Influence of peripheral vision on object categorization in central vision

Alexia Roux-Sibilon

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Audrey Trouilloud

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Louise Kauffmann

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France
University Grenoble Alpes, CNRS, Grenoble INP,
GIPSA-lab, Grenoble, France

Nathalie Guyader

University Grenoble Alpes, CNRS, Grenoble INP,
GIPSA-lab, Grenoble, France

Martial Mermillod

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Carole Peyrin

University Grenoble Alpes,
University of Savoie Mont Blanc, CNRS, LPNC,
Grenoble, France

Predictive models of visual recognition state that predictions based on the rapid processing of low spatial frequencies (LSF) may guide the subsequent processing of high spatial frequencies (HSF). While the HSF signal necessarily comes from central vision, most of the LSF signal comes from peripheral vision. The present study aimed at understanding how LSF in peripheral vision may be used to generate predictive signals that guide visual processes in central vision. In two experiments, participants performed an object categorization task in central vision while a semantically congruent or incongruent scene background was displayed in peripheral vision. In Experiment 1, results showed a congruence effect when the peripheral scene was displayed before the object onset. In Experiment 2, results showed a congruence effect only when the peripheral scene was intact, thus carrying a semantic meaning, but not when it was phase-scrambled, thus carrying only low-level information. The study suggests that the low resolution of peripheral vision facilitates the processing of foveated objects in the visual scene, in line with predictive models of visual recognition.

Introduction

Visual recognition in humans is considerably efficient and fast. Complex stimuli such as objects and natural scenes are robustly processed and categorized despite their infinite variability. Data on the functional neuroanatomy of visual pathways (Van Essen & DeYoe, 1995) and neurophysiological recordings in primates (De Valois, Albrecht, & Thorell, 1982; Shams & Von Der Malsburg, 2002; Shapley & Lennie, 1985) suggest that the visual system is able to rapidly extract the low spatial frequencies (LSF) from a visual scene through fast magnocellular pathways. LSF signals convey coarse information about the global shape and

structure of the scene. The processing of this information precedes the processing of a high spatial frequency (HSF) signal, which conveys finer information about the scene such as edges and object details through slowest parvocellular pathways. Accordingly, theories of visual perception have proposed that visual information is integrated in a *coarse-to-fine* manner (Bar, 2003; Hegdé, 2008; Kauffmann, Ramanoël, & Peyrin, 2014; Schyns & Oliva, 1994).

In this theoretical framework, it was also hypothesized that the brain uses the rapidly available LSF to predict visual inputs, especially within the orbitofrontal cortex (Bar & Aminoff, 2003; Kauffmann et al., 2014; Kveraga, Boshyan, & Bar, 2007; Peyrin et al., 2010). The predictive signal would be back-projected, via top-down connections, to occipito-temporal visual areas to guide bottom-up processes. Predictions could then influence the subsequent processing of HSF. As experimental evidence, a combined magnetoencephalography (MEG) and functional magnetic resonance imaging (fMRI) study (Bar et al., 2006) demonstrated earlier activations in the orbitofrontal cortex than in the occipito-temporal cortex during recognition of object images, this early activity depending on the presence of LSF in the image. Recent fMRI studies investigating the effective connectivity between these regions using dynamic causal modeling showed that a magnocellular signal (e.g., achromatic and low-luminance contrast drawings, LSF-filtered scenes) increases the connectivity strength from the orbitofrontal cortex to the inferotemporal cortex (Kauffmann, Chauvin, Pichat, & Peyrin, 2015; Kveraga et al., 2007). Petras, ten Oever, Jacobs, and Goffaux (2019) used a classifier trained to discriminate between EEG scalp patterns evoked by LSF inputs and EEG scalp patterns evoked by HSF inputs in order to tease apart LSF and HSF contribution to the neural response evoked by broadband stimuli. They then tested the classifier on EEG scalp patterns evoked by intact and phase-scrambled broadband faces. Results showed an early LSF dominance followed by a reduced HSF dominance in response to intact, but not to scrambled, face stimuli. Given the fact that LSF was informative about HSF only in intact images, the reduction in the late contribution of HSF suggests that the first LSF parsing elicits a robust representation of the face that subsequently reduces (unnecessary) HSF processing.

However, these studies often use small stimuli displayed in central vision and therefore miss the selectivity of peripheral vision for LSF processing. Indeed, the processing of spatial frequencies is nonhomogeneous throughout the visual field (Curcio & Allen, 1990; Curcio, Sloan, Kalina, & Hendrickson, 1990). In the retina, the density of midget ganglion cells, tuned to HSF, is greater in the fovea, while the density of parasol ganglion cells, tuned to LSF, increases with retinal eccentricity. Therefore, the receptive fields in peripheral vision are too large to capture HSF signals. The counterpart is that the visual system has a considerable amount of LSF information rapidly available in peripheral vision that can contribute to the generation of predictive signals. In this view, the processing of details in central vision could benefit from predictive signals originating from both the overlapping central LSF signal and the eccentric LSF signal.

The objective of the present study is to test whether LSF information in peripheral vision can be used to generate predictive signals that guide visual processes in central vision. In real-life conditions, observers often foveate relevant elements (e.g., a face during a social interaction, an object while performing an action) while these are embedded within a coherent context encompassing the whole visual field. Central vision is thus more suited for object perception and peripheral vision for scene perception. This distinction is found in the visual cortex where object-selective areas (in the lateral occipito-temporal cortex) respond more strongly to central rather than peripheral visual input, while scene-selective areas (in the medial occipito-temporal cortex) respond more strongly to peripheral rather than central visual input (Arcaro, McMains, Singer, & Kastner, 2009; Baldassano, Fei-Fei, & Beck, 2016; Levy, Hasson, Avidan, Hendler, & Malach, 2001; Malach, Levy, & Hasson, 2002). In the present study, we asked participants to categorize an object in central vision while a semantically related (predictive/congruent) or unrelated (nonpredictive/incongruent) scene background was presented in peripheral vision. We did not filter out the spatial frequency content from the scene, since peripheral vision acts as a natural low-pass filter. We expected that LSF information available in the broadband scene displayed in peripheral vision would influence the explicit categorization of the object in central vision (resulting in better performance in congruent than in incongruent trials). A large number of studies investigated the interactions between objects and context (Bar & Ullman, 1993; Boucart, Moroni, Szaffarczyk, & Tran, 2013; Davenport, 2007; Davenport & Potter, 2004; Joubert, Fize, Rousselet, & Fabre-Thorpe, 2008; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Rémy et al., 2013; Sun, Simon-Dack, Gordon, & Teder, 2011). Typically, categorization performance is better when objects are within a congruent rather than incongruent context. A congruent context can also facilitate object processing in patients with age-related macular degeneration, a retinal disease causing central vision loss due to the destruction of macular photoreceptors (Boucart et al., 2013). However, because the distinction between central and peripheral vision was not the aim of these studies, authors did not control either the position of the object on the background scene, or the retinal size

of the object and the scene (resulting in rather small scenes and big objects).

Here, we examined how scene information influences object categorization when object information is restricted to central vision and scene information to peripheral vision. We manipulated the time during which the scene was processed before object onset (Experiment 1). We expected that if information in peripheral vision is used to generate a predictive signal that shapes visual processing in central vision, the peripheral influence on object categorization (i.e., the congruence effect) would be greater when peripheral information can be accumulated and result in a sharper representation of the scene—that is, when the scene is presented longer before object onset. To test whether low-level aspects of scene processing are sufficient to influence visual processing in central vision, or whether achieving a semantic representation of the scene is needed (Bar, 2003; Kauffmann, Bourgin, Guyader, & Peyrin, 2015), we presented the scenes in two versions: intact and phase-scrambled (Experiment 2). The phase-scrambled version of the scene is made by disrupting the phase of the scene image while preserving low-level (spatial frequency and orientation distribution) information. We expected that if the semantic representation of the scene is used to generate predictions, the peripheral scene would influence object categorization when it is intact more than when its phase is scrambled.

Experiment 1

Methods

Participants

Fifty right-handed participants (43 women; $M \pm SD = 21 \pm 4$ years, range: 18–26 years) with normal or corrected-to-normal vision participated in the experiment. Most of them were psychology students who received course credits for their participation. They gave their 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 (Grenoble-Alps Research Ethics Committee, Community University Grenoble Alps, IRB00010290).

