such as information on the content of spatial frequencies and orientation or on the spatial configuration). Thus, when a semantically incongruent peripheral scene leads to erroneous predictions about the category of the central scene, the resolution of the conflict should be easier if the physical information also leads to an erroneous prediction (peripheral scene PhyDis). On the contrary, a peripheral scene for which physical properties are similar to the one of the central scene could lead the visual system to believe that the prediction is relevant. In such a case, this would result in an additional conflict, increasing the activation of the inferior frontal gyrus.

Our experiment therefore brings another experimental evidence of the role of the inferior frontal gyrus in visual perception and, in particular, in the processing of low resolution information to be used for predictive mechanisms. Despite a growing body of evidence allowing to interpret the involvement of this region in predictive visual processing, its precise causal role, as well as the type of information that it represents, has yet to be explored. Moreover, influential predictive coding theories of visual processing do not include the need for such a cortical hub to trigger predictions. For instance, hierarchical predictive coding models based on the proposition of Rao and Ballard (1999; see also Spratling, 2017; Huang & Rao, 2011) state instead that neuronal activity is predicted at each stage of visual processing, as a result of extra classical receptive fields mechanisms and feedback loops from adjacent higher areas. The idea of a cortical hub for prediction in the place of the inferior frontal gyrus has now to be theoretically conciliated with more classical views of predictive coding.

Finally, when contrasting the incongruent condition to the congruent condition of PhyDis scenes ([congruent-PhyDis > incongruent-PhyDis] contrast), we observed activation within the right occipital and temporal cortices. Occipito-temporal activations were also observed when we assessed the physical similarity effect for congruent scenes ([congruent-PhyDis > congruent-PhySim] contrast). In both contrasts, the activation is driven by the presentation of two scenes of the same category (e.g., two natural scenes) but whose spatial arrangement and spectral properties are different (e.g., a forest scene with a lot of energy in HSF and in the vertical orientations in central vision vs. a beach scene with large uniform blobs in LSF in peripheral vision). Again, in the theoretical framework proposed above, this result could be interpreted as reflecting the fact that, even when the two scenes are semantically congruent, the physical properties of the peripheral scene may still be used to generate predictions about the physical properties of the category of the central scene. When erroneous (i.e., PhyDis condition), these predictions would lead to a prediction error: The semantic content of the central scene would be accurately predicted, but its physical properties would not correspond to what was expected for this category based on the peripheral scene. This would result in a conflict and greater recruitment of occipito-temporal regions to disentangle between alternative interpretations, suggesting the role of these regions in coding prediction error related to the physical properties of scenes.

Conclusion

These two experiments showed that information in peripheral vision is automatically processed when categorizing scenes in central vision and that it influences this categorization. We interpreted these results in the context of predictive models of visual recognition in which the inferior frontal gyrus would be part of a cortical system that manages sensory predictions. Indeed, predictions would rely on low-level visual information related to the physical characteristics of the scene (amplitude spectrum, spatial configuration). This was suggested by the larger semantic interference effect in the condition of physical similarity, but also the observation of impaired categorization of the central scene when it was semantically congruent to the peripheral scene, but PhyDis. This study leads us to question the nature of the physical characteristics used to activate the predictions. Is this information contained in the amplitude spectrum? Or is it rather spatial information such as the configuration of the scene described by the arrangement of the blobs? Roux-Sibilon et al. (2019) recently observed that the interference effect of a scene background on the categorization of an object in central vision disappears when the amplitude spectrum of the scene background was preserved, but the spatial configuration was altered (phase scrambling of scene images). This result thus suggests that information contained in the amplitude spectrum is not necessarily sufficient to trigger predictions and that phase information, conveying the spatial configuration of the scene, may be critical to trigger peripheral predictions. Using stimuli that take into account the cortical magnification factor, future studies should explicitly manipulate the physical characteristics of peripheral vision that could ensure this function.

Uncorrected Proof

Acknowledgments

This work was performed on the IRMaGe platform member of France Life Imaging network (grant ANR-11-INBS-0006) and

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

Reprint requests should be sent to Carole Peyrin, Laboratoire de Psychologie et NeuroCognition (LPNC), CNRS UMR 5105 - Université Grenoble Alpes, BSHM - 1251 Av Centrale CS40700, 38058 Grenoble Cedex 9, France, or via e-mail: carole.peyrin@univ-grenoble-alpes.fr.

