Journal of Experimental Psychology: Human Perception and Performance

The Spatiotemporal Dynamics of Scene Gist Recognition

Adam M. Larson, Tyler E. Freeman, Ryan V. Ringer, and Lester C. Loschky

Online First Publication, November 18, 2013. doi: 10.1037/a0034986

Journal of Experimental Psychology:
Human Perception and Performance

0096-1523/13/$12.00 DOI: 10.1037/a0034986

The Spatiotemporal Dynamics of Scene Gist Recognition

Adam M. Larson
University of Findlay

Tyler E. Freeman, Ryan V. Ringer, and
Lester C. Loschky
Kansas State University

Viewers can rapidly extract a holistic semantic representation of a real-world scene within a single eye fixation, an ability called recognizing the *gist* of a scene, and operationally defined here as recognizing an image's basic-level scene category. However, it is unknown how scene gist recognition unfolds over both time and space—within a fixation and across the visual field. Thus, in 3 experiments, the current study investigated the spatiotemporal dynamics of basic-level scene categorization from central vision to peripheral vision over the time course of the critical first fixation on a novel scene. The method used a window/scotoma paradigm in which images were briefly presented and processing times were varied using visual masking. The results of Experiments 1 and 2 showed that during the first 100 ms of processing, there was an advantage for processing the scene category from central vision, with the relative contributions of peripheral vision increasing thereafter. Experiment 3 tested whether this pattern could be explained by spatiotemporal changes in selective attention. The results showed that manipulating the probability of information being presented centrally or peripherally selectively maintained or eliminated the early central vision advantage. Across the 3 experiments, the results are consistent with a zoom-out hypothesis, in which, during the first fixation on a scene, gist extraction extends from central vision to peripheral vision as covert attention expands outward.

Keywords: scene gist, scene categorization, spatiotemporal processing, attention, time course

When viewing a scene image, people may describe it with such terms as "a bedroom" or "a person raking leaves." Viewers can arrive at this semantic categorization of a real-world scene extremely rapidly—within a single eye fixation—and the theoretical construct for this process is often called "scene gist recognition" (Biederman, Rabinowitz, Glass, & Stacy, 1974; Castelhano & Henderson, 2008; Fei-Fei, Iyer, Koch, & Perona, 2007; Greene & Oliva, 2009; Larson & Loschky, 2009; Loschky & Larson, 2010; Malcolm, Nuthmann, & Schyns, 2011; Oliva & Torralba, 2006; Potter, 1976; Rousselet, Joubert, & Fabre-Thorpe, 2005). The scene gist construct is important for theories of scene perception because recognizing the gist of a scene affects later theoretically important processes such as attentional selection (Eckstein, Drescher, & Shimozaki, 2006; Gordon, 2004; Torralba, Oliva, Castelhano, & Henderson, 2006), object recognition (Bar & Ullman, 1996; Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce & Pollatsek, 1992; Davenport & Potter, 2004; but see Hollingworth & Henderson, 1998), and long-term memory for scenes (Brewer & Treyens, 1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989).

Scene gist recognition has been operationalized in numerous ways, but usually in terms of the ability to classify a briefly flashed scene image at some level of abstraction, from the highly specific (e.g., "a baby reaching for a butterfly"; Fei-Fei et al., 2007; Intraub, 1981; Potter, 1976, pp. 509–510), to the basic-level scene category (e.g., front yard; (Loschky, Hansen, Sethi, & Pydimari, 2010; Malcolm et al., 2011; Oliva & Schyns, 2000; Renninger & Malik, 2004; Rousselet et al., 2005), to the superordinate-level scene category (e.g., natural; Goffaux et al., 2005; Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Loschky & Larson, 2010), to whether the scene contains an animal (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Evans & Treisman, 2005; Fei-Fei, VanRullen, Koch, & Perona, 2005; Kirchner & Thorpe, 2006; Rousselet, Fabre-Thorpe, & Thorpe, 2002), to the scene's emotional valence (e.g., positive; Calvo, 2005; Calvo, Nummenmaa, & Hyona, 2008). Thus, the theoretical construct of *scene gist* has been operationalized in many different ways regarding the semantic information that viewers acquire from a scene. Similar to many previous studies, the current study operationally defines scene gist recognition as viewers' ability to accurately categorize real-world scenes at the basic level (e.g., Joubert et al., 2007; Loschky & Larson, 2010; Loschky et al., 2007, 2010; Malcolm et al., 2011; McCotter, Gosselin, Sowden, & Schyns, 2005; Oliva & Schyns, 2000; Rousselet et al., 2005).

As noted, a fundamental constant in the scene gist construct is that the specified semantic information is acquired within a single fixation. The fact that scene gist acquisition occurs within a single eye fixation has been shown by eye movement research in which the very first eye movement in a visual search task, which imme-

Adam M. Larson, Department of Psychology, University of Findlay; Tyler E. Freeman, Ryan V. Ringer, and Lester C. Loschky, Department of Psychological Sciences, Kansas State University.

Author contributions: All authors conceived and designed the experiments; the experiments were performed by Adam M. Larson and Ryan V. Ringer; the data were analyzed by Adam M. Larson, Tyler E. Freeman, and Ryan V. Ringer; and Lester C. Loschky, Adam M. Larson, Tyler E. Freeman, and Ryan V. Ringer contributed to writing the article.

Correspondence concerning this article should be addressed to Adam M. Larson, Department of Psychology, 200 Howard St., University of Findlay, Findlay, OH 45840. E-mail: larson@findlay.edu

diately follows the first fixation (typically placed at the center of the image prior to onset of the image), generally goes directly to an expected location based on the semantic category of the scene. For example, it has been shown when the search target was "chimney," the first saccade on a scene often went directly to the roof of a house, even when there was no chimney in the picture (Eckstein et al., 2006; see also Torralba et al., 2006). Other studies have strongly suggested that gist recognition occurs within a single fixation by presenting extremely briefly flashed and backward-masked scene images (to control processing time) and asking viewers to categorize the scenes at the basic level. These studies have shown asymptotic basic-level scene categorization performance at stimulus onset asynchronies (SOAs) of 100 ms (Biederman et al., 1974; Potter, 1976) and an inflection point after SOAs as little as 35 to 50 ms (Bacon-Macé et al., 2005; Loschky & Larson, 2010; Loschky, Larson, Smerchek, & Finan, 2008; Loschky et al., 2007, 2010), which is far less than the average duration of a fixation during scene viewing of 330 ms (Rayner, 1998). Thus, in order to build on the findings from the latter studies, the current study investigates the time course of scene gist processing by varying processing time over a wide range of masking SOAs whose maximum is roughly equal to a single fixation.

Although much is known about how humans process and represent the gist of scenes, a key unknown is how scene gist recognition unfolds over both time and space—within a single eye fixation and across the visual field. Specifically, because viewers acquire scene gist within the temporal window of a single fixation, any given piece of visual information in a scene has a fixed retinal eccentricity during that critical first fixation. Thus, because retinal eccentricity greatly affects visual processing (for review, see Strasburger, Rentschler, & Juttner, 2011; Wilson, Levi, Maffei, Rovamo, & DeValois, 1990), the retinal eccentricity of scene information must play a key role in scene gist acquisition (Larson & Loschky, 2009). However, this raises the novel question addressed in this study: Does the spatial variability in processing across the visual field undergo important changes over the time course of the critical first fixation on a scene? Interestingly, three plausible alternative hypotheses regarding such spatiotemporal variability in scene gist processing are suggested by the existing literature.

The first hypothesis stems from the well-known differences in the speed of information transmission between central and peripheral vision from the retina to the brain, with peripheral visual information reaching the lateral geniculate nucleus (LGN) of the thalamus and the primary visual cortex (V1) before central visual information (Nowak, Munk, Girard, & Bullier, 1995; Schmolesky et al., 1998). Because peripheral vision is particularly important for scene gist recognition (Larson & Loschky, 2009), it is possible and plausible that this temporal processing advantage for peripheral vision may underlie the incredible speed of gist recognition (Calvo et al., 2008; Girard & Koenig-Robert, 2011).

A second plausible alternative hypothesis is based on eye movement and attention research, which has shown that covert attention starts centrally in foveal vision at the start of each fixation, and, over time, extends out to the visual periphery (Henderson, 1992; White, Rolfs, & Carrasco, in press). This central-to-peripheral spatiotemporal order of visual processing could extend to the process of scene gist recognition during the critical first fixation on a scene, producing an advantage for central vision early in the first fixation.

Finally, a third plausible alternative hypothesis is based on the idea that the rapid extraction of scene gist within a single fixation occurs in the near absence of attention (Fei-Fei et al., 2005; Li, VanRullen, Koch, & Perona, 2002; Otsuka & Kawaguchi, 2007; Rousselet et al., 2002) and in parallel across the field of view (Rousselet, Fabre-Thorpe, & Thorpe, 2002). In that case, neither central nor peripheral vision would be expected to play a larger role early or late in processing scene gist, but instead would be assumed to play equivalent roles throughout the first fixation.

These three plausible alternative hypotheses cover three logical possibilities, either (a) an early advantage for peripheral over central vision, (b) the reverse, namely, an early advantage for central over peripheral vision, or (c) no advantage for either central or peripheral vision. Below, we describe the research supporting each of these alternative hypotheses in greater detail.

Central Versus Peripheral Vision and the Spatiotemporal Dynamics of Scene Gist Recognition

The visual field can be roughly divided into two mutually exclusive regions: central and peripheral vision. Central vision, which includes both foveal and parafoveal vision,[1] is contained within a roughly 5° radius of fixation (Osterberg, 1935; cited in Strasburger et al., 2011, p. 3). We follow standard convention in studies of visual cognition by defining peripheral vision as the remainder of the visual field beyond central vision's 5° radius (e.g., Hollingworth, Schrock, & Henderson, 2001; Holmes, Cohen, Haith, & Morrison, 1977; Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981; Shimozaki, Chen, Abbey, & Eckstein, 2007; van Diepen & Wampers, 1998).

Our first of three alternative hypotheses regarding the spatiotemporal dynamics of scene gist acquisition is related to the fact that the vast majority of information in real-world scenes is contained within peripheral vision, which has lower spatial resolution but a finer temporal resolution and faster information transmission to visual cortex than central vision (Livingstone & Hubel, 1988; Nowak et al., 1995; Strasburger et al., 2011; Wilson et al., 1990). Neurophysiological studies of macaques have shown that visual information transmitted by the magnocellular retinal ganglion cells reaches the LGN and V1 approximately 20 ms faster than information transmitted by the parvocellular retinal ganglion cells (Nowak et al., 1995; Schmolesky et al., 1998). This advantage has been estimated to be substantially larger (90 ms) for peripheral vision in humans, as shown for discrimination of Gabor orientation in central vision (4° eccentricity) versus peripheral vision (10° eccentricity) (Carrasco, McElree, Denisova, & Giordano, 2003). The visual transmission advantage for peripheral vision could be critical for processing real-world scene images, including the recognition of a scene's basic-level category, especially at the early stages of scene processing.