Stimuli

Stimuli were images consisting of a combination of an object and a scene background. All images used to create the stimuli (scenes and objects) were downloaded from the image bank Pixabay (https://pixabay.com/) under Creative Commons Zero license, or from Google Images searches of copyright-free pictures. The scenes used as backgrounds were photographs representing either outdoor natural landscapes (e.g., savannah, field, park ice), or indoor views (e.g., living room, corridor). They were converted into 256-level grayscale and rescaled to 1024×768 pixels. Stimuli were displayed on a 30-in. monitor (Dell Ultrasharp) with a resolution of 2560×1600 pixels and a refresh rate of 60 Hz. We estimated the gamma function of the monitor by measuring the luminance values of the display for different values of uniform gray-level stimuli generated in MATLAB (MathWorks, Natick, MA) with a calibration tool (Spyder5ELITE, Datacolor, Rotkreuz, Switzerland). Based on the estimated function, we gamma-corrected luminance values of each scene to linearize indirectly the gamma function. Mean luminance and root mean square (RMS) contrast (standard deviation of the luminance) of the scenes were then equalized to obtain a mean luminance of 0.51 for luminance values ranging from 0 to 1 (i.e., mean luminance of 130 on a gray-level scale) and a mean RMS of 0.24 (i.e., 61 on a gray-level scale). These values correspond to the average values of luminance and RMS contrast of all scene images (landscapes and indoor views). Objects were photographs of animals (e.g., cow, penguin) and pieces of furniture (e.g., armchair, coffee table) that were cut out of natural images using Adobe Photoshop. Each object was rescaled to 128 pixels on its larger side (either width or height, depending on the object).

We created 20 groups of four object-scene stimuli (80 stimuli) by combining one scene background with one object. The four combinations resulted in two congruent (animal + outdoor context and furniture + indoor context) and two incongruent (animal + indoor context and furniture + outdoor context) object-scene associations (Figure 1). Scenes and objects forming congruent associations were combined based on their real-world congruency (e.g., cow + field, penguin + park ice). Each object was pasted on the corresponding congruent and incongruent backgrounds, at the center on the horizontal axis, but in the lower part on the vertical axis (centered at 576 pixels from the top of the image) in order to respect the usual position of an object in a photograph. The object was isolated from the background scene by a circular gray patch (diameter of about 250 pixels) whose edges were smoothly blended into the scene. This allowed us to restrict contextual information to peripheral vision only, and to control for possible local contrast effects on the perception of the object. Stimuli can be download from https://osf.io/mfhx5/.

Procedure

The experiment was programmed on E-Prime 2.0 software (E-Prime Psychology Software Tools Inc.,
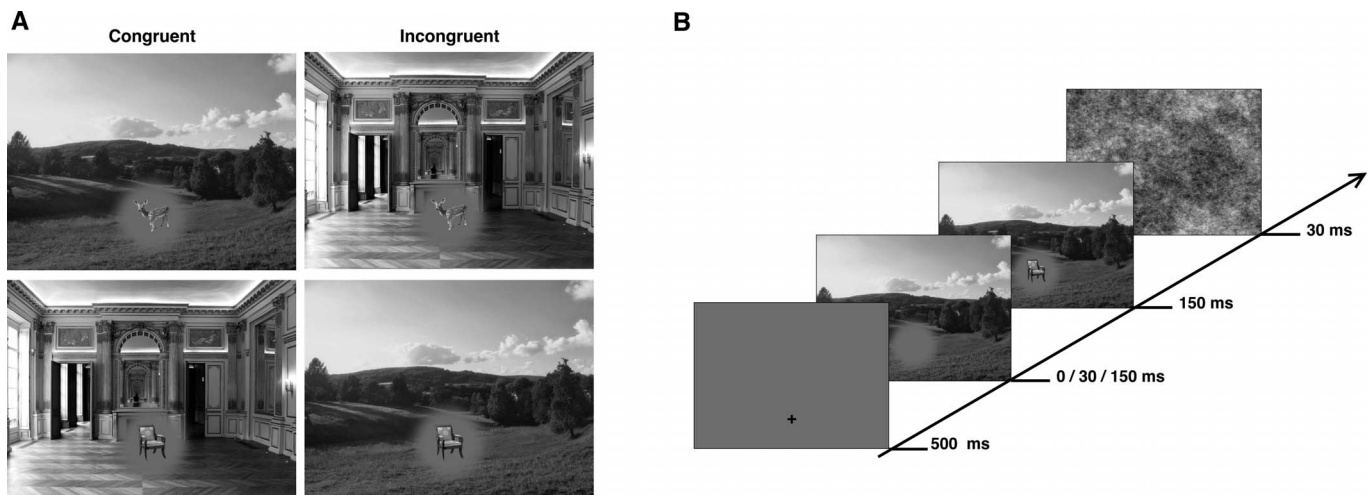
Figure 1. (A) Example of a group of four scene + object stimuli used in Experiment 1. (B) Schematic of the experimental procedure. In the SOA-0 condition, the scene + object stimulus appeared directly, without any pre-exposure of the peripheral scene.

Pittsburgh, PA). Participants viewed stimuli at a distance of 50 cm. At this distance, the background scenes subtended 24° of visual angle horizontally and 18° vertically. The larger side of the objects (either the width or the height) subtended 3° and the gray circular patch subtended about 6° in diameter and was centered at ~13° from the top of the image. The participant's head was supported by a chinrest in order to maintain distance.

Participants were instructed to maintain their fixation on a little white fixation cross, centered on the screen location at which the object was presented (i.e., at the center of the screen horizontally and in the lower part vertically). Each trial began with the fixation cross presented for 500 ms, followed by a stimulus sequence, and by a backward mask (1/f noise) for 30 ms. Interstimulus interval was of 2,500 ms. The stimulus sequence depended on the experimental condition. In the stimulus onset asynchrony (SOA)-0 condition, the scene-object stimulus was presented for 150 ms. In the SOA-30 condition, the scene-object stimulus was also presented for 150 ms, but was preceded by the scene image alone (including a circular gray patch masking central visual information) for 30 ms. Similarly, in the SOA-150 condition, the scene-object stimulus was presented for 150 ms, and was preceded by the scene image alone for 150 ms. The three scene SOA conditions were blocked and block order was counterbalanced across participants by a Latin square procedure. Participants had to categorize the object as animal or furniture by pressing two keys on a keyboard. They received no special instruction about the background, but if they questioned it, they were told that the background did not matter. Response keys were counterbalanced across participants. Accuracy and response time (RT) were recorded at each trial. There were 240 trials (2 Congruence × 3 Scene-SOA × 20 Stimuli Groups × 2 Object Categories). Participants could pause between scene SOA blocks, and the experiment lasted about 20 min. Prior to completing the experimental trials, participants underwent a short training session (12 trials) to be familiarized with the stimuli and the task.

Results

For each participant, correct RTs (in ms) were log transformed and then trimmed by removing trials for which RT was inferior or superior to the condition average ± 2.5 SDs (1.76% of trials were excluded). Mean correct RTs (mRT), mean log transformed correct RTs (mLog[RT]) and mean correct response rates (mCR), with standard deviations, for each experimental condition are reported in Table 1. We conducted two repeated measures ANOVA on mCR and mLog(RT) with congruence (congruent or incongruent stimuli) and scene SOA (SOA-0, SOA-30, SOA-150) as within-subject factors. Further pairwise comparisons were tested with two-tailed paired-samples t tests, and trend analyses with polynomial contrasts, using a Bonferroni adjustment of alpha level to correct for multiple tests ($0.05/8 = 0.006$ for the eight pairwise tests and trend analyses performed). Effect size was estimated by calculating Cohen's d for within subject design. Statistical analyses were conducted in R.

Analysis of mCR showed that mCR was larger in the congruent condition ($M \pm SD$: 0.96 ± 0.05) than in the incongruent condition (0.95 ± 0.05; $F[1, 49] = 6.52$, $p =$

	Congruent SOA-0	Incongruent SOA-0	Congruent SOA-30	Incongruent SOA-30	Congruent SOA-150	Incongruent SOA-150
mCR	0.96 ± 0.20	0.96 ± 0.20	0.97 ± 0.18	0.96 ± 0.20	0.96 ± 0.19	0.95 ± 0.23
mRT	567 ± 159	568 ± 158	539 ± 149	543 ± 140	527 ± 135	547 ± 130
mLog(RT)	2.74 ± 0.10	2.74 ± 0.10	2.72 ± 0.10	2.72 ± 0.10	2.71 ± 0.10	2.73 ± 0.09

Table 1. Mean correct response rate (mCR), mean correct response times in milliseconds (mRT), and mean log transformed correct response times (mLog[RT]), with standard deviations, for each experimental condition (congruence, scene SOA).