Author Contributions

Carole Peyrin: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Visualization; Writing–Original draft; Writing–Review & editing. Alexia Roux-Sibilon: Formal analysis; Methodology; Visualization; Writing–Original draft; Writing–Review & editing. Audrey Trouilloud: Investigation; Methodology; Project administration. Sarah Khazaz: Investigation; Methodology; Project administration. Malena Joly: Investigation; Methodology; Project administration. Cédric Pichat: Formal analysis; Investigation; Methodology; Software. Muriel Boucart: Conceptualization; Writing–Original draft; Writing–Review & editing. Alexandre Krainik: Funding acquisition; Investigation; Project administration; Resources. Louise Kauffmann: Conceptualization; Methodology; Software; Supervision; Writing–Original draft; Writing–Review & editing.

Funding Information

Carole Peyrin, Agence Nationale de la Recherche (http://dx.doi.org/10.13039/501100001665), grant numbers: ANR-11-INBS-0006, ANR-15-IDEX-02.

Uncorrected Proof

REFERENCES

Bar, M. (2003). A cortical mechanism for triggering top–down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*, 600–609. **DOI:** https://doi.org/10.1162/089892903321662976, **PMID:** 12803970

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*, 280–289. **DOI:** https://doi.org/10.1016/j.tics.2007.05.005, **PMID:** 17548232

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top–down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 449–454. **DOI:** https://doi.org/10.1073/pnas.0507062103, **PMID:** 16407167, **PMCID:** PMC1326160

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. **DOI:** https://doi.org/10.18637/jss.v067.i01

Boucart, M., Moroni, C., Thibaut, M., Szaffarczyk, S., & Greene, M. (2013). Scene categorization at large visual eccentricities. *Vision Research*, *86*, 35–42. **DOI:** https://doi.org/10.1016/j.visres.2013.04.006, **PMID:** 23597581

Curcio, C. A., & Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology*, *300*, 5–25. **DOI:** https://doi.org/10.1002/cne.903000103, **PMID:** 2229487

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *Journal of Physiology*, *159*, 203–221. **DOI:** https://doi.org/10.1113/jphysiol.1961.sp006803, **PMID:** 13883391, **PMCID:** PMC1359500

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception?. *Trends in Cognitive Sciences*, *22*, 764–779. **DOI:** https://doi.org/10.1016/j.tics.2018.06.002, **PMID:** 30122170

De Valois, R., Albrecht, D., & Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*, 545–559. **DOI:** https://doi.org/10.1016/0042-6989(82)90113-4

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*, 553–567. **DOI:** https://doi.org/10.1017/S0140525X00031575

Friston, K. (2005). A theory of cortical response. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *360*, 815–836. **DOI:** https://doi.org/10.1098/rstb.2005.1622, **PMID:** 15937014, **PMCID:** PMC1569488

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. *Neuroimage*, *10*, 607–619. **DOI:** https://doi.org/10.1006/nimg.1999.0498, **PMID:** 10547338

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. **DOI:** https://doi.org/10.1167/18.9.9, **PMID:** 30208433

Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 580–593. **DOI:** https://doi.org/10.1002/wcs.142, **PMID:** 26302308

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*, 2394–2405. **DOI:** https://doi.org/10.1162/jocn_a_00861, **PMID:** 26244724

Kauffmann, L., Chauvin, A., Guyader, N., & Peyrin, C. (2015). Rapid scene categorization: Role of spatial frequency order, accumulation mode and luminance contrast. *Vision Research*, *107*, 49–57. **DOI:** https://doi.org/10.1016/j.visres.2014.11.013, **PMID:** 25499838

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. **DOI:** https://doi.org/10.1016/j.bandc.2015.07.004, **PMID:** 26232267

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 1–14. **DOI:** https://doi.org/10.3389/fnint.2014.00037, **PMID:** 24847226, **PMCID:** PMC4019851

Kauffmann, L., Roux-Sibillon, 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*, 853–867. **DOI:** https://doi.org/10.1080/13506285.2017.1347590

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26. **DOI:** https://doi.org/10.18637/jss.v082.i13

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*, 13232–13240. **DOI:** https://doi.org/10.1523/JNEUROSCI.3481-07.2007, **PMID:** 18045917, **PMCID:** PMC6673387

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision*, *9*, 6. **DOI:** https://doi.org/10.1167/9.10.6, **PMID:** 19810787

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 180° visual field. *Journal of Vision*, *19*, 1–21. **DOI:** https://doi.org/10.1167/19.5.15, **PMID:** 31100131