[1] Although these two visual areas are referred to as central vision, there are important anatomical and perceptual differences between them. The fovea contains the greatest concentration of cones (Curcio, Sloan, Packer, Hendrickson, & Kalina, 1987), whereas the parafovea has the largest concentration of rods (Curcio, Sloan, Kalina, & Hendrickson, 1990). Likewise, visual acuity is greater in the fovea than the parafovea (Westheimer, 1982; Wilson et al., 1990).

Recent studies have shown the importance of peripheral vision for processing real-world scene images. Larson and Loschky (2009) showed that the central 5° of an image could be completely removed from a scene with no decrease in basic-level scene categorization performance. Conversely, presenting only the central 5° of a scene, while blocking scene information beyond that, produced worse categorization performance than when the entire scene image was presented. Similarly, Boucart and colleagues (Boucart, Moroni, Thibaut, Szaffarczyk, & Greene, 2013) showed the usefulness of peripheral vision for scene categorization by presenting scene images to the left and right of fixation and having viewers indicate the side with the target category. Performance was good (73% accuracy), even for scenes presented at up to 70° eccentricity. Similar results have been shown for animal detection in scenes using far peripheral vision (Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001). These results show that peripheral vision conveys critical information for basic-level scene categorization despite its low spatial resolution. Given that information from peripheral vision is transmitted to the LGN and V1 faster than information presented in central vision, this could produce better scene gist recognition in peripheral vision than central vision at the earliest stages of scene processing (Calvo et al., 2008; Girard & Koenig-Robert, 2011).

Interestingly, a separate body of literature on attention in scenes suggests a second alternative hypothesis regarding the spatiotemporal processing of scene gist in the first fixation. Henderson (1992) has argued for the sequential attention model, in which attention starts in central vision for each eye fixation and is later sent to the target of the next saccade in the visual periphery toward the end of the fixation. Research on reading processes has shown evidence consistent with this hypothesis (Rayner et al., 1981; Rayner, Liversedge, & White, 2006; Rayner, Liversedge, White, & Vergilino-Perez, 2003), and later research found similar findings for visual search in scenes and scene memory (Glaholt, Rayner, & Reingold, 2012; Rayner, Smith, Malcolm, & Henderson, 2009; van Diepen & d'Ydewalle, 2003). For example, van Diepen and d'Ydewalle (2003) found that in a nonobject search task, masking foveal information early in a fixation was more detrimental than masking peripheral information. This deleterious effect was observed with both the search task and eye movement measures, suggesting that at the beginning of each fixation, information from the center of vision was processed first, followed by the information contained in the visual periphery. However, van Diepen and d'Ydewalle (2003) argued that their findings might not apply to other tasks:

> Given the task demands in the present experiments, objects of moderate size were of primary importance, and had to be inspected foveally. Conceivably, in other tasks a much larger part of the stimulus is processed at the beginning of fixations (e.g., when the ***scene identity*** has to be determined). Obviously, in the latter tasks the area that will affect fixation durations can be expected to be much larger than just the foveal stimulus. (p. 97; emphasis added)

That is, although there is evidence of the sequential allocation of attention from central to peripheral vision during extended visual search in scenes, it may not apply to the rapid acquisition of scene gist at the beginning of the first fixation on a scene, when peripheral vision may be more important.

A more recent study by Glaholt et al. (2012) further tested the sequential attention model with real-world scenes in both visual search and scene recognition memory tasks. In that study, either central vision or the entire image was masked, gaze contingently, after varying SOAs within the first 100 ms of each fixation (Glaholt et al., 2012). That study showed that loss of central vision (with a 3.71°-radius mask) only affected early processing. For scene recognition memory, a central mask with a 0-ms masking SOA reduced performance, but a 50-ms SOA did not. For visual search, a central mask with a 50-ms masking SOA degraded foveal target identification, but a 100-ms SOA did not. In contrast, masking the entire image disrupted both visual search and scene recognition memory as late as a 75-ms masking SOA. Thus, whereas central vision was only important early in a fixation, peripheral vision remained important until later in fixations for both visual search and scene recognition memory. These results are consistent with the sequential attention model, but do not speak to van Diepen and d'Ydewalle's (2003) argument that the foveal-to-peripheral sequential attention model may not apply to scene gist recognition on the first fixation on a scene. Thus, this remains an important untested hypothesis regarding the spatiotemporal dynamics of scene gist acquisition.

An interesting question is whether the sequential attention model can be interpreted as a spatiotemporally constrained version of the zoom-lens model (Eriksen & St. James, 1986; Eriksen & Yeh, 1985), in which covert attention zooms out on each fixation. Regardless, if the sequential attention model indeed applies to scene gist recognition on the first fixation on a scene (contrary to van Diepen & d'Ydewalle's, 2003, suggestion), then it suggests that basic-level scene categorization should be best in central vision at the early stages of a fixation, whereas performance should converge at later processing times.

Finally, a third alternative hypothesis regarding the spatiotemporal dynamics of scene gist acquisition is suggested by research on the role of attention in scene gist acquisition. This research questions whether attentional processes—here, the spatiotemporal dynamics of visual attention across the visual field—underlie scene gist recognition, or whether preattentive processes, which work across the entire field of view in parallel, underlie scene gist recognition. Several recent studies have suggested that scene categorization requires little, if any, attentional resources (Fei-Fei et al., 2005; Li et al., 2002; Otsuka & Kawaguchi, 2007; Rousselet et al., 2002), whereas other studies suggest that attention may yet play a role in obtaining meaningful scene information (Cohen, Alvarez, & Nakayama, 2011; Evans & Treisman, 2005; Walker, Stafford, & Davis, 2008).

If scene gist recognition requires attention and is affected by attentional processes, then the spatiotemporal dynamics of attention during the initial fixation on a scene should produce differences in basic-level scene categorization between central and peripheral vision at early processing times. Conversely, if scene gist acquisition is an attention-free process, based primarily on preattentive processes operating in parallel across the entire field of view (Fei-Fei et al., 2005; Li et al., 2002; Otsuka & Kawaguchi, 2007; Rousselet et al., 2002), then no differences should be found between the utility of central versus peripheral vision for scene gist acquisition over the time course of the critical first fixation. This constitutes a well-founded and plausible null hypothesis regarding

the role of the spatiotemporal dynamics of visual attention in scene gist acquisition.

In sum, the goals of the current study were to determine (a) whether there is any difference in the utility of central versus peripheral vision in acquiring the basic-level scene category of a scene over the time course of a single fixation, and (b) if such differences exist, whether they are more consistent with the idea that peripheral vision is processed most quickly, and thus dominates early scene categorization, or whether processing expands from central vision outward over the course of a single fixation consistent with the sequential attention model.

General Method of the Study

The current study used a "window" and "scotoma" paradigm (see Figure 1) to evaluate the relative contributions of central versus peripheral vision to scene gist recognition over time (Larson & Loschky, 2009). We define a *window* as a circular viewable region encompassing the central portion of a scene, while blocking the more eccentric peripheral information (McConkie & Rayner, 1975; van Diepen, Wampers, & d'Ydewalle, 1998). Conversely, a *scotoma* blocks out the central portion of a scene and shows only the peripheral information (Rayner & Bertera, 1979; van Diepen, et al., 1998). An inherent difficulty in such a method is that any difference in scene categorization could potentially be explained simply in terms of a difference in the amount of viewable information available in each condition. For example, if it were shown that peripheral vision had an advantage over central vision, one could argue that this was because peripheral vision had more information. Conversely, if central vision showed an advantage, this advantage could similarly be argued to be due to cortical magnification of foveal and parafoveal information. Thus, to control for such potentially confounding spatial attributes inherent to central and peripheral vision, it is first necessary to determine what we call the "critical radius"—that is, the radius that perfectly divides the central and peripheral regions of a scene into two mutually exclusive regions, each of which produces *equivalent scene categorization performance* when given *unlimited processing time within a single fixation* (i.e., when images are *unmasked*, and thus sensory memory generally lasts until a saccade is made; Larson & Loschky, 2009). Given a critical radius for which unlimited processing time in a single fixation produces equal performance for information presented in both window and scotoma conditions, then we can ask whether *limiting* processing time produces any difference in scene categorization performance between those window and scotoma conditions based on the critical radius.

Figure 1. Example image in a window and a scotoma viewing condition. Note that the radius for the window is the same as the radius for the scotoma. This is the critical radius used in the current study, which produces equal scene gist recognition in both the window and scotoma conditions when presented without a visual mask (i.e., when given unlimited processing time within a single fixation).

It is important to note from the outset that such a research strategy is highly conservative, with a strong bias toward finding no difference between the window and scotoma conditions, given that the critical radius is defined in terms of producing equivalent performance between the two conditions (when there is no masking). Therefore, if variations in processing time *do* produce differences even when using the critical radius, then we can be confident that those differences do *not* stem from an imbalance in the amount of image content provided within the window and scotoma conditions, respectively. It is for this reason that we took the conservative strategy of using the critical radius to balance the functional value of viewable imagery in the window and scotoma conditions.

General Hypotheses

If the greater neural transmission speed of peripheral vision influences early differences in scene gist acquisition, then we would expect a scene categorization performance advantage for the scotoma condition over the window condition at early processing times. Conversely, if the sequential attention model applies to acquiring scene gist over the course of a single fixation, then we would expect a scene categorization performance advantage for the window condition at early processing times compared with the scotoma condition. Finally, if scene gist acquisition is a largely parallel and preattentive process, then window and scotoma conditions should be equally useful for scene categorization throughout the critical first fixation, and varying the processing times for window and scotoma conditions should produce no advantage for either condition, whether early or late in processing.

We conducted three experiments to explore the spatiotemporal dynamics of scene gist acquisition in a single fixation and their relationship to visual attention. Across the three experiments, our data suggest that at the beginning of a fixation, attention is allocated to central vision. However, within the first 100 ms of scene processing, attention expands to encompass peripheral areas of the visual field. These findings are consistent with a combination of the sequential attention model and a spatiotemporally constrained interpretation of the zoom lens model of attention that we call the *zoom-out hypothesis*. These novel results, and the theoretical advance they provide, place fundamental spatiotemporal constraints on any theory of the processes involved in scene gist acquisition.