0.014, $d = 0.36$). The main effect of Scene-SOA was not significant ($F[2, 98] = 1.37$, $p = 0.258$), and this factor did not interact with the congruence of stimuli ($F[2, 98] = 2.22$, $p = 0.114$; Figure 2). It should be noted that the global mCR was high (0.96 ± 0.20), with distributions of individual observations reflecting a ceiling effect.

Analysis of mLog(RT) showed that participants were faster to categorize objects in congruent stimuli (2.723 ± 0.102) than in incongruent stimuli (2.731 ± 0.097; $F[1, 49] = 39.44$, $p < 0.001$, $d = 0.88$). The main effect of scene SOA was also significant ($F[1, 98] = 14.40$, $p < 0.001$, $d = 0.54$). Both linear and quadratic trends were significant (linear contrast: $t[49] = 4.22$, $p < 0.001$, $d = 0.60$; quadratic contrast: $t[49] = 2.83$, $p = 0.007$, $d = 0.40$), indicating that participants were faster to categorize the objects in both the SOA-30 (2.721 ± 0.010) and SOA-150 (2.719 ± 0.095) than in the SOA-0 condition (2.741 ± 0.103). Importantly, scene SOA interacted with the congruence of the stimuli ($F[1, 98] = 12.4$, $p < 0.001$, $d = 0.50$; Figure 2). We tested the difference between the congruent and incongruent trials for each scene SOA condition. It was significant for the SOA-150 condition (congruent: 2.710 ± 0.096; incongruent: 2.728 ± 0.092; $t[49] = 7.60$, $p < 0.001$, $d = 1.08$), but neither for the SOA-30 condition (congruent: 2.718 ± 0.104; incongruent: 2.723 ± 0.096; $t[49] = 2.68$, $p = 0.010$) nor for the SOA-0 condition (congruent: 2.740 ± 0.103; incongruent: 2.741 ± 0.102; $t[49] = 0.64$, $p = $

0.525). Post hoc trend analyses were conducted to examine the effect of scene SOA on congruent and incongruent stimuli separately. For the congruent condition, the linear trend was significant ($t[49] = 5.44$, $p < 0.001$, $d = 0.77$), but the quadratic trend was not significant ($t[49] = 1.89$, $p = 0.064$). This suggests that RT decreases linearly when scene SOA increases in the congruent condition. In the incongruent condition, the linear trend was not significant ($t[49] = 2.42$, $p = 0.019$) but the quadratic trend was ($t[49] = 3.36$, $p < 0.006$). We then tested for a linear modulation of the difference between incongruent and congruent trials by the scene SOA factor. The linear contrast was significant ($t[49] = 4.70$, $p < 0.001$, $d = 0.31$) and the quadratic contrast was not significant ($t[49] = 1.24$, $p = 0.223$), suggesting that the peripheral scene influence on object categorization increased linearly with longer scene SOA.

Results of Experiment 1 can be summarized following two main observations. First, contrary to our hypothesis, no effect of congruence was observed for the SOA-0 and SOA-30 conditions, suggesting that the representation of the scene was not robust enough to generate strong predictions in those conditions and thus influence the object categorization. Second, there was a congruence effect in the SOA-150 condition (on RTs). This effect suggests that when the scene background was sufficiently processed before the object onset, participants automatically processed information in peripheral
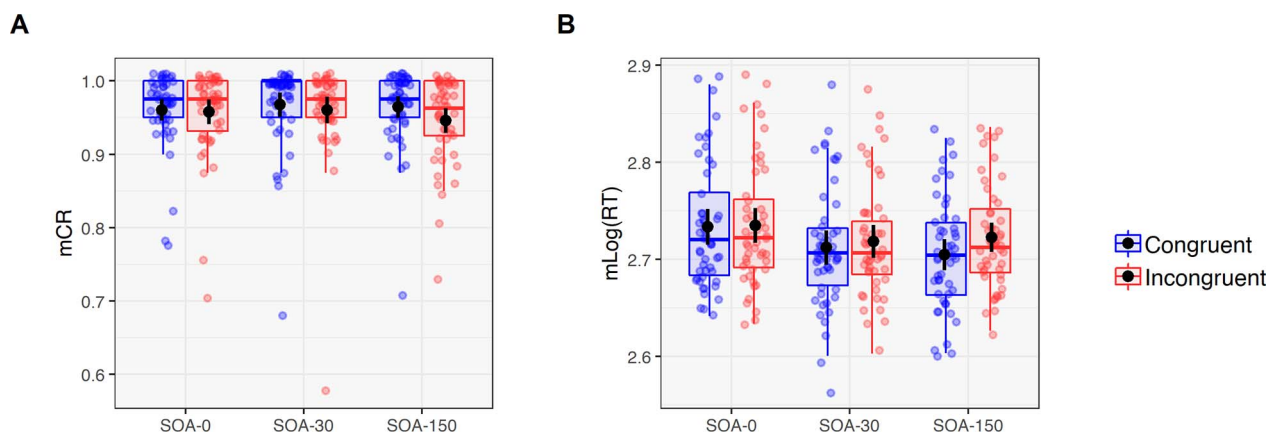

Figure 2. (A) Mean correct response rates (mCR) and (B) mean log transformed correct RTs (mLog[RT]) for the categorization of the object according to the congruence between the object and the scene in peripheral vision (congruent, incongruent) and the scene SOA (SOA-0, SOA-30, SOA-150). Black dots and error bars indicate means and 95% *CI*s. Color dots are individual observations (slightly jittered for better visualization).

vision although it was irrelevant to the task. This result is consistent with the hypothesis of predictive influences from peripheral vision. The effect, however, is rather small in practice, since the raw observed difference between the congruent and incongruent conditions was 20 ms. However, the effect seems to be robust and should be detected easily in subsequent experiments, as suggested by the large Cohen's *d* (1.08). Thus, results of Experiment 1 are consistent with the hypothesis of predictive influences from peripheral vision, but suggest that the influence is rather small. Moreover, the mechanisms underlying the congruence effect are not clear. First, as we did not test an experimental condition without a context, we cannot know whether object categorization was facilitated in the congruent condition or impaired in the incongruent condition. In addition, it is possible—at least in a certain proportion of trials—that the recognition of the scene category itself may have directly driven the response, without any feedback modulation on the processing of the object. An extreme example of such an "overhasty" response strategy would be a situation where a participant always responds *animal* when an outdoor scene is recognized and *furniture* when an indoor scene is recognized, regardless of the object. In this case, the proportion of correct responses would be 100% in the congruent condition and 0% in the incongruent condition, resulting in a maximum congruence effect. Considering this, the congruence effect measured in our experiment might not be entirely due to the *integration* of the peripheral visual input during the categorization of the object in central vision. Unfortunately, Experiment 1 did not allow us to dissociate an actual object categorization influenced by the scene from mere scene-consistent responses. Finally, we observed a ceiling effect of accuracy, suggesting that the design of this experiment made the task too easy, making it difficult to detect interactions on accuracy.

Experiment 2

We conducted a second experiment to understand better how peripheral vision influences central vision processing considering the limits described above. We used the same experimental paradigm as in Experiment 1, but the scenes were always presented 150 ms before the object onset (i.e., the condition for which we observed a significant congruence effect in Experiment 1). First, in order to rule out the possibility of a ceiling effect, we parametrically decreased the visibility of objects by manipulating a phase-scrambling parameter. Objects were presented at seven different levels of visibility. The parametric manipulation allowed us to map psychometric functions of mCR as a function of visibility levels. Another aim of this manipulation was