Lukavsky, J. (2019). Scene categorization in the presence of a distractor. *Journal of Vision*, *19*, 1–11. **DOI:** https://doi.org/10.1167/19.2.6, **PMID:** 30735564

Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental Brain Research*, *227*, 367–377. **DOI:** https://doi.org/10.1007/s00221-013-3517-1, **PMID:** 23604577

Musel, B., Chauvin, A., Guyader, N., Chokron, S., & Peyrin, C. (2012). Is coarse-to-fine strategy sensitive to normal aging? *PLoS One*, *7*, e38493. **DOI:** https://doi.org/10.1371/journal.pone.0038493, **PMID:** 22675568, **PMCID:** PMC3366939

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 1–17. **DOI:** https://doi.org/10.3758/CABN.7.1.1, **PMID:** 17598730

Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In K. S. Rockland, J. H. Kaas, & A. Peters (Eds.), *Extrastriate cortex in primates* (pp. 205–241). Boston: Springer. **DOI:** https://doi.org/10.1007/978-1-4757-9625-4_5

Nowak, L. G., Munk, M., Girard, P., & Bullier, J. (1995). Visual latencies in areas v1 and v2 of the macaque monkey. *Visual Neuroscience*, *12*, 371–384. **DOI:** https://doi.org/10.1017/S095252380000804X, **PMID:** 7786857

Oliva, A. (2005). Gist of a scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 251–256). Burlington, MA: Elsevier Academic. **DOI:** https://doi.org/10.1016/B978-012375731-9/50045-8

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175. **DOI:** https://doi.org/10.1023/A:1011139631724

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, *21*, 147–160. **DOI:** https://doi.org/10.1068/p210147, **PMID:** 1513664

Pelli, D. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, *18*, 445–451. **DOI:** https://doi.org/10.1016/j.conb.2008.09.008, **PMID:** 18835355, **PMCID:** PMC3624758

Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *Neuroimage*, *186*, 103–112. **DOI:** https://doi.org/10.1016/j.neuroimage.2018.10.086, **PMID:** 30403971

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., et al. (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*, 2768–2780. **DOI:** https://doi.org/10.1162/jocn.2010.21424, **PMID:** 20044901

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*, 79–87. **DOI:** https://doi.org/10.1038/4580, **PMID:** 10195184

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*, 7. **DOI:** https://doi.org/10.1167/19.14.7, https://doi.org/10.31234/osf.io/fp4rk

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

Skottun, B., De Valois, R., Grosof, D., Movshon, J. A., Albrecht, D. G., & Bonds, A. B. (1991). Classifying simple and complex cells on the basis of response modulation. *Vision Research*, *31*, 1079–1086. **DOI:** https://doi.org/10.1016/0042-6989(91)90033-2, **PMID:** 1909826

Spratling, M. W. (2017). A hierarchical predictive coding model of object recognition in natural images. *Cognitive Computation*, *9*, 151–167. **DOI:** https://doi.org/10.1007/s12559-016-9445-1, **PMID:** 28413566, **PMCID:** PMC5371651

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391–412. **DOI:** https://doi.org/10.1088/0954-898X_14_3_302, **PMID:** 12938764

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*, 190–198. **DOI:** https://doi.org/10.1111/nyas.12680, **PMID:** 25728836

Trouilloud, A., Kauffmann, L., Roux-Sibilon, A., Rossel, P., Boucart, M., Mermillod, M., et al. (2020). Rapid scene categorization: From coarse peripheral vision to fine central vision. *Vision Research*, *170*, 60–72. **DOI:** https://doi.org/10.1016/j.visres.2020.02.008, **PMID:** 32259648

Wässle, H., Grünert, U., Röhrenbeck, J., & Boycott, B. B. (1990). Retinal ganglion cell density and cortical magnification factor in the primate. *Vision Research*, *30*, 1897–1911. **DOI:** https://doi.org/10.1016/0042-6989(90)90166-I

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160–168. **DOI:** https://doi.org/10.1016/j.tics.2011.02.005, **PMID:** 21420894, **PMCID:** PMC3070834

Uncorrected Proof

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

During the preparation of your manuscript, the questions listed below arose. Kindly supply the necessary information.

1. Please spell out both occurrences of HSF.
2. Please verify the changes made in « «(Congruence, Physical Similarity, Congruence × Physical Similarity, Congruence × Physical Similarity × Visual Field).” Results section.
3. Please spell out TFE.
4. Please confirm if Funding Information are complete and accurate.
5. Please provide details for the following unlisted reference(s):
   Friston et al., 1995

END OF ALL QUERIES

Uncorrected Proof