Experiment 1

Experiment 1 was designed to investigate the relative utility of information in central versus peripheral vision for scene gist acquisition over the time course of single fixation. We used a window/scotoma paradigm to selectively present scene information to either central or peripheral vision, respectively, together with visual masking to vary processing time. This enabled us to test our three competing hypotheses that the transmission speed

advantage for peripheral vision, the sequential attention model, or parallel and preattentive processes across the visual field would best explain the spatiotemporal dynamics of scene gist recognition.

Our window and scotoma stimuli were constructed using a *critical radius* that produced *equivalent basic-level scene categorization* in *both the window and scotoma conditions* when the stimuli were *unmasked*, in order to functionally equalize the viewable information presented in the window and scotoma conditions. By using the critical radius, however, we greatly reduced the chances of rejecting the null hypothesis when comparing the window and scotoma conditions. Thus, even relatively small differences found between the two conditions as a function of SOA would indicate differences in the spatiotemporal dynamics of scene gist recognition.

Method

Participants. There were 56 participants (33 female), whose ages ranged from 18 to 32 years old ($M = 19.59$, $SD = 2.02$). All had normal or corrected-to-normal vision (20/30 or better), gave their institutional review board-approved informed written consent, and received course credit for participating.

Design. The experiment used a 2 (window vs. scotoma) $\times$ 6 (processing time) within-subjects design. There were 28 practice trials, followed by 240 recorded trials.

Stimuli. Window and scotoma stimuli were created from circularly cropped scene images having a diameter of 21.9° (i.e., a maximal radius/retinal eccentricity of 10.95°) at a viewing distance of 63.5 cm, using a forehead and chin rest. We interpolated the size of the critical radius based on the prior results of Larson and Loschky (2009), and confirmed through pilot testing that a critical radius of 5.54° (170 pixel radius) produced equal performance in both the window and scotoma conditions when stimuli were presented for 24 ms unmasked. Window images presented 25.6% of the viewable scene area inside the critical radius, while 74.4% of the viewable scene area was presented outside the critical radius in the scotoma condition. We used a total of 240 images, which were comprised of 10 scene categories (five natural: beach, desert, forest, mountain, river; five man-made: farm, home, market, pool, street). Thus, the 240 scene images were randomly assigned to each viewing condition, and each scene was presented only once. The circular scene stimuli were presented on a 17-in. ViewSonic Graphics Series CRT monitor (Model G90fb).

Masks were scene texture images generated using the Portilla and Simoncelli (2000) algorithm. These types of masks have been shown to be highly effective at disrupting scene gist processing because they contain second-order and higher order image statistics similar to real-world scenes but do not contain any recognizable information (Loschky et al., 2010). Masks were identical in shape and size to the stimuli they masked—window stimuli were masked by window masks, and scotoma stimuli were masked by scotoma masks (see Figure 2). This was done to avoid metacontrast masking, which tends to produce Type B (u-shaped) masking functions (Breitmeyer & Ogmen, 2006; Francis & Cho, 2008). All images, including targets and masks, were equalized in terms of their mean luminance and Root Mean Square (RMS) contrast (see Loschky et al., 2007, for details on equalizing mean luminance and RMS contrast). Scene information contained outside the window, or inside the scotoma, was replaced by neutral gray equal to the mean luminance value of our stimuli. The same gray value was used for the blank screens and the background of the fixation point and category label.

Procedures. After completing a preliminary visual near acuity test (using a Snellen chart), participants were seated in front of a computer monitor. Participants were first familiarized with the 10 scene categories by showing them four sample images from each category together with their respective category labels. Participants then completed 30 unrecorded practice trials before completing the 240 experimental trials. The sample and practice stimuli were not used again in the main experiment.

Figure 2 shows a schematic of a trial in each of the two viewing conditions (window vs. scotoma). We used an EyeLink 1000 remote eyetracking system with a forehead and chin rest to maintain a constant viewing distance. The eyetracker was programmed with a "fixation failsafe" algorithm to ensure that participants were fixated in the center of the screen. If the participant was not fixated within a 1° $\times$ 1° bounding box at the center of the image when they pressed the gamepad button to initiate a trial, the trial was recycled and did not initiate. Thus, the participant was always fixating the center of the screen when they initiated a trial. After a 48-ms delay, the target image was flashed for 24 ms, and following the prescribed interstimulus interval (ISI) of 0, 71, 165, 259, or 353 ms (which produced the target-to-mask SOAs of 24, 94, 188, 282, or 376 ms,[2] or a no-mask condition), the mask was presented for 24 ms. It was predicted that the longest masking SOA, which is only slightly longer than the average fixation durations on scene images (330 ms), would be equivalent to the no-mask condition. This is because the retinal image in the no-mask condition would be masked by a saccade (Irwin, 1992) after, on average, 330 ms. The shortest SOA was based on the shortest stimulus duration that produced above-chance performance with this task and stimuli based on pilot testing. The other SOAs were chosen to provide roughly equal steps between the two extremes. Following the mask presentation, there was a 750-ms blank screen, and then a category label was presented until the participant responded using a handheld gamepad. The category label was a valid description of the scene on 50% of trials (and invalid on the other 50%). If the label was valid, participants were instructed to press the "yes" button on their game pad, and otherwise to press the "no" button. The presentation of window and scotoma scene image conditions were randomized throughout the experiment. All scene image categories appeared equally often, in random order. All category labels appeared equally often, and invalid labels were randomly selected from the remaining nine categories without replacement.

Results

Precursors. Due to poor task performance, we eliminated data from participants whose average accuracy was at or below the fifth percentile (<51.98% correct; two participants). The fixation data for each participant was then filtered spatially and temporally to ensure that the point of fixation was within the 1° $\times$ 1°

[2] The monitor's 85-Hz refresh rate allowed images to be presented for multiples of 11.76 ms. SOAs were calculated based on this refresh rate, with the reported SOAs rounded to the nearest whole number.

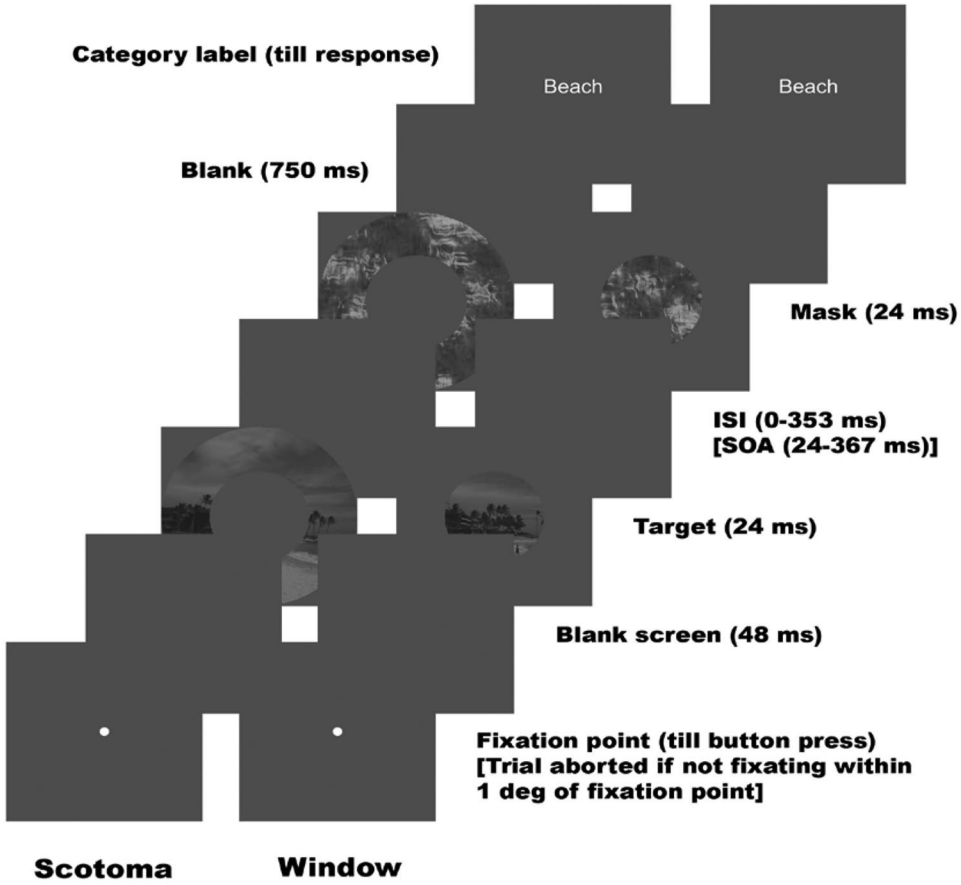

Figure 2. Trial schematic. Both window and scotoma conditions are illustrated for comparison.

bounding box at the center of the image for the entire time period from the onset of the target to the offset of the mask, and that there was only a single eye fixation during this time period. Any trials that did not meet these criteria were discarded. Overall, 17.1% of the trials were removed from the analysis, resulting in a total of 11,337 trials that satisfied the experimental constraints. A greater proportion of trials were eliminated from the 376-ms SOA (22% of these trials were removed; 1,654 trials satisfied the experimental constraints) and the 282-ms SOA (16% of the trials were removed; 1,799 trials satisfied the experimental constraints) compared with the remaining SOAs (<11% of the trials were removed; 1,919 to 1,993 trials satisfied the experimental constraints per SOA condition). This was due to the fact that viewers were more likely to spontaneously make an eye movement within the 282-ms and 376-ms SOA conditions than at the shorter SOAs (<188-ms SOA).

Main analyses. As assumed by use of the critical radius, basic-level scene categorization performance did not differ between the window and scotoma conditions in the no-mask condition, $t(53) = 0.57$, $p = .57$, Cohen's $d = .07$. This equivalence was shown by calculating the reciprocal of the JZS Bayes factor (= 0.125) from the t value, which showed substantial evidence in favor of the null (Wetzels et al., 2011).[3] Thus, crucially important for our method, the critical radius produced equal performance in the window and scotoma conditions when processing time lasted for a single eye fixation (i.e., when there no mask, and thus the subject's next eye movement masked their retinal image of the scene).

The remaining data were analyzed with a 2 (viewing condition: window vs. scotoma) × 5 (SOA: 24, 94, 188, 282, and 376) within-subjects factorial ANOVA, and a trend analysis was performed to determine if there were any differences in the psychophysical functions between the two viewing conditions over time. Trend analyses are only reported for interval scaled independent variables with more than two levels. The results of window and scotoma performance across SOAs are shown in Figure 3. As expected, scene categorization performance increased with processing time, as shown by a large and significant main effect of

[3] The JZS Bayes Factor can be used to determine the degree of *support* for the null hypothesis (Rouder, Speckman, Sun, Morey, & Iverson, 2009). Wetzels et al. (2011, Table 1) gives interpretations of 1/(JZS Bayes Factor) as calculated by Rouder et al. (2009). The null hypothesis is stated below as H0, and the alternative hypothesis is stated as HA. Those values and interpretations are as follows: A 1/Bayes factor >100 = decisive evidence for HA; 30–100 = very strong evidence for HA; 10–30 = strong evidence for HA; 3–10 = substantial evidence for HA; 1–3 = anecdotal evidence for HA; 1 = no evidence; 1/3–1 = anecdotal evidence for H0; 1/10–1/3 = substantial evidence for H0; 1/30–1/10 = strong evidence for H0; 1/100–1/30 = very strong evidence for H0; <1/100 = decisive evidence for H0.