to promote the congruence effect. The rationale was that when the input is poor or ambiguous, the weight of predictions should be stronger compared to that of the inputs (Kok & de Lange, 2015). Therefore, information in peripheral vision should have a stronger influence when the central visual information is poor. In this context, studies have shown that predictive processes enhance visual perception when the visual stimulation is noisy or incomplete (e.g., Brandman & Peelen, 2017; Tang et al., 2018; Teufel, Dakin, & Fletcher, 2018; Wyatte, Curran, & O'Reilly, 2012). Secondly, in order to consider the direction of the congruence effect (facilitation in the congruent condition, hindrance in the incongruent condition, or both), we included a baseline condition. In this condition, we combined the object with a meaningless image background (1/f noise). Thirdly, in order to distinguish the part of the response based on the peripheral scene alone from the part of the response actually due to the influence of the peripheral scene on object categorization, we included trials for which no object was presented. In those trials, the object region was simply filled with 1/f noise. In that case, there was no correct response. However, since these trials were randomly embedded among object trials (some of them being of low visibility), participants did not realize the absence of objects. We tested if the tendency to rely on the scene when no object was present correlated with the congruence effect. A positive correlation would suggest that the congruence effect is partly due to a mere processing of the scene, without exerting feedback modulation during the processing of the object. In Experiment 2, we were also interested by the nature of the peripheral influence. In the peripheral visual field, the visual system extracts low-level visual features (spatial frequencies and orientations), whose processing allows the construction of high-order semantic representations. As can be seen in Figure 3, pieces of furniture and indoor scenes tend to have similar amplitude spectra. The energy is mainly distributed on vertical and horizontal, but also few oblique, orientations (due to the viewpoint perspective), ranging from the lowest to the highest spatial frequencies. In the same way, animals and outdoor scenes tend to have similar amplitude spectra, with energy more sparsely distributed throughout orientations, and mostly on the lowest spatial frequency range. It is thus possible that the influence of peripheral vision on central object categorization is only based on low-level visual features, rather than on a higher order semantic representation of the scene. Both influences are plausible. For example, predictive coding theories of vision (Friston & Stephan, 2007; Lee & Mumford, 2003; Rao & Ballard, 1999) propose that predictions flow between hierarchical areas within the visual cortex, where low-level aspects are represented. In predictive models of visual recognition (Bar, 2003; Kauffmann et
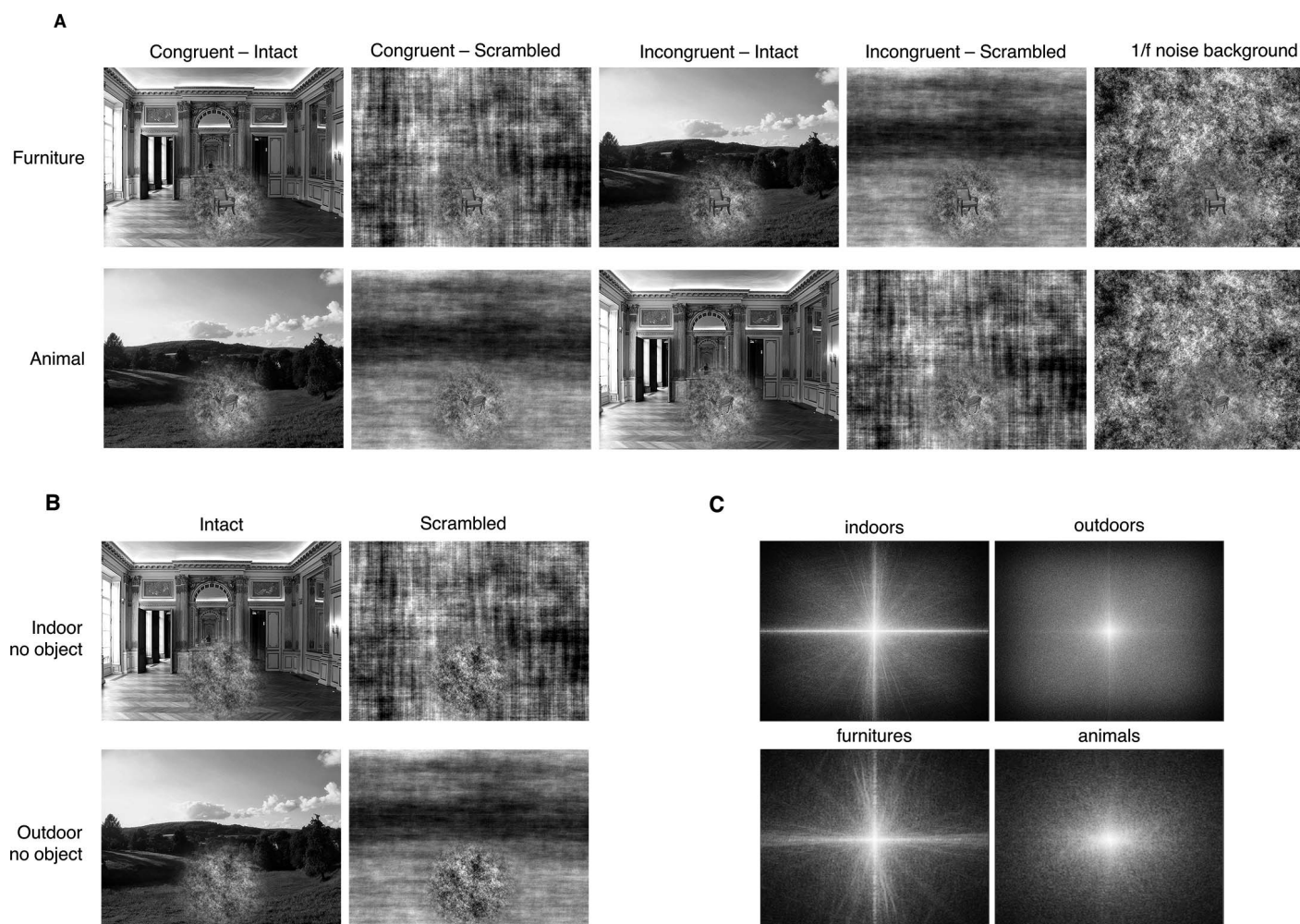
Figure 3. Example of a set of stimuli used in Experiment 2. (A) The object was a piece of furniture (first row) or an animal (second row) pasted on a 1/f noise and embedded in a congruent or incongruent scene background whose phase was either intact or scrambled, or in a meaningless 1/f noise background. (B) "Scene-alone" condition, in which no object was present in the 1/f noise. (C) The mean amplitude spectrum of indoor scenes is similar to that of furniture, while the mean amplitude spectrum of outdoor scenes is similar to that of animals. For illustration purposes, contrast and phase coherence of objects were slightly increased in the present figure. See Figure 4 for a zoom in on the object.

al., 2014; Kveraga et al., 2007; Peyrin et al., 2010), predictions are triggered in the inferofrontal cortex, where semantic aspects would be represented. To test for the influence of the two types of information (low-level and semantic) we manipulated the presence of a semantic content in the background scene. In the *intact* condition, we used the original scene image. In the *scrambled* condition, we suppressed the semantic information by scrambling the phase spectrum of the intact scene image via random permutation. This procedure is known to preserve orientation and spatial frequency content while preventing the processing of any semantic content (Goffaux et al., 2010; Woodhead, Wise, Sereno, & Leech, 2011). If the peripheral influence is due to low-level visual features only, we expected to observe an effect of congruence both in the intact and scrambled condition. On the contrary, if the peripheral influence involves semantic representations as well, we expected to observe a greater effect of congruence in the intact condition than in the scrambled condition.

Methods

Participants

Eighteen right-handed participants (14 women; $M \pm SD$: 21 ± 3 years; range: 19–30 years) who did not participated in Experiment 1 were included. They were psychology students who received course credits for their participation. The sample size was chosen based on a power analysis with estimated effect size of 1.08 (effect size of the congruence effect in the SOA-150 condition in Experiment 1) to achieve power of 0.99 at
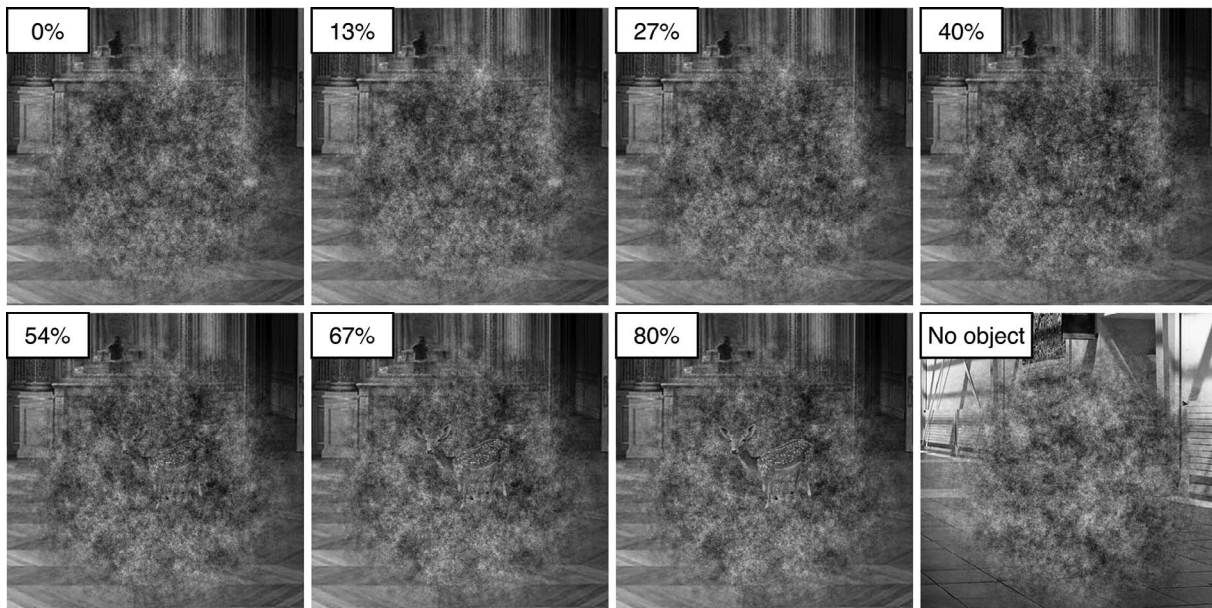
Figure 4. Portion of an indoor scene where participants fixated during the experiment, containing an object (a fawn) or no object. Participants viewed each object in seven versions, ranging from 0% to 80% of phase coherence, within 1/f noise. For illustration purposes, contrast and phase coherence of objects were slightly increased in the present figure.

alpha level of 0.05. All participants had normal or corrected-to-normal vision. They gave their informed written consent before participating in the study.

Stimuli

In the same way as for Experiment 1, we used 20 groups of four object-scene associations. Stimuli were created in MATLAB. We first gamma-corrected each object and pasted it on a 1/f noise image. We varied the visibility of the object by manipulating a phase-scrambling parameter. For each object, seven versions of the object-on-1/f image were thus created by parametrically adding coherence in the phase structure of the object, using Ales, Farzin, Rossion, and Norcia's (2012) MATLAB function. This function interpolates the phase of the Fourier transformation between the object-on-1/f image and a noise image of the same size. The seven versions ranged from 0% to 80% of phase coherence (100% being the original object on 1/f noise), increasing linearly in 13.33% steps (Figure 4). We chose to limit maximum coherence to 80% in order minimize the ceiling performance observed in Experiment 1 with fully visible objects. Each of the seven object-on-1/f images was pasted on five different backgrounds (intact-congruent, intact-incongruent, scrambled-congruent, scrambled-incongruent, and noise-baseline). Intact scene images were gamma-corrected and equalized to obtain a mean luminance of 0.51 and mean RMS contrast of 0.24. Scrambled scene images were created by scrambling the phase of the intact scenes in the Fourier domain via random permutation. The noise background for the baseline condition was a 1/f noise image of the same size as the scenes (i.e., 1024 × 768 pixels, or 24° × 18°). Object-on-1/f images were then progressively blended to the different backgrounds with a circular patch (Figure 3). Given this procedure, the circular patch had a noisy structure, allowing a more natural blending between the object and the scene compared to Experiment 1 where the circular patch was uniformly gray. We also created stimuli in which the scene backgrounds (intact and scrambled) were presented without an object (the part of the image in central vision was only filled with 1/f noise).

Procedure

The experiment was programmed on E-Prime 2.0 (E-Prime Psychology Software Tools Inc., Pittsburgh, PA). Stimuli were displayed on the same 30-in. monitor as in Experiment 1. At a viewing distance of 50 cm, the scenes sized 24° of visual angle horizontally and 18° vertically, the larger side of the objects sized 3°, and the circular patch sized about 8°. The participant's head was supported by a chinrest in order to maintain distance. They were instructed to maintain their fixation on a little white fixation cross, centered on the screen location at which the object was presented (i.e., at the center of the screen horizontally and in the lower part vertically). The fixation cross was displayed for a random duration from 500 to 1,500 ms (on average ~1,000 ms) in order to disrupt the predictability of trials' rhythm. Each trial began with the fixation cross, followed by a first stimulus (always a background

	Congruent intact scene	Incongruent intact scene	Congruent scrambled scene	Incongruent scrambled scene	Baseline
Threshold	3.44 ± 0.65	3.76 ± 0.70	3.59 ± 0.52	3.62 ± 0.68	3.57 ± 0.51
mRT	522 ± 57	550 ± 54	541 ± 62	533 ± 58	533 ± 52
mLog(RT)	2.70 ± 0.06	2.73 ± 0.04	2.72 ± 0.05	2.71 ± 0.04	2.71 ± 0.04

Table 2. Mean threshold values, mean correct response times in milliseconds (mRT), and mean log transformed correct response times (mLog[RT]), with standard deviations, for each experimental condition (Congruence, Background).

without an object) for 150 ms, then by a second stimulus for 150 ms, and finally by a backward mask (1/f noise) for 30 ms. The second stimulus depended on the experimental condition: intact background with a congruent object, intact background with an incongruent object, scrambled background with a congruent object, scrambled background with an incongruent object, baseline 1/f noise background with an object, intact background without an object, or scrambled background without an object. The first stimulus was thus the background of the second stimulus. The interstimulus interval was 3000 ms on average. It should be noted that in the second stimulus, there was always a 1/f noise circular patch in central vision (including an object or not). Thus, in order to maintain a coherent percept in central vision throughout the trial, we also blended a 1/f noise circular patch in the background of the first stimulus.

As in Experiment 1, the task was to categorize the object as animal or furniture by pressing two keys on a keyboard. Since the visibility of the object was low in many trials, participants were encouraged to rely on their "intuition" and were instructed to respond at random when they could not see any object (without favoring one of the two response keys). Response keys were counterbalanced across participants. Accuracy and RTs were recorded at each trial. Experimental conditions were fully randomized. There were 1,720 trials: 280 intact-congruent trials (20 animal object/intact outdoor and 20 furniture object/intact indoor × 7 levels of object phase coherence), 280 intact-incongruent trials (20 animal object/intact indoor and 20 furniture object/intact outdoor × 7 levels of object phase coherence), 280 scrambled-congruent trials (20 animal object/scrambled outdoor and 20 furniture object/scrambled indoor × 7 levels of object phase coherence), 280 scrambled-incongruent trials (20 animal object/scrambled indoor and 20 furniture object/scrambled outdoor × 7 levels of object phase coherence), 280 noise-baseline trials (20 animal object/noise and 20 furniture object/noise × 7 levels of object phase coherence), 160 intact scene alone (80 intact outdoor and 80 intact indoor), and 160 scrambled scene alone (80 scrambled outdoor and 80 scrambled indoor). The experiment was split into two 1-hr experimental sessions, including pauses. Prior to completing the experimental trials, participants underwent a training session (20 trials) to be familiarized with the stimuli and the task.

Results

Data analysis

For each participant, correct RTs were log transformed and then trimmed by removing trials for which the RT was inferior or superior to the condition average ± 2.5 SDs (0.34% of trials were excluded). Data used for the analysis of RT only included phase coherence levels for which mCR (averaged across conditions and participants) exceeded 0.75 at group level (Levels 4, 5, 6, and 7 of phase-coherence). Weibull psychometric functions were fitted to each participant's mCR through maximum likelihood estimation for the congruent and incongruent conditions of the intact and scrambled scene conditions, as well as for the baseline condition, using the *Quickpsy* package in R (Linares & Lopez-Moliner, 2016). For each experimental condition, threshold values were derived from the psycho-metric functions at a mCR of 0.75 ($1 - p$(chance) / 2). Threshold values, mRT and mlog(RT), with standard deviations, for each experimental condition (congruence and background) are reported in Table 2.

Congruence effect and nature of influences: Intact versus scrambled scenes

To assess if the congruence effect was due to low-level or semantic features in the scene background, we conducted a repeated measures ANOVA on the threshold values and mLog(RT) with congruence (congruent or incongruent stimuli) and background (intact or scrambled scenes) as within-subject factors. Further pairwise comparisons were tested with two-tailed paired-sample t tests using a Bonferroni adjustment of alpha level to correct for multiple tests (0.05 / 2 = 0.025 for the two tests performed with each dependent variable). Effect size was estimated by calculating Cohen's d for within subject design.