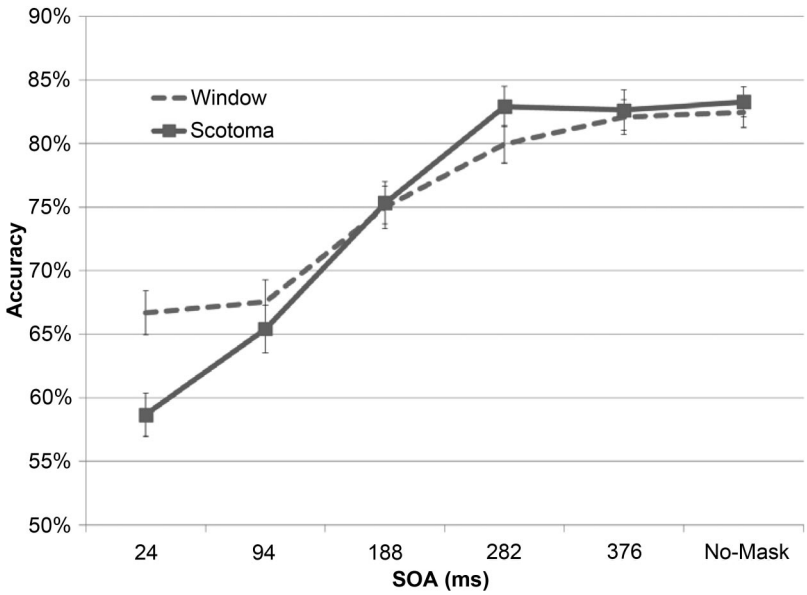

Figure 3. Scene gist accuracy as a function of image type (window vs. scotoma) and processing time (SOA in ms). Error bars represent the standard error for each data point. *No-mask* is the unmasked condition.

processing time, $F(4, 212) = 77.40$, $p < .001$, Cohen's $f = .752$,[4] as well as significant linear trend, $F(1, 53) = 257.16$, $p < .001$. There was no main effect detected in the window versus scotoma comparison, $F(1, 53) = 1.53$, $p = .22$, Cohen's $f = .031$.

Our chief interest was whether or not there would be a significant interaction between the window/scotoma viewing conditions and processing time. As shown in Figure 3, there was a significant interaction, $F(4, 212) = 4.05$, $p = .003$, Cohen's $f = .150$, such that there was an advantage for the window conditions over the scotoma conditions, but only at the shortest SOA.[5] This was verified with multiple Bonferroni-corrected t tests (critical p value = .01), which compared scene categorization performance between the window and scotoma conditions at each SOA (24 ms: $t[53] = 3.19$, $p = .002$, Cohen's $d = 0.45$; all other ts < 1.82, ps $> .07$). This is consistent with the hypothesis that processing of scene gist begins in central vision and expands outward over time. As predicted based on our use of the critical radius stimuli, the longest SOA (376 ms) produced identical performance between window ($M = 82.1\%$, $SD = .10$) and scotoma conditions ($M = 82.6\%$, $SD = .12$), $t(53) = 0.34$, $p = .74$, Cohen's $d = 0.04$, JZS Bayes Factor = 8.86. The longest SOA ($M = 82.4\%$, $SD = .09$) also produced identical results to the no-mask condition ($M = 82.8\%$, $SD = .07$), $t(53) = 0.33$, $p = .74$, Cohen's $d = 0.03$, JZS Bayes Factor = 8.89, which was necessarily predicted by our assumption that a 376-ms masking SOA is equivalent to the complete processing time for a single fixation afforded by the no-mask condition.

Discussion

The results of Experiment 1 showed that the window condition produced moderately and significantly better scene categorization performance than the scotoma condition at the earliest processing time (24-ms SOA). This advantage was gone by the 94-ms SOA, and thereafter processing was equivalent between both conditions. These results are inconsistent with our hypothesis based on the peripheral vision transmission speed advantage, which predicted that the scotoma condition would be better than the window condition at the early stages of processing. Likewise, these results are inconsistent with the hypothesis that the basic-level scene category can be recognized in the near absence of attention in parallel across the field of view (Fei-Fei et al., 2005; Li et al., 2002; Otsuka & Kawaguchi, 2007; Rousselet et al., 2002). However, contrary to the suggestion of van Diepen and d'Ydewalle (2003), these results are consistent with the sequential attention model (Henderson, 1992, 1993) as applied to scene gist recognition. This assumes that attention starts at the point of fixation at the beginning of a fixation, and later extends to the next saccade target in the visual periphery. The data are also consistent with a combination of (a) the sequential attention model, and (b) a spatiotemporally constrained version of the zoom lens model of attention (Eriksen & St. James, 1986; Eriksen & Yeh, 1985), in which attention zooms out from the center of vision over the course of a fixation, which we call the zoom-out hypothesis. Thus, at the beginning of the stimulus presentation (24-ms SOA), participants were more accurate when information was presented at the center of vision than if it was presented in the visual periphery (beyond 5.54° eccentricity in the scotoma condition), suggesting that attention was focused near the center of vision, in which there was information in the window condition but not the scotoma condi-

[4] Cohen's f magnitudes for small, medium, and large effect sizes are generally given as .10, .25, and .40, respectively (Cohen, 1988).

[5] The trend analyses confirmed the results of the ANOVA by showing that the interaction was significant as a linear trend, $F(1, 53) = 7.09$, $p = .010$. As this suggests, and as shown in Figure 3, as processing time increased, performance increased at different rates for the window and scotoma scene images.

tion. However, after an additional 70-ms processing time (94-ms SOA), information presented either centrally or peripherally produced equal performance, suggesting that, by that time, attention had expanded to encompass the entire image in both the window and scotoma conditions. Thereafter, performance monotonically increased at an equivalent rate in both conditions as processing time increased further.

It is worth noting that asymptotic scene categorization performance was not reached until roughly the 282-ms SOA, which is considerably longer than in most studies assessing basic-level scene categorization using visual masking to control processing time. A simple explanation for this is that both the window and scotoma conditions were missing scene information that would otherwise be present in a whole scene image, and thus extra processing time was required to reach asymptotic performance. The fact that a whole image condition requires less processing time points to the importance of processing across the entire visual field (or at least the entire image), and thus the advantage for central information early in a fixation should be considered in relative rather than absolute terms.

Experiment 2

Experiment 1 assessed the spatiotemporal dynamics of scene gist acquisition, operationalized in terms of basic-level scene categorization, between 24 and 376 ms processing time. However, the data from Experiment 1 showed that the important scene-processing differences between central and peripheral vision were in the first 100 ms of processing. The advantage for information presented centrally at 24-ms SOA disappeared by 94 ms, and thereafter performance was similar across all SOAs up to and including 376 ms. Thus, a key question is what happens to the spatiotemporal dynamics of scene gist acquisition during the first 100 ms of viewing a scene? Specifically, when does information presented to central and peripheral vision become equally useful? Experiment 2 addressed these questions.

Method

Participants. There were 85 participants (48 females), whose ages ranged from 18 to 32 years old ($M = 19.51$, $SD = 2.05$). All had normal or corrected-to-normal vision (20/30 or better), gave their institutional review board-approved informed written consent, and received course credit for participating.

Design. The design of Experiment 2 was the same as Experiment 1 except that the SOAs in Experiment 2 were more densely sampled from the first 100 ms of processing, when meaningful differences in Experiment 1 were observed. As in Experiment 1, target and mask images were presented for 24 ms. However, the ISIs were 0, 12, 24, 47, 71, and 353 ms (producing SOAs of 24, 35, 47, 71, 94, and 376 ms). Because no difference was found between the 376-ms SOA and the no-mask condition in Experiment 1, the 376-ms SOA served the same function as a no-mask condition in Experiment 2 (i.e., providing a single eye fixation's processing time). Thus, the SOAs used in Experiment 2 reflect a range of processing time from the minimum processing time necessary for above chance performance (24-ms SOA) to the point in processing in Experiment 1 when information presented to both central and peripheral vision first produced equal performance (94-ms SOA). The other SOAs were chosen to provide roughly equal steps between the two extremes. As in Experiment 1, participants completed familiarization and practice trials before the recorded trials.

Stimuli and procedures. All stimuli and procedures were the same as in Experiment 1, except for the different SOAs.

Results

Precursors. Data cleaning procedures were the same as in Experiment 1. Due to poor task performance, we eliminated data from participants whose average accuracy was at or below the fifth percentile (<55.96% correct; four participants). The fixation data for each subject was then filtered spatially and temporally to ensure that the point of fixation was within the $1° \times 1°$ bounding box at the center of the image for the entire time period from the onset of the target to the offset of the mask, and that there was only a single eye fixation during this time period. Any trials that did not meet these criteria were discarded. This resulted in a total of 17,444 trials that satisfied the experimental conditions after removing 15.5% of the trials from the analysis. Similar to Experiment 1, a greater proportion of data was eliminated from the 376-ms SOA (19% of the trials were removed; 2,541 trials satisfied the experimental constraints) than the remaining SOAs (<12% of the trials were removed per SOA condition; 2,932 to 3,011 trials satisfied the experimental constraints per SOA condition).

Main analyses. As assumed by the use of the critical radius, and seen in Table 1, scene categorization performance in the window and scotoma conditions was not different at the longest SOA (376 ms), $t(80) = 0.81$, $p = .42$, Cohen's $d = .08$, with the JZS Bayes Factor (= 0.121) indicating substantial evidence in favor of the null (Wetzels et al., 2011). Thus, when given the equivalent of a single eye fixation to process the scenes, equal performance was found in the two viewing conditions.