Analysis of thresholds values showed that main effects of congruence ($F[1, 17] = 4.26$, $p = 0.055$) and background ($F[1, 17] < 1$) were not significant but that these two factors interacted ($F[1, 17] = 7.54$, $p = 0.014$, $d = 0.65$; Figure 5). Follow-up pairwise comparisons
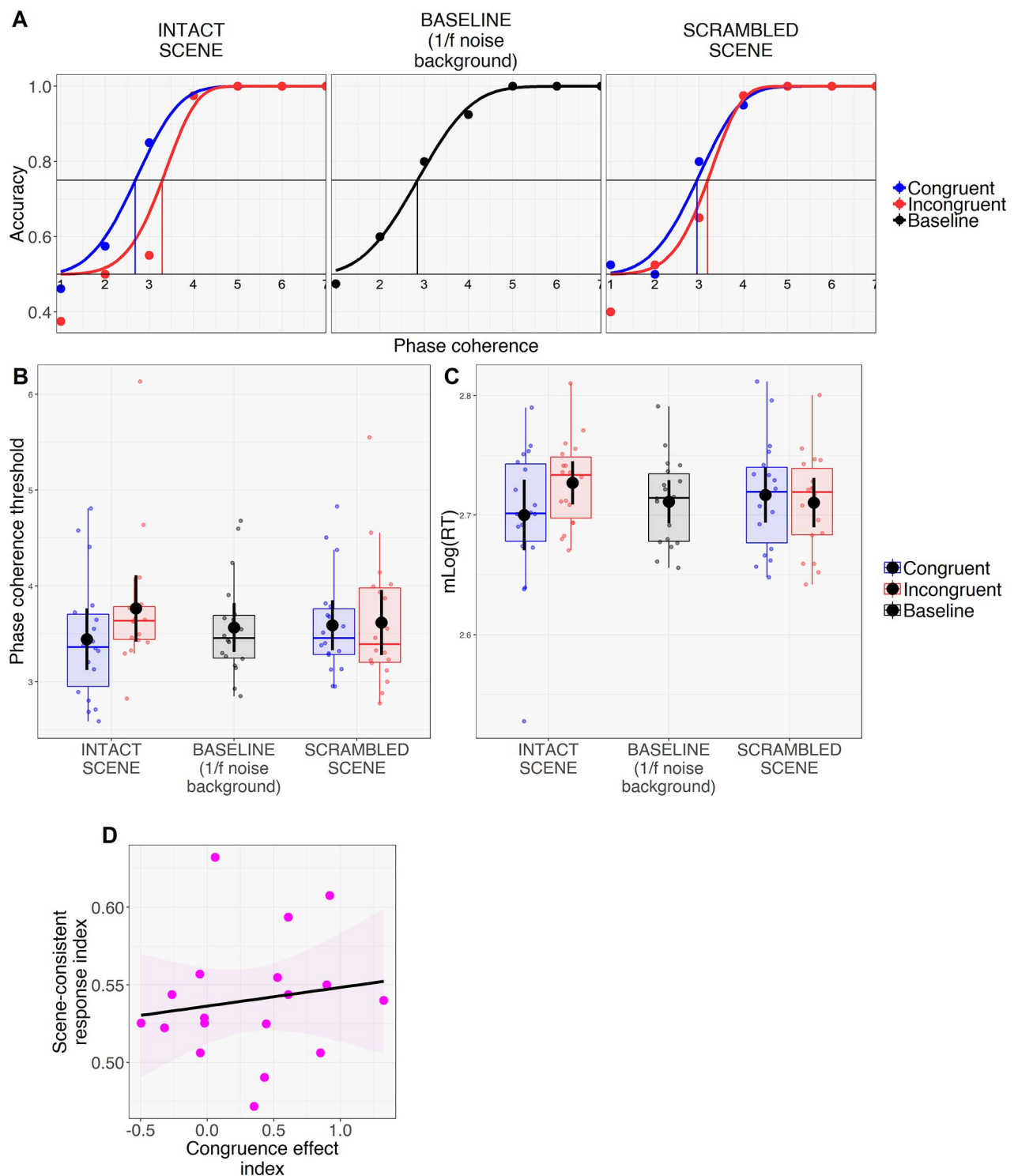
Figure 5. (A) For illustrative purposes, the figure shows a Weibull model fitted to the accuracy of one representative subject for the congruent and incongruent stimuli of the intact and scrambled background conditions, and for the baseline condition (for which the background was just 1/f noise). (B) Mean correct response rates (mCR) and (C) mean of log transformed correct RTs (mLog[RT]) plotted according to the experimental condition. Black dots and error bars indicate means and 95% *CI*s. Color dots are the individual observations. (D) Congruence effect index (difference in thresholds between the congruent and incongruent stimuli of the intact scene condition) as a function of the scene-consistent response index (the tendency of participants to use the category of the scene to infer the category of the object, when there is no object in the stimulus). Shading is 95% *CI*.

showed that thresholds were significantly lower in the congruent (3.44 ± 0.65) than in the incongruent (3.76 ± 0.70) condition for the intact scenes (t[17] = 2.73, p = 0.014, d = 0.65) while there was no difference between the congruent and incongruent conditions for the scrambled scenes (congruent: 3.59 ± 0.53; incongruent: 3.62 ± 0.68; t[17] = 0.36, p = 0.723).

Analysis of mLog(RT) gave very similar results with no main effect neither of congruence (F[1, 17] = 1.56, p = 0.229) nor of the background (F[1, 17] < 1), but there was an interaction between the two factors (F[1, 17] = 7.13, p = 0.016, d = 0.63; Figure 5). However, follow-up pairwise comparisons showed no main effect of congruence neither with intact scenes (congruent: 2.70 ± 0.06; incongruent: 2.73 ± 0.04; t[17] = 2.12, p = 0.049) nor with scrambled scenes (congruent: 2.72 ± 0.05; incongruent: 2.71 ± 0.04; t[17] = 0.90, p = 0.383).

These results partially support those of Experiment 1 by showing that visual information in the periphery is automatically processed and can decrease the level of perceptual quality needed to recognize objects in central vision when congruent, compared to when incongruent. However, unlike Experiment 1, we did not observe the congruence effect on RTs. Furthermore, it seems that peripheral visual information needs to carry a semantic meaning to influence the processing of the object, as no congruence effect was observed when the phase of the peripheral scene was altered (scrambled scene condition). Since there was no congruence effect with scrambled scenes, the following analyses only consider the intact scene condition.

Direction of the congruence effect: Comparison to the baseline

To evaluate the direction of the congruence effect (facilitation in the congruent condition, hindrance in the incongruent condition, or both), we compared threshold values and mLog(RT) of the baseline condition (meaningless 1/f noise background) to those of the congruent and incongruent conditions with two-tailed paired-sample t tests (alpha level Bonferroni corrected: 0.05 / 2 = 0.025 for the two tests performed with each dependent variable). For threshold values, these tests showed no difference neither between the baseline (3.57 ± 0.51) and the congruent condition (t[17] = 1.37, p = 0.190), nor between the baseline and the incongruent condition (t[17] = 1.99, p = 0.063). For mLog(RT), there was no difference between the baseline (2.71 ± 0.04) and the congruent condition (t[17] = 0.89, p = 0.388), but participants were significantly slower in the incongruent than in the baseline condition (t[17] = 2.72, p = 0.015, d = 0.64).

Relation between the congruence effect and the scene-consistent response effect

In trials where no object was presented, we labeled as "scene-consistent responses" each *animal* response when an outdoor scene was presented and each *furniture* response when an indoor scene was presented. In theory, participants are supposed to respond at random between animal and furniture in the no-object trials. The proportion of scene-consistent responses thus represents the tendency of participants to use the category of the scene to infer the category of the object, when there is no object in the stimulus. We first tested if the proportion of scene-consistent responses, hereafter named *scene-consistent response index*, was different from chance level (i.e., a proportion of 0.5) in the intact condition using a one-sample t test. The proportion of scene-consistent responses averaged across participants was 0.54 ± 0.04, and was statistically different from chance level (t[17] = 4.28, p < 0.001, d = 1.01), indicating that when they could not identify any object, participants' responses were in part driven by the scene category. Then, we tested the Kendall correlation between the scene-consistent response index and the congruence effect index calculated on threshold values (incongruent *minus* congruent) of the intact scene condition. There was no correlation between these two variables (r = 0.16, p = 0.36; Figure 5). To avoid the limitation of concluding on the null hypothesis, we tested a Bayesian correlation using the *Psycho* package in R (Makowski, 2018). This Bayesian correlation analysis indicates anecdotal evidence (BF = 1.76) in favor of an absence of a positive association between the scene-consistent response index and the congruence effect index (r = 0.11, median absolute deviation = 0.21, 90% CI [−0.23, 0.46]). The correlation can be considered as large, moderate, small, or very small with respective probabilities of 2.94%, 15.58%, 32.78%, and 18.25%. As a reminder, the effect of the scene congruence on object categorization observed in our experiments may originate from two processes. The first would be that the implicit categorization of the scene automatically drives the response (e.g., the participant responds *animal* after having recognized an outdoor scene, independently of the recognition of the object). The second would be an *integration* of the peripheral visual input to the recognition processes occurring in central vision (i.e., the process we actually aimed to measure). In light of this, we observed no strong evidence for a link between the congruence effect index and the scene-consistent response index, suggesting that the effect of congruence on object categorization observed in our experiments was mostly due to the *integration* of the peripheral visual input to the object recognition processes in central vision, rather than being directly driven by the categorization of the scene.