A 2 (viewing condition: window vs. scotoma) × 5 (SOA: 24, 35, 47, 71, and 94 ms) within-subjects factorial ANOVA was used to analyze performance between the two viewing conditions over the first 100 ms of scene processing. As shown in Figure 4, as expected, basic-level scene categorization increased with processing time, $F(4, 320) = 7.02$, $p < .001$,

Table 1
Descriptive Statistics for Window and Scotoma Conditions in Experiment 2 at Each Level of Processing Time

Processing time (SOA in ms)	Window		Scotoma		Effect	
	M	*SD*	*M*	*SD*	*M*	*SE*
24	0.668	0.109	0.634	0.119	0.032	0.016
35	0.643	0.108	0.589	0.112	0.053	0.015
47	0.636	0.123	0.607	0.120	0.029	0.019
71	0.658	0.113	0.623	0.136	0.035	0.018
94	0.691	0.122	0.654	0.119	0.036	0.019
376	0.793	0.133	0.807	0.103	−0.013	0.016

Note. Performance at each processing time (SOA) < 100 ms for both window and scotoma conditions was significantly different from performance at the longest SOA (376 ms; the control condition), all ts > 5.77, all ps < .001 (Bonferroni corrected $\alpha = .005$). The "Effect" column is the mean difference between the Window condition and Scotoma condition for a given level of SOA. SOA = stimulus onset asynchronies.

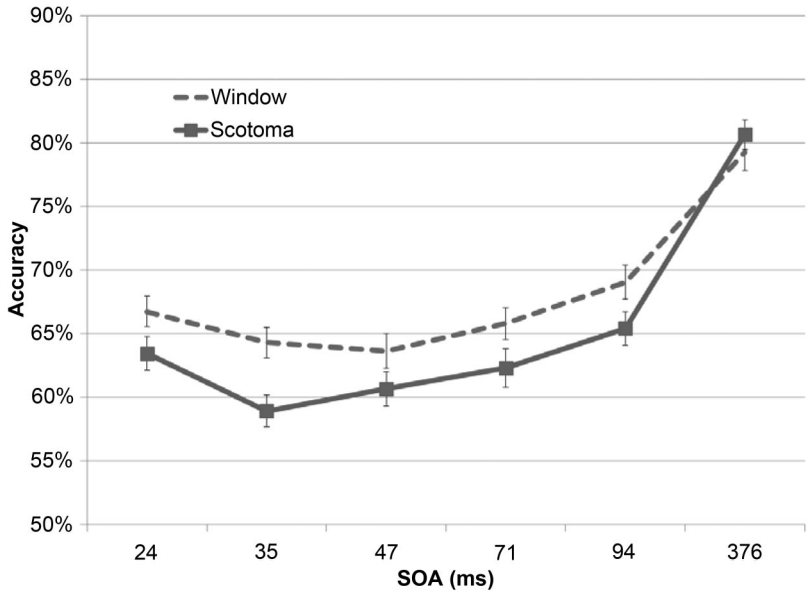

Figure 4. Scene gist accuracy as a function of image type (window vs. scotoma) and processing time (SOA in ms). Error bars represent the standard error for each data point.

Cohen's $f = .172$. Of greater interest, however, the window condition produced better scene categorization performance than the scotoma condition, $F(1, 80) = 23.69$, $p < .001$, Cohen's $f = 167$. However, the interaction between viewing condition and processing time did not affect scene categorization accuracy, $F(4, 320) = 0.27$, $p = .90$.[6] The lack of an interaction indicates that the advantage for the window image condition over the scotoma image was present over the entire first 100 ms of processing.

This result is consistent with both reading and scene perception research showing an advantage for information presented to central vision at the beginning of a fixation (Glaholt et al., 2012; Rayner et al., 1981, 2003, 2006; van Diepen & d'Ydewalle, 2003). Additionally, this pattern is consistent with the conclusion drawn from Experiment 1 that the scene information presented in central vision is used earlier than the information presented in the visual periphery.

Discussion

The results of Experiment 2 suggest that during the first 100 ms of scene viewing, information presented in central vision produced better scene gist recognition than that presented in peripheral vision. The results of Experiment 1 showed that performance for scene information presented in either central or peripheral vision was equal by the 94-ms SOA, however Experiment 2 showed that the central vision advantage was in fact present over the entire first 100 ms. This finding is consistent with previous research showing that in the early processing stages of an eye fixation, information processed by central vision is important for reading (Rayner et al., 1981, 2003, 2006), visual search in scenes, and scene memory (Glaholt et al., 2012; van Diepen & d'Ydewalle, 2003), whereas information in peripheral vision becomes increasingly important later in a fixation. Such results are therefore consistent with the sequential model of attention and the zoom-out hypothesis as applied to scene gist recognition, in which focal attention starts centrally and expands over time to encompass information in peripheral vision. Both explain the overall better performance for central vision during the first 100 ms of scene processing and the converging performance between information presented to central and peripheral vision over time. Thus, these results are also consistent with the idea that attention does indeed affect scene categorization and scene gist over the critically important first 100 ms of processing. Nevertheless, we draw this conclusion cautiously because we did not directly manipulate attention in either Experiment 1 or 2. Instead, we manipulated processing time and the availability of task-relevant information as a function of retinal eccentricity. Experiment 3 addressed this issue by directly manipulating viewers' attention while they carried out the scene categorization task in either the window or scotoma conditions.

Experiment 3

Experiments 1 and 2 suggest an advantage for basic-level scene categorization in central vision (i.e., the window condition) at early processing times (<100-ms SOA); however, this advantage disappeared for later processing times, resulting in equivalent scene categorization performance for information presented in central and peripheral vision. These results are consistent with the idea that while attempting to recognize the gist of rapidly presented scenes, attention starts centrally and expands to the visual periphery over time. This was predicted

[6] Trend analyses show that processing time produced a significant linear trend for accuracy, $F(1, 80) = 6.11$, $p = .016$. Additionally, the interaction between processing time and window/scotoma conditions on accuracy showed no evidence of a linear trend, $F(1, 80) = .04$, $p = .84$.

by our proposed zoom-out hypothesis, which derives from a combination of the sequential and zoom lens models of attention. However, Experiments 1 and 2 did not explicitly manipulate attention. Thus, Experiment 3 did just that, explicitly manipulating attention in order to test whether attentional allocation underlies the changing spatiotemporal patterns of scene categorization performance found in Experiments 1 and 2. Specifically, if the early advantage for the window condition is due to attention starting in central vision at the beginning of the first fixation, then that advantage should be eliminated if it is possible to strategically allocate attention to the visual periphery in the scotoma condition.

Attention was manipulated to be either focused in central vision or spread out to peripheral vision through a probability manipulation of the type often used in studies of covert attention (Geng & Behrmann, 2005; Yantis & Egeth, 1999; Yantis & Jonides, 1984). Such studies have shown that attention tends to be allocated to the area most likely to contain target information. Thus, we manipulated the proportion of trials in which either the window or the scotoma conditions were presented. In Experiments 1 and 2, the window and scotoma trials occurred randomly with equal probability. However, in Experiment 3, participants were assigned either to (a) the window-dominant condition, in which 80% of the trials presented window stimuli, and scotoma images composed the remaining 20%; or (b) the scotoma-dominant condition, in which these proportions are reversed.

Using this probability-based attentional manipulation, we can determine whether spatial attention modulates the spatiotemporal patterns of scene categorization performance observed in Experiments 1 and 2. If the early central vision advantage is in fact the result of an attentional process, then the early central advantage should remain in the window-dominant condition, but should be completely eliminated in the scotoma-dominant condition, because attention would be strategically spread across both central and peripheral vision.

Importantly, this logic should also take into account the fact that, given sufficient processing time, performance should be equal in both window and scotoma conditions, due to our use of the critical radius. Thus, the critical tests between the proposed alternative explanations must be at the shortest SOAs (e.g., 35-ms SOA), for which any spatiotemporal effects of attention on scene gist recognition would be expected to be greatest.

Method

Participants. There were 112 participants (80 females) with ages ranging from 17 to 36 ($M = 19.5$, $SD = 2.79$). All had normal or corrected-to-normal vision (20/30 acuity or better), gave their institutional review board-approved informed written consent (with written parental consent also given for those under the age of 18), and received course credit for participating.

Design. The experiment used a mixed design: 2 (window vs. scotoma) × 4 (processing time: 35-, 70-, 105-, 376-ms SOA), both within-subjects, × 2 (attentional bias: 80% window vs. 80% scotoma) between subjects, with participants randomly assigned to the attentional bias conditions.

Stimuli. An additional eight images per basic-level scene category were added to the image set used in Experiments 1 and 2.

Procedure. Procedures were the same as in Experiments 1 and 2, with the following exceptions. In the 80% window condition, participants completed 320 experimental trials (equivalent to the total number of images), of which a randomly selected 80% of trials (i.e., 256) were presented in the window condition, with the remaining 20% of trials (i.e., 64) presented in the scotoma condition. In the 80% scotoma condition, the ratio of window to scotoma trials was reversed. Participants were informed of the probabilities of viewing a window or scotoma scene for their respective condition. Additionally, in order to prepare participants for their attention condition, the 32 practice trials had the same ratio of window to scotoma trials as their condition (i.e., 26 images in the dominant condition and the remaining six images from the other condition).

Results

Precursors. Data cleaning procedures were the same as in Experiments 1 and 2. Due to poor task performance, we eliminated data from participants whose average accuracy was at or below the fifth percentile (<60.4% correct; six participants). The fixation data for each subject was then filtered spatially and temporally to ensure that the point of fixation was within the 1° × 1° bounding box at the center of the image for the entire time period from the onset of the target to the offset of the mask, and that there was only a single eye fixation during this time period. Any trials that did not meet these criteria were discarded. Data cleaning resulted in a total of 29,276 trials that were included in the analysis, after deleting 19.0% of the trials. As in Experiments 1 and 2, a greater proportion of trials were eliminated from the 376-ms SOA (23% of the trials were removed; 6,525 trials satisfied the experimental constraints) compared with the remaining SOAs (<11% of the trials were removed per SOA condition; 7,538 to 7,612 trials satisfied the experimental constraints per SOA condition).

Main analyses. As assumed from our use of the critical radius, scene categorization performance did not differ between the window and scotoma viewing conditions at 376-ms SOA, in either the 80% window, $t(48) = 1.44$, $p = .16$, Cohen's $d = .17$ JZS Bayes Factor = 3.30, or 80% scotoma conditions, $t(56) = 0.33$, $p = .74$, Cohen's $d = .04$, JZS Bayes Factor = 9.12, as shown in Table 2, with both JZS

Table 2
Descriptive Statistics for Window and Scotoma Scenes at Each Level of Processing for the Window- and Scotoma-Dominant Attentional Conditions in Experiment 3

Processing time (SOA in ms)	Window		Scotoma		Effect	
	M	*SD*	*M*	*SD*	*M*	*SE*
Window-dominant condition						
35	0.692	0.073	0.593	0.124	0.099	0.020
71	0.674	0.074	0.646	0.111	0.028	0.018
106	0.715	0.078	0.666	0.153	0.049	0.023
376	0.796	0.066	0.768	0.146	0.027	0.018
Scotoma-dominant condition						
35	0.644	0.131	0.631	0.059	0.012	0.018
71	0.686	0.142	0.659	0.070	0.026	0.018
106	0.698	0.132	0.719	0.083	−0.021	0.017
376	0.825	0.114	0.830	0.068	−0.005	0.017

Note. The "Effect" column is the mean difference between the window condition and scotoma condition for a given level of SOA. SOA = stimulus onset asynchronies.

Bayes Factors providing good evidence for the null. As before, this equivalence between window and scotoma conditions, when given the equivalent of one eye fixation's processing time, means that any difference observed between these two conditions at earlier processing times cannot be the result of differences in viewable information available to each. Rather, these differences must be due to spatiotemporal differences in processing.

The remaining data were submitted to a 2 (attentional manipulation: window- vs. scotoma-dominant conditions) × 2 (viewing condition: window vs. scotoma image) × 3 (SOA: 35, 71, 106) mixed factorial ANOVA. As shown in Figure 5, scene categorization increased with processing time (SOA), $F(2, 208) = 18.39$, $p < .001$, Cohen's $f = .165$.[7] Of greater interest, however, and consistent with Experiments 1 and 2, the window conditions produced better categorization performance than the scotoma conditions, as evidenced by a significant main effect of viewing condition, $F(1, 104) = 17.83$, $p < .001$, Cohen's $f = .115$. There was no main effect of the attentional manipulation on scene categorization accuracy, $F(1, 104) = 0.62$, $p = .433$. Likewise, there was no Viewing Condition × Processing Time interaction, $F(2, 208) = 2.30$, $p = .10$, Cohen's $f = .045$.[8]

The Attentional Manipulation × Processing Time interaction was not significant, $F(2, 208) = 0.78$, $p = .46$, nor were there any significant trends, $ps \geq .24$. However, of critical importance was the significant Attentional Manipulation × Viewing Condition interaction, $F(1, 104) = 11.71$, $p = .001$, Cohen's $f = .009$, namely, that the difference between the window and scotoma viewing conditions was affected by the probability-based manipulation of spatial attention. In the window-dominant condition, scene categorization performance was better for the window images ($M = .69$, $SD = 0.05$) than the scotoma images ($M = .64$, $SD = 0.08$), $t(48) = 4.93$, $p < .001$, Cohen's $d = 0.53$, but this advantage was eliminated in the scotoma-dominant condition (window: $M = .68$, $SD = 0.08$; scotoma: $M = .67$, $SD = 0.05$), $t(56) = 0.55$, $p = .59$, Cohen's $d = 0.11$, JZS Bayes Factor = 0.121, providing substantial support for the null. Thus, the early advantage for scene gist processing in central vision was alternatively eliminated or maintained by the allocation of spatial attention.

The three-way interaction between attentional manipulation, viewing condition, and processing time (SOA) failed to reach significance, $F(2, 208) = 2.56$, $p = .08$, Cohen's $f = .049$, suggesting that the two-way interaction between attentional manipulation and viewing condition did not differ over the first 100 ms of processing. However, a trend analysis showed a significant quadratic trend for the three-way interaction, $F(1, 104) = 5.16$, $p = .025$, which may be explained by visual inspection of Figure 5. Figure 5 shows that the window advantage in the window-dominant condition is greatest at the shortest SOA, whereas in the scotoma-dominant condition, there is no difference between window and scotoma conditions at any SOA. This observation was tested with a pair of 2 (viewing condition: window vs. scotoma image) × 3 (processing time: 35-, 71-, 106-ms SOA) within-subjects factorial ANOVAs for each respective attentional manipulation condition. The window-dominant condition had significant main effects for both processing time, $F(2, 96) = 5.54$, $p = .005$, Cohen's $f = .18$, and viewing condition, $F(1, 48) = 23.88$, $p < .001$, Cohen's $f = .28$, and a marginally significant interaction between processing time and the viewing condition, $F(2, 96) = 3.06$, $p = .052$, Cohen's $f = .118$, due to the greatest difference being at the 35-ms SOA. Conversely, the scotoma-dominant condition only showed a main effect of processing time, $F(2, 112) = 14.36$, $p < .001$, Cohen's $f = .28$, but no main effect for viewing condition, $F(1, 56) = .39$, $p = .53$, nor a significant Processing Time × Viewing Condition interaction, $F(2, 112) = 1.67$, $p = .19$, Cohen's $f = .063$. These results are consistent with the hypothesis that the observed spatiotemporal differences in scene categorization in Experiments 1 and 2 were due to differential allocation of attention over time and space. Specifically, the current results showed that the early central vision advantage is modulated by attention, such that it was completely eliminated if covert attention was strategically reallocated over the entire visual field, as shown in the scotoma-dominant condition.

Discussion

Experiment 3 examined whether the spatiotemporal pattern of scene gist acquisition, as seen in Experiments 1 and 2, could be reasonably explained in terms of an attentional mechanism. If so, then manipulating the spatial distribution of attention could potentially eliminate the central advantage by strategically reallocating some of the attentional resources from central vision to peripheral vision. Thus, attention would be allocated over the entire visual field, resulting in equivalent performance between these regions. To test this explanation, Experiment 3 used a probability-based attentional manipulation in which either centrally or peripherally presented scene information was presented 80% of the time. The results from Experiment 3 showed that the central advantage was indeed eliminated when scene information was more likely to be presented in the visual periphery in the scotoma-dominant condition. This suggests that attention was reallocated based on the likely spatial location of scene information. This, in turn, is consistent with the hypothesis that the spatiotemporal pattern of scene categorization performance observed in Experiments 1 and 2 can be explained in terms of differential allocation of covert attention—namely, the early scene categorization advantage for central vision, was either maintained or eliminated by differential attentional allocation.

The results also show that observers can draw on attention instructions (emphasizing central vs. peripheral processing) to impact performance very early in visual processing. Early literature demonstrated volitional flexibility in directing attention to specified pictures during rapid serial visual presentations (Intraub, 1984). The current study shows a similar volitional flexibility in attention to different spatial areas during presentation of a single image.

General Discussion

The current study examined the spatiotemporal dynamics of scene gist recognition, namely, the rapid acquisition of a semantic representation of a scene within a single eye fixation.

[7] A significant linear trend was found for processing time on scene gist accuracy, $F(1, 104) = 35.74$, $p < .001$.

[8] The trend analysis did show a significant linear interaction between viewing condition and processing time, $F(1, 104) = 4.25$, $p = .042$.

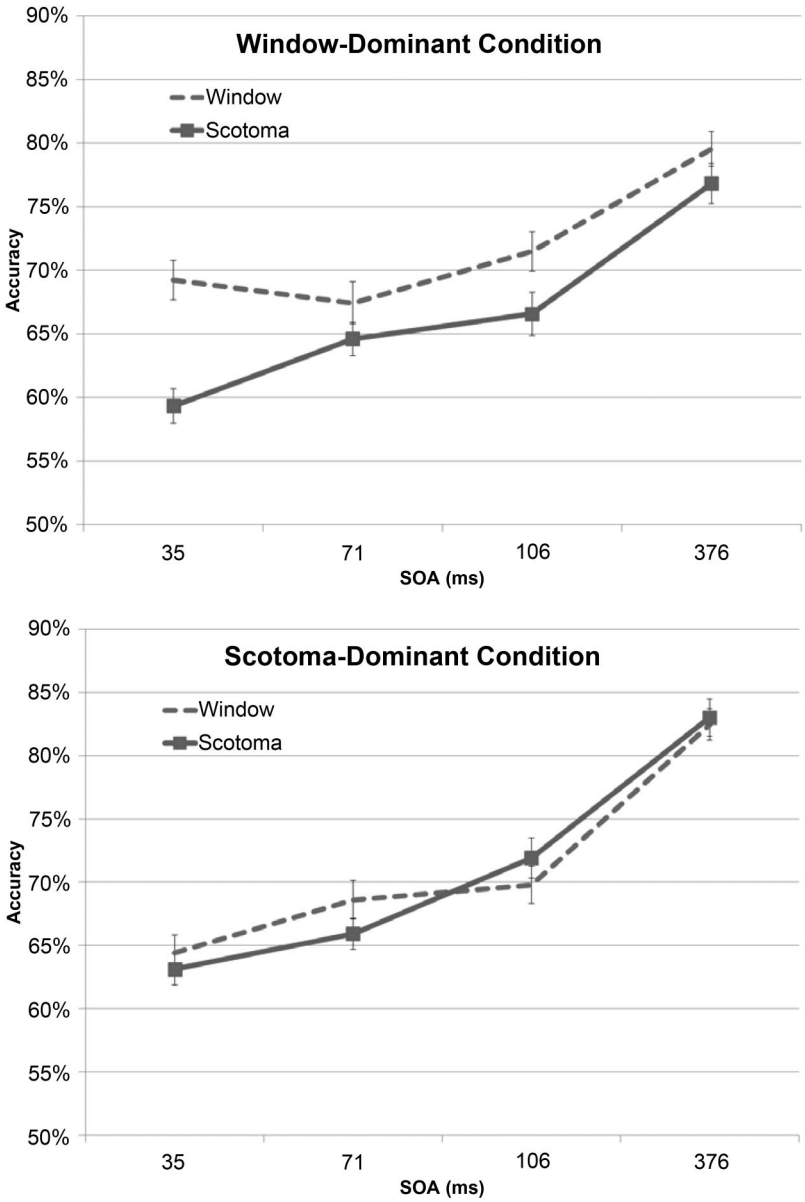

Figure 5. Experiment 3 results showing the window-dominant condition and the scotoma-dominant condition. Error bars represent the standard error for each data point.

The fact that scene gist acquisition occurs within a single fixation—during which time all elements of the retinal image have fixed locations—led us to investigate whether different regions of the visual field, specifically, central versus peripheral vision, contribute differentially to this process over the course of the critical first fixation on a scene. The three experiments in the current study strongly suggest that during the first fixation on a scene, processing of the gist of a scene is influenced by covert attention, which begins focused in central vision and rapidly expands outward into the visual periphery within the first 100 ms of viewing.

Over the past decade, there has been an ongoing debate regarding the role of attention in scene gist recognition. Several previous studies have shown that rapid scene categorization can occur in the near absence of attention for peripheral scenes when attention is focused centrally (Fei-Fei et al., 2005; Li et al., 2002), or multiple scenes in parallel across the field of view (Rousselet et al., 2002). Such findings suggest that attention has little, if any, effect on recognizing scene gist and that both central and peripheral scene information should be equally useful for recognizing scene gist over the course of a single fixation. Conversely, several more recent studies have challenged these conclusions by showing that rapid scene categorization is affected by attentional limits (Cohen et al., 2011; Evans & Treisman, 2005; Walker et al., 2008) and that there are limits to the parallel processing of scene gist (Rousselet, Thorpe, & Fabre-Thorpe, 2004).