Discussion

Most of the time in natural vision, eye movements direct central vision toward relevant objects, while peripheral vision continuously extracts coherent contextual information. The objective of this study was twofold: to explore how LSF extracted in peripheral vision can shape the processing of central visual inputs, and to test the predictive nature of these influences. In two independent experiments, we examined how information conveyed by LSF in peripheral vision could influence the processes leading to object categorization in central vision. Participants categorized objects surrounded by congruent information (i.e., indoor scenes for furniture, outdoor scenes for animals) or incongruent information (the other way around). Importantly in our experimental design, contextual information was restricted to peripheral vision (beyond 6° in Experiment 1 and 8° in Experiment 2) and object information to central vision. We did not filter out the HSF in peripheral vision (peripheral vision acts as a natural low-pass filter), nor did we filter out the LSF of the object in central vision (we were not interested in LSF and HSF integration at the object level). Unfortunately, we were not able to record eye movements during the experiments in order to control that participants maintained their gaze on the object location when the scene was presented. However, given that the task was to attend and categorize the object in central vision, it would have been inefficient and even demanding to make eye movements to other locations without impairing object categorization. Indeed, saccadic eye movements are usually initiated within 100–150 ms (Fischer & Weber, 1993). If participants made a saccade in the scene (presented during 150 ms in the SOA-150 condition) prior to making a new saccade toward the object (also presented during 150 ms), this puts severe constraints on (a) the time left to process visual information in the scene between a first saccade on the scene and an immediate second saccade on the object or (b) the time left to actually initiate a new saccade toward the object and categorize it. Yet, mean correct response rate was high (in Experiment 1 and in Experiment 2 when object visibility was reasonable) suggesting that eye movements did not impair object categorization in the vast majority of trials. As expected, we found an effect of the congruence of the scene in both experiments indicating that contextual information in peripheral vision can be processed automatically during a central visual field directed task. We speculate that even when irrelevant for the ongoing task, peripheral vision is used to provide information on the type of object and details that are expected to be perceived in central vision, by generating a predictive signal that controls visual processes in central vision, in line with predictive coding theories (Friston & Stephan,

2007; Lee & Mumford, 2003; Rao & Ballard, 1999) and recent models of visual recognition (Bar, 2003; Kauffmann et al., 2014; Peyrin et al., 2010). Moreover, in Experiment 2, the visibility threshold to accurately categorize the object was lower when the peripheral context was congruent than incongruent. In real-life conditions, where information in peripheral vision is usually congruent with central vision, predictive processes based on peripheral vision could thus boost the perceptual processing of poorly visible objects in the scene. Such mechanisms surely can be a benefit to visual recognition in low lighting and masking conditions.

Still, it is easy to overinterpret this kind of results in terms of predictive coding. We can think of another account, already discussed by Brandman and Peelen (2017), of how contextual scene information may influence object processing in this type of experimental paradigms. Without the need for predictive feedback mechanisms, object and context could be processed mostly independently, in parallel, and only later their semantic representations would be integrated. There are empirical arguments for this view. For example, the fact that visual regions in the occipito-temporal cortex are preferentially involved in the visual processing of objects (the fusiform and inferior temporal gyri; Grill-Spector, 2003; Grill-Spector, Kourtzi, & Kanwisher, 2001) and scenes (e.g., the parahippocampal cortex; Epstein, 2005; Epstein & Kanwisher, 1998) can be interpreted as the existence of partly independent representations of object and scene information. The approach adopted by Brandman and Peelen (2017) in a fMRI-MEG study to disentangle the two accounts (predictive feedback vs. parallel processing) was to pixelize the object, rending its recognition practically impossible without a context. Behavioral results of a classification task showed that the recognition of such objects was much better when embedded in congruent scenes, than when isolated. The decoding accuracy of a classifier of BOLD activity in object and scene selective cortex was also strongly enhanced when the objects were in congruent contexts. These findings suggest that contextual information helped object categorization by shaping the *perceptual* representation of the object, via predictive feedbacks flowing through neurons of the visual cortex, and therefore are not compatible with the hypothesis of a strictly independent and parallel construction of semantic scene and object representations.

In our study, we used a different approach to test for the predictive hypothesis. In Experiment 1, we manipulated the robustness of the scene representation by presenting it for different durations before the object onset. In the object-background literature, we had no knowledge of any study having directly manipulated the sequential apparition of scene and object in order to

test for predictive aspects of object and scene integration. We speculated that the more it is possible to accumulate information in peripheral vision, the stronger the predictive signal would be and the more it would influence object recognition in central vision. We found that the effect of congruence increased linearly with scene duration, a result that is also in favor of a predictive, rather than parallel, view of background influence. In Experiment 2, mean visibility threshold for the baseline condition (noise background) was approximately in between mean threshold for the congruent condition and that for the incongruent condition (see Figure 5), even though the baseline condition did not significantly differ from either condition. For RTs, we did observe a difference between the incongruent condition and the baseline. Overall, the pattern of results is not clear-cut and, as it stands, rather suggests that an incongruent context delayed the object categorization. Therefore, we did not replicate the facilitation effect observed with congruent background by Brandman and Peelen (2017) with poorly visible object. Nonetheless, the control condition in their study was simply the object passed onto a uniform background. We rather embedded the object in a noise background following the relationship between amplitude and spatial frequency that is typical of natural scenes ($1/f^2$; Ruderman & Bialek, 1994). It is thus possible that the mere presence of noise in peripheral vision boosted the perception of the object, for example via mechanisms of stochastic resonance (McDonnell & Ward, 2011).

A substantial problem with sequential presentation is that we cannot be sure about the mechanisms responsible for a given participant's response. As mentioned above, a given object categorization can be (a) "overhasty," i.e., driven directly (and erroneously in incongruent trials) by the recognition of the scene category, or (b) actually driven by the integration of the scene to the object representation. In Experiment 2, we wanted to ensure that the congruence effect was due to the second account. We included trials in which no object was present, while participants were not aware of the manipulation. Looking at behavioral responses to those stimuli, we measured by how much participants had a tendency to use the category of the scene to infer the category of the object even though there was no object in the stimulus. First, we found that participants' responses were indeed driven by the scene category (0.54% ± 0.04% of responses were consistent with the scene category, significantly different from chance). In a predictive coding framework, related effects have been shown, where feedback activity carrying information about the surrounding context in the visual cortex was measured without any feedforward input (Smith & Muckli, 2010). However, we found that the tendency to respond based on the scene

category did not correlate with the size of the congruence effect, suggesting that the latter account (b) was more likely to drive our results compared to the former account (a).

The other aim of our study was to characterize further the nature of the peripheral influence. We manipulated the presence of a semantic content in the background scene by presenting it either in its intact version or in a scrambled version made by disrupting the phase of the image. We found no effect of congruence when the phase of the scene was scrambled, suggesting that the mere low-level attributes of scenes do not influence object categorization.[1] In fact, this would make sense if we think of situations where some artificial objects are typically encountered in natural environments (e.g., benches in public gardens) or natural objects in artificial environments (e.g., a cat in a living room). In such situations, the low-level visual information (i.e., the amplitude spectrum) available in peripheral vision would not be consistent with that of the object. Thus, our results suggest that predictions rather contain more mid- or high-level information and that they are initiated in a higher level cortex (for example, in the orbitofrontal cortex or in regions of the ventral pathway) that represents semantics aspects of the environment, as postulated by some models of visual recognition (Bar, 2003; Fabre-Thorpe, 2011; Kauffmann, Bourgin, et al., 2015). Consistently, studies by Loschky and collaborators (Loschky & Larson, 2008; Loschky et al., 2007) have shown that the Fourier amplitude spectrum is insufficient for scene gist recognition. This suggests that the scene must be recognized at some level (e.g., coarse gist recognition) in order to influence the object processing. The mere low-level signal would not be sufficient. Yet, it is difficult to conclude on the null hypothesis. It is thus possible that the physical information contained in the scrambled scenes alone does influence object processing at a lower extend, but that we were not able to measure this effect here. For example, this can be due simply to a lack of statistical power, since we planned our effect size for Experiment 2 based on the congruent effect of Experiment 1, where only intact scenes where presented. It is also possible that phase scrambling is not a good model of low-level predictive processes in the brain, and that the spatial distribution of low-level features—available in the phase spectrum—is in fact important for such processes. Stojanoski and Cusack (2014) have compared the neural activity at the earliest stages of the visual system induced by intact images, phase-scrambled images, as well as diffeomorphic images (meaningless images resembling images printed on a distorted rubber sheet). They have shown that these regions do not respond to phase-scrambled stimuli in the same way as they do to intact ones. On the contrary, neural activity induced by diffeomorphic

images is indistinguishable to that induced by intact images. Diffeomorphic stimuli may therefore be a better way to model low-level visual processes.