Importantly, whether one can identify multiple scenes simultaneously in their visual periphery, or be able to shift their locus of attention from central vision to peripheral vision within a single image, might involve different aspects of attention. The current study investigated changes in processing within a single scene over time. Experiments 1 and 2 showed an asymmetry in processing across time and space that favored central vision early in a fixation, which is *inconsistent* with the claim that scene gist is processed homogeneously across the field of view by preattentive processes over the course of a single fixation. Furthermore, Experiment 3 provided evidence consistent with the hypothesis that the processing asymmetry observed in Experiments 1 and 2 was due to the spatiotemporal dynamics of attentional allocation over the course of a fixation, by showing that spreading attention into the visual periphery eliminated the central advantage, whereas focusing attention in the center of vision maintained the central advantage.

Indeed, the condition least likely to be attended—the scotoma condition at the shortest SOA when information was expected to be presented in a central window—produced the worst performance. Nevertheless, although scene categorization in this condition was extremely poor (.59 accuracy), it was significantly above chance (.50), $t(48) = 5.25$, $p < .001$, Cohen's $d = 1.52$, and thus the results of the current study do not show that attention is *necessary* for scene gist recognition. It should be noted, however, that the attention manipulation in Experiment 3 involved divided attention, rather than inattention, and thus is not a rigorous test of the necessity of attention for scene gist recognition (Cohen et al., 2011; Mack & Rock, 1998) and leaves ample room for future investigation of the role of attention in scene gist processing. The above-chance performance in the condition least likely to be attended is also consistent with the hypothesis that on certain trials, for certain participants, attention was at least moderately allocated to the periphery, thus producing performance slightly, though significantly, above chance.

The current study also sheds light on the role of peripheral vision in scene gist recognition. Previous studies have shown the importance of information from peripheral vision for rapid scene-processing tasks, including animal detection and scene gist recognition (Boucart et al., 2013; Larson & Loschky, 2009; Thorpe et al., 2001; Tran, Rambaud, Despretz, & Boucart, 2010). Furthermore, physiological evidence shows that information from peripheral vision, transmitted primarily by the magnocellular pathway, activates early cortical visual areas prior to information from central vision, primarily from the parvocellular pathway (Carrasco et al., 2003; Livingstone & Hubel, 1988; Nowak et al., 1995). Together, such findings suggest that scene information from peripheral vision may be used earlier for scene gist recognition than information from central vision. However, in the current study, although peripheral vision provided sufficient information for scene gist recognition, that information was used relatively later in a fixation than information from central vision. A question for further studies is why the processing speed advantage for peripheral vision via the magnocellular pathway does not translate into earlier processing of information for scene gist recognition from peripheral vision.

The results of the current study most strongly support the zoom-out hypothesis, which states that during gist recognition, attention proceeds from central vision to peripheral vision over the course of the first fixation on a scene. Other research has investigated the effects on various scene perception tasks of masking either central vision or peripheral vision early in a fixation, and has shown that masking central vision has greater effects on visual search and memory for scenes during the first 50 ms of fixations, whereas masking peripheral vision continues to have effects up through 70 to 100 ms into a fixation (Glaholt et al., 2012; van Diepen & d'Ydewalle, 2003). Those results suggest that scene information in central vision is extracted earlier in a fixation than is information in peripheral vision (Glaholt et al., 2012; van Diepen & d'Ydewalle, 2003). Importantly, the current study contradicts a prediction proposed by van Diepen and d'Ydewalle (2003), namely, that scene identification (gist recognition) would not follow the sequential attention model, but instead would be immediately more biased to peripheral vision. Instead, the current results are consistent with the idea that a wide array of visual tasks, ranging from reading, to visual search, to scene memory, to scene gist recognition, all begin with attention located in central vision, which then proceeds outward into peripheral vision over the first 100 ms of a fixation.

The results of the current study and those of van Diepen and d'Ydewalle (2003) and Glaholt et al. (2012) are consistent with the zoom-out hypothesis, which derives from both the sequential attention model for eye-movements (Henderson, 1992, 1993) and the zoom-lens model of attention (Eriksen & St. James, 1986; Eriksen & Yeh, 1985; Müller, Bartelt, Donner, Villringer, & Brandt, 2003; Seiple, Clemens, Greenstein, Holopigian, & Zhang, 2002). The sequential attention model argues that attention starts focused on the foveal object at the beginning of each fixation and then moves to a parafoveal or peripheral target object before the end of the fixation, and is thus an object-based theory of attention. The zoom-lens model argues that attention can vary from highly focused at the center of vision to broadly diffuse from central to peripheral vision, and is thus a space-based theory of attention. The zoom-out hypothesis argues that scene gist processing proceeds from central vision to peripheral vision over the course of a single fixation, and thus is essentially a space-based account. However, the current study cannot speak to whether the spread of attention outward from the center of vision is spatially relatively uniform or instead is nonuniform and object-based. This is a critical question for future research.

It could be argued that the observed spatiotemporal processing differences in scene categorization in Experiments 1 through 3 could be due to content differences between the central and peripheral image regions of our photographic stimuli. Although possible, this seems highly unlikely for the following reasons. First, we used a critical radius that perfectly divided scene images into central and peripheral regions that produced equal performance when presented for the duration of a single fixation. The results of Experiments 1 through 3 confirmed this assumption of equality after processing for a single fixation, with the central advantage only occurring during the first 100 ms of a fixation. Thus, any content advantage for central vision would have to occur only early in a fixation.

Second, Experiment 3 showed that the central advantage was eliminated when attention was strategically allocated to the visual periphery. The latter finding is hard to reconcile with claims that the central advantage is due to differences in content, but is consistent with the zoom-out hypothesis.[9]

However, an important question for future research is which information is used in central vision and peripheral vision earlier and later in a fixation. Such studies should take into account various well-known differences in processing between central and peripheral vision. For example, central vision is more sensitive than peripheral vision to higher spatial frequencies (Banks, Sekuler, & Anderson, 1991; Cannon, 1985; Loschky, McConkie, Yang, & Miller, 2005; Peli & Geri, 2001; Pointer & Hess, 1989), color information (Mullen & Losada, 1999; Nagy & Wolf, 1993; Rovamo & Iivanainen, 1991), and phase information (Knight, Shapiro, & Lu, 2008), among others. Previous studies have already contributed to our understanding of the roles in scene gist recognition of lower versus higher spatial frequencies (McCotter et al., 2005; Oliva & Schyns, 1997; Schyns & Oliva, 1994), color information (Castelhano & Henderson, 2008; Goffaux et al., 2005; Loschky & Simons, 2004; Oliva & Schyns, 2000; Steeves et al., 2004), and phase information (Guyader, Chauvin, Peyrin, Hérault, & Marendaz, 2004; Joubert, Rousselet, Fabre-Thorpe, & Fize, 2009; Loschky et al., 2007, 2010; Loschky & Larson, 2008; McCotter et al., 2005; Wichmann, Braun, & Gegenfurtner, 2006). Studies that examine the extraction of different types of scene information from central versus peripheral vision over time will be critical to understanding how scene representations are constructed and stored in memory. By knowing which types of information are processed by central versus peripheral vision, and the time course of their use, we can build computational models of scene categorization that more closely reflect how humans perceive and understand their visual environment.

Another interesting question for further research is whether the results shown in the current experiments generalize to more natural viewing conditions in which the viewer is not required to fixate the center of the screen prior to the onset of the image. In natural viewing, the first fixation on a scene is simply wherever the viewer happens to fixate. Thus, one might argue that requiring the participants in the current study to fixate the center of the screen prior to onset of the image could have enforced the zoom-out pattern of attention that starts from central vision. This question cannot be answered by the results of the current studies. However, if requiring viewers to fixate the center of the screen prior to the onset of an image caused attention to zoom out, then this must also hold for the vast majority of previous studies on scene gist recognition, which have similarly required central fixation.

In sum, the present study suggests that the spatiotemporal dynamics of attention across the visual field affect viewers' gist recognition during the first fixation on a scene. The results provide strong support for the zoom-out hypothesis, in which rapid gist extraction from a scene unfolds from central to peripheral vision via visual selective attention. This provides important spatiotemporal constraints for theories of the rapid understanding of real-world scenes.

[9] Additionally, previous research suggests that diagnostic objects in central vision should not be responsible for our reported effects. Specifically, Davenport and Potter (2004) presented scene images with a semantically consistent or inconsistent object presented at the center of vision. They showed that a semantically inconsistent object resulted in worse scene categorization performance, but a semantically consistent object did not facilitate scene categorization compared with the same scene without the central object. Thus, if central objects affected our results, then only semantically inconsistent objects should have had such an influence, and they should have reduced the central advantage we found. However, there were no semantically inconsistent objects within our scene categories.

References

Bacon-Macé, N., Macé, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research, 45,* 1459–1469. doi:10.1016/j.visres.2005.01.004

Banks, M. S., Sekuler, A. B., & Anderson, S. J. (1991). Peripheral spatial vision: Limits imposed by optics, photoreceptors, and receptor pooling. *Journal of the Optical Society of America, A, Optics, Image & Science, 8,* 1775–1787. doi:10.1364/JOSAA.8.001775

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception, 25,* 343–352. doi:10.1068/pp.250343

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

Biederman, I., Rabinowitz, J., Glass, A., & Stacy, E. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103,* 597–600. doi:10.1037/h0037158

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

Boyce, S. J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 531–543. doi:10.1037/0278-7393.18.3.531

Breitmeyer, B. G., & Ogmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision*. Oxford, UK: Clarendon Press.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 207–230. doi:10.1016/0010-0285(81)90008-6

Calvo, M. G. (2005). Relative contribution of vocabulary knowledge and working memory span to elaborative inferences in reading. *Learning and Individual Differences, 15,* 53–65. doi:10.1016/j.lindif.2004.07.002

Calvo, M. G., Nummenmaa, L., & Hyona, J. (2008). Emotional scenes in peripheral vision: Selective orienting and gist processing, but not content identification. *Emotion, 8,* 68–80. doi:10.1037/1528-3542.8.1.68

Cannon, M. W. (1985). Perceived contrast in the fovea and periphery. *Journal of the Optical Society of America, A, Optics, Image & Science, 2,* 1760–1768. doi:10.1364/JOSAA.2.001760

Carrasco, M., McElree, B., Denisova, K., & Giordano, A. M. (2003). Speed of visual processing increases with eccentricity. *Nature Neuroscience, 6,* 699–700. doi:10.1038/nn1079

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 660–675. doi:10.1037/0096-1523.34.3.660

Cohen, J. (1988). *Statistical Power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, M. A., Alvarez, G. A., & Nakayama, K. (2011). Natural-scene perception requires attention. *Psychological Science, 22,* 1165–1172. doi:10.1177/0956797611419168