All that being said, it seems important to highlight some matters regarding the effect of congruence in our experiments. Initially, we hypothesized that low-resolution information in peripheral vision could be used to help perceptual processing of the object in central vision via predictive processes. However, the effect of congruence was not observed when the scene was simultaneously presented with the object (SOA-0 condition), nor when the scene was presented 30 ms before the onset of the object (SOA-30 condition). Yet, according to spatial frequency–based models of visual recognition, LSF are processed faster than HSF and thus could influence the processing of the latter when perceived simultaneously. It seems that the visual system must have enough time to process peripheral LSF for it to influence object recognition (here, in the SOA-150 condition). Second, in the SOA-150 condition, the difference in RT between congruent and incongruent conditions was only of 20 ms, as pointed out before. Third, in Experiment 2, the congruence effect on RT observed in Experiment 1 did not replicate (participants were not significantly faster to categorize the object in the congruent than incongruent condition), although we observed the effect on thresholds.

Therefore, it is possible that peripheral vision only plays a moderate role in the predictive processes leading to object recognition, and that LSF available in central vision are preferentially used to initiate such predictive processes. Nonetheless, it should be noted that our task was a superordinate (animal vs. furniture) categorization task. A basic level categorization task could benefit more from low-resolution contextual information in peripheral vision. Future studies should compare directly LSF based predictions from central vision to those from peripheral vision, and also investigate other levels of object categorization (e.g., basic level). On the other hand, it is possible that LSF are not a good descriptor of peripheral vision. For example, crowding and ensemble perception are important properties of peripheral vision that we have not considered here. Influential models of peripheral vision such as Balas, Nakano, and Rosenholtz (2009) suggest that such mechanisms lead to consider peripheral vision as qualitatively different from central vision. Peripheral vision would represent summary statistics of different features of the environment (orientations, size, textures, hue) within pooling regions. Therefore, summary statistics could be a better descriptor of peripheral vision than LSF.

To conclude, this study shows that low-resolution information in peripheral vision can be unintentionally processed and integrated to information in central vision, if available for a sufficient amount of time. This effect is not likely to be due to response strategy, but rather on mechanisms of periphery-based predictions integrated during the processing of local elements in central vision. The study also suggests that predictions rely on the semantic processing of peripheral information, since mere low-level information did not influence the processing of the object in central vision. Such mechanisms could boost perception of poorly visible objects. Given that the effects are rather small and unstable, further studies manipulating the level of object categorization and comparing predictions coming from central and peripheral vision are needed to explore the issue further.

Footnote

[1] It should be noted that the absence of a congruence effect in the scrambled condition does not preclude that the mere presence of noise in this condition may have boosted the perception of the object, as we previously suggested for the baseline condition (1/f noise background).

References

Ales, J. M., Farzin, F., Rossion, B., & Norcia, A. M. (2012). An objective method for measuring face detection thresholds using the sweep steady-state visual evoked response. *Journal of Vision, 12*(10):

18, 1–18, https://doi.org/10.1167/12.10.18. [PubMed] [Article]

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience*, *29*(34), 10638–10652.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12):13, 1–18, https://doi.org/10.1167/9.12.13. [PubMed] [Article]

Baldassano, C., Fei-Fei, L., & Beck, D. M. (2016). Pinpointing the peripheral bias in neural scene-processing networks during natural viewing. *Journal of Vision*, *16*(2):9, 1–14, https://doi.org/10.1167/16.2.9. [PubMed] [Article]

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609.

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron*, *38*(2), 347–358.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, *103*(2), 449–454.

Bar, M., & Ullman, S. (1993). Spatial context in recognition. *Perception*, *25*(3), 343–352.

Boucart, M., Moroni, C., Szaffarczyk, S., & Tran, T. H. C. (2013). Implicit processing of scene context in macular degeneration. *Investigative Ophthalmology and Visual Science*.

Brandman, T., & Peelen, M. V. (2017). Interaction between scene and object processing revealed by human fMRI and MEG decoding. *Journal of Neuroscience*, *37*(32), 7700–7710.

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.

Davenport, J. L. (2007). Consistency effects between objects in scenes. *Memory & Cognition*, *35*(3), 393–401.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 545–559.

Epstein, R. (2005). The cortical basis of visual scene processing. *Visual Cognition*, *12*(6), 954–978.

Epstein, R., & Kanwisher, N. (1998, April 9). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*, 243.

Fischer, B., & Weber, H. (1993). Express saccades and visual attention. *Behavioral and Brain Sciences*, *16*(3), 553–567.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, *159*(3), 417–458.

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2010). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476.

Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*(2), 159–166.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422.

Hegdé, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology*, *84*(4), 405–439.

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, 1–18, https://doi.org/10.1167/8.13.11. [PubMed] [Article]

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.

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.

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.

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 37.

Kok, P., & de Lange, F. P. (2015). Predictive coding in sensory cortex. In *An introduction to model-based cognitive neuroscience* (pp. 221–244). New York, NY: Springer.

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.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America, A*, *20*(7), 1434–1448.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature Neuroscience*, *4*(5), 533–539.

Linares, D., & Lopez-Moliner, J. (2016). quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, *8*(1), 122–131.

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, 1–9, https://doi.org/10.1167/8.1.4. [PubMed] [Article]

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), 1431–1450.

Makowski, D. (2018). The Psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 470.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in Cognitive Sciences*, *6*(4), 176–184.

McDonnell, M. D., & Ward, L. M. (2011). The benefits of noise in neural systems: Bridging theory and experiment. *Nature Reviews Neuroscience*, *12*(7), 415.

Petras, K., ten Oever, S., Jacobs, C., & Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *NeuroImage*, *186*, 103–112.

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., ... 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.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Rémy, F., Saint-Aubert, L., Bacon-Macé, N.,

Vayssière, N., Barbeau, E., & Fabre-Thorpe, M. (2013). Object recognition in congruent and incongruent natural scenes: A life-span study. *Vision Research*, *91*, 36–44.

Ruderman, D. L., & Bialek, W. (1994). Statistics of natural images: Scaling in the woods. In *Advances in Neural Information Processing Systems* (pp. 551–558).

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–200.

Shams, L., & Von Der Malsburg, C. (2002). The role of complex cells in object recognition. *Vision Research*, *42*(22), 2547–2554.

Shapley, R., & Lennie, P. (1985). Spatial frequency analysis in the visual system. *Annual Review of Neuroscience*, *8*(1), 547–581.

Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proceedings of the National Academy of Sciences*, *107*(46), 20099–20103.

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision*, *14*(12):6, 1–16, https://doi.org/10.1167/14.12.6. [PubMed] [Article]

Sun, H.-M., Simon-Dack, S. L., Gordon, R. D., & Teder, W. A. (2011). Contextual influences on rapid object categorization in natural scenes. *Brain Research*, *1398*, 40–54.

Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J. O., ... Kreiman, G. (2018). Recurrent computations for visual pattern completion. *Proceedings of the National Academy of Sciences*, *115*(35), 8835–8840.

Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, *8*(1), 10853.

Van Essen, D., & DeYoe, E. A. (1995). Concurrent processing in the primate visual cortex. In *The Cognitive Neurosciences* (pp. 383–400).

Woodhead, Z. V. J., Wise, R. J. S., Sereno, M., & Leech, R. (2011). Dissociation of sensitivity to spatial frequency in word and face preferential areas of the fusiform gyrus. *Cerebral Cortex*, *21*(10), 2307–2312.

Wyatte, D., Curran, T., & O'Reilly, R. (2012). The limits of feedforward vision: Recurrent processing promotes robust object recognition when objects are degraded. *Journal of Cognitive Neuroscience*, *24*(11), 2248–2261.