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology, 292,* 497–523. doi:10.1002/cne.902920402

Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E., & Kalina, R. E. (1987). Distribution of cones in human and monkey retina: Individual variability and radial asymmetry. *Science, 236,* 579–582. doi:10.1126/science.3576186

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

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17,* 973–980. doi:10.1111/j.1467-9280.2006.01815.x

Eriksen, C. W.St., & James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225–240. doi:10.3758/BF03211502

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 583–597. doi:10.1037/0096-1523.11.5.583

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1476. doi:10.1037/0096-1523.31.6.1476

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7,* 1–29. doi:10.1167/7.1.10

Fei-Fei, L., VanRullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12,* 893–924. doi:10.1080/13506280444000571

Francis, G., & Cho, Y. S. (2008). Effects of temporal integration on the shape of visual backward masking functions. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1116–1128. doi:10.1037/0096-1523.34.5.1116

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics, 67,* 1252–1268. doi:10.3758/BF03193557

Girard, P., & Koenig-Robert, R. (2011). Ultra-rapid categorization of Fourier-spectrum equalized natural images: Macaques and humans perform similarly. *PLoS ONE, 6,* e16453. doi:10.1371/journal.pone.0016453

Glaholt, M. G., Rayner, K., & Reingold, E. M. (2012). The mask-onset delay paradigm and the availability of central and peripheral visual information during scene viewing. *Journal of Vision, 12*(1), 9. doi:10.1167/12.1.32

Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition, 12,* 878–892. doi:10.1080/13506280444000562

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 760–777. doi:10.1037/0096-1523.30.4.760

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

Guyader, N., Chauvin, A., Peyrin, C., Hérault, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies, 327,* 313–318. doi:10.1016/j.crvi.2004.02.006

Henderson, J. M. (1992). Visual attention and eye movement control during reading and picture viewing. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (Vol. 10, pp. 260–283). New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-2852-3_15

Henderson, J. M. (1993). Visual attention and saccadic eye movements. In J. V. R. Gery d'Ydewalle (Ed.), *Perception and cognition: Advances in eye movement research. Studies in visual information processing* (Vol. 4, pp. 37–50). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127,* 398–415. doi:10.1037/0096-3445.127.4.398

Hollingworth, A., Schrock, G., & Henderson, J. M. (2001). Change detection in the flicker paradigm: The role of fixation position within the scene. *Memory & Cognition, 29,* 296–304. doi:10.3758/BF03194923

Holmes, D. L., Cohen, K. M., Haith, M. M., & Morrison, F. J. (1977). Peripheral visual processing. *Perception & Psychophysics, 22,* 571–577. doi:10.3758/BF03198765

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 604–610. doi:10.1037/0096-1523.7.3.604

Intraub, H. (1984). Conceptual masking: The effects of subsequent visual events on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 115–125. doi:10.1037/0278-7393.10.1.115

Irwin, D. E. (1992). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 307–317. doi:10.1037/0278-7393.18.2.307

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision, 9*(1), 1–16.

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

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776. doi:10.1016/j.visres.2005.10.002

Knight, E., Shapiro, A., & Lu, Z.-L. (2008). Drastically different percepts of five illusions in foveal and peripheral vision reveal their differences in representing visual phase. *Journal of Vision, 8,* 967. doi:10.1167/8.6.967

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *Journal of Vision, 9*(10), 1–16.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 99,* 9596–9601. doi:10.1073/pnas.092277599

Livingstone, M., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science, 240,* 740–749. doi:10.1126/science.3283936

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimari, T. (2010). The role of higher-order image statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics, 72,* 427–444. doi:10.3758/APP.72.2.427

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), 1–9.

Loschky, L. C., & Larson, A. M. (2010). The natural/man-made distinction is made prior to basic-level distinctions in scene gist processing. *Visual Cognition, 18,* 513–536. doi:10.1080/13506280902937606

Loschky, L. C., Larson, A. M., Smerchek, S., & Finan, S. (2008). The superordinate natural/man-made distinction is perceived before basic level distinctions in scene gist recognition. *Journal of Vision, 8*(6), 738. doi:10.1167/8.6.738

Loschky, L. C., McConkie, G. W., Yang, J., & Miller, M. E. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition, 12,* 1057–1092. doi:10.1080/13506280444000652

Loschky, L. C., Sethi, A., Simons, D. J., Pydimari, T., Ochs, D., & Corbeille, J. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 1431–1450. doi:10.1037/0096-1523.33.6.1431

Loschky, L. C., & Simons, D. J. (2004). The effects of spatial frequency content and color on scene gist perception [Abstract]. *Journal of Vision, 4*(8), 881. doi:10.1167/4.8.881

Mack, A., & Rock, I. (1998). *Inattentional blindness* (Vol. 6). Cambridge, MA: MIT Press.

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2011). Ordinate and subordinate level categorizations of real-world scenes: An eye movement study. *Journal of Vision, 11*(11), 1112. doi:10.1167/11.11.1112

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17,* 578–586. doi:10.3758/BF03203972

McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition, 12,* 938–953. doi:10.1080/13506280444000599

Mullen, K. T., & Losada, M. A. (1999). The spatial tuning of color and luminance peripheral vision measured with notch filtered noise masking. *Vision Research, 39,* 721–731. doi:10.1016/S0042-6989(98)00171-0

Müller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A physiological correlate of the "zoom lens" of visual attention. *The Journal of Neuroscience, 23,* 3561–3565.

Nagy, A. L., & Wolf, S. (1993). Red–green color discrimination in peripheral vision. *Vision Research, 33,* 235–242. doi:10.1016/0042-6989(93)90161-O

Nowak, L. G., Munk, M. H. J., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience, 12,* 371–384. doi:10.1017/S095252380000804X

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

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

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155,* 23–36.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmologica (Suppl), 6,* 11–96.

Otsuka, S., & Kawaguchi, J. (2007). Natural scene categorization with minimal attention: Evidence from negative priming. *Perception & Psychophysics, 69,* 1126–1139. doi:10.3758/BF03193950

Peli, E., & Geri, G. A. (2001). Discrimination of wide-field images as a test of a peripheral-vision model. *Journal of the Optical Society of America, A, Optics, Image Science, and Vision, 18,* 294–301. doi:10.1364/JOSAA.18.000294

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 587–595. doi:10.1037/0278-7393.15.4.587

Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research, 29,* 1133–1151. doi:10.1016/0042-6989(89)90061-8

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40,* 49–71. doi:10.1023/A:1026553619983

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522. doi:10.1037/0278-7393.2.5.509

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science, 206,* 468–469. doi:10.1126/science.504987

Rayner, K., Inhoff, A. W., Morrison, R. E., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 167–179. doi:10.1037/0096-1523.7.1.167

Rayner, K., Liversedge, S. P., & White, S. J. (2006). Eye movements when reading disappearing text: The importance of the word to the right of fixation. *Vision Research, 46,* 310–323. doi:10.1016/j.visres.2005.06.018

Rayner, K., Liversedge, S. P., White, S. J., & Vergilino-Perez, D. (2003). Reading disappearing text: Cognitive control of eye movements. *Psychological Science, 14,* 385–388. doi:10.1111/1467-9280.24483

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science, 20,* 6–10. doi:10.1111/j.1467-9280.2008.02243.x

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research, 44,* 2301–2311.

Rouder, J. N., Speckman, P., Sun, D., Morey, R., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. doi:10.3758/PBR.16.2.225

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629.

Rousselet, G. A., Joubert, O. R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition, 12,* 852–877. doi:10.1080/13506280444000553

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research, 44,* 877–894. doi:10.1016/j.visres.2003.11.014

Rovamo, J., & Iivanainen, A. (1991). Detection of chromatic deviations from white across the human visual field. *Vision Research, 31,* 2227–2234. doi:10.1016/0042-6989(91)90175-5

Schmolesky, M. T., Wang, Y. C., Hanes, D. P., Thompson, K. G., Leutgeb, S., Schall, J. D., & Leventhal, A. G. (1998). Signal timing across the macaque visual system. *Journal of Neurophysiology, 79,* 3272–3278.

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

Seiple, W., Clemens, C., Greenstein, V. C., Holopigian, K., & Zhang, X. (2002). The spatial distribution of selective attention assessed using the multifocal visual evoked potential. *Vision Research, 42,* 1513–1521. doi:10.1016/S0042-6989(02)00079-2

Shimozaki, S. S., Chen, K. Y., Abbey, C. K., & Eckstein, M. P. (2007). The temporal dynamics of selective attention of the visual periphery as measured by classification images. *Journal of Vision, 7*(12), 1–20. doi:10.1167/7.12.10

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

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision, 11,* 13. doi:10.1167/11.5.13

Thorpe, S. J., Gegenfurtner, K. R., Fabre-Thorpe, M., & Bulthoff, H. H. (2001). Detection of animals in natural images using far peripheral

vision. *European Journal of Neuroscience, 14,* 869–876. doi:10.1046/j.0953-816x.2001.01717.x

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786. doi:10.1037/0033-295X.113.4.766

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

van Diepen, P. M., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition, 10,* 79–100. doi:10.1080/713756668

van Diepen, P. M., & Wampers, M. (1998). Scene exploration with Fourier-filtered peripheral information. *Perception, 27,* 1141–1151. doi:10.1068/pp.271141

van Diepen, P. M., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337–355). Oxford, UK: Anonima Romana. doi:10.1016/B978-008043361-5/50016-X

Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Scenes with multiple foreground objects. *Journal of Vision, 8,* 1–12. doi:10.1167/8.4.21

Westheimer, G. (1982). The spatial grain of the perifoveal visual field. *Vision Research, 22,* 157–162. doi:10.1016/0042-6989(82)90177-8

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011a). Statistical evidence in experimental psychology: An empirical comparison using 855 t tests. *Perspectives on Psychological Science, 6,* 291–298. doi:10.1177/1745691611406923

White, A. L., Rolfs, M., & Carrasco, M. (in press). Adaptive deployment of spatial and feature-based attention before saccades. *Vision Research, 85,* 26–35. doi:10.1016/j.visres.2012.10.017

Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. (2006). Phase noise and the classification of natural images. *Vision Research, 46,* 1520–1529. doi:10.1016/j.visres.2005.11.008

Wilson, H. R., Levi, D., Maffei, L., Rovamo, J., & DeValois, R. (1990). The perception of form: Retina to striate cortex. In L. Spillmann & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 231–272). San Diego, CA: Academic Press.

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 661–676. doi:10.1037/0096-1523.25.3.661

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601–621. doi:10.1037/0096-1523.10.5.601

Received October 1, 2012
Revision received September 27, 2013
Accepted October 1, 2013 ■