The contributions of central versus peripheral vision to scene gist recognition

Adam M. Larson   Department of Psychology, Kansas State University, USA

Lester C. Loschky   Department of Psychology, Kansas State University, USA

Which region of the visual field is most useful for recognizing scene gist, central vision (the fovea and parafovea) based on its higher visual resolution and importance for object recognition, or the periphery, based on resolving lower spatial frequencies useful for scene gist recognition, and its large extent? Scenes were presented in two experimental conditions: a "Window," a circular region showing the central portion of a scene, and blocking peripheral information, or a "Scotoma," which blocks out the central portion of a scene and shows only the periphery. Results indicated the periphery was more useful than central vision for maximal performance (i.e., equal to seeing the entire image). Nevertheless, central vision was more efficient for scene gist recognition than the periphery on a per-pixel basis. A critical radius of 7.4° was found where the Window and Scotoma performance curves crossed, producing equal performance. This value was compared to predicted critical radii from cortical magnification functions on the assumption that equal V1 activation would produce equal performance. However, these predictions were systematically smaller than the empirical critical radius, suggesting that the utility of central vision for gist recognition is less than predicted by V1 cortical magnification.

Introduction

People can recognize the gist of a real world scene, here operationalized as the ability to categorize it at the basic level with a single word or phrase, after viewing it for an extremely short period of time (i.e., within a single fixation), for example a masked duration of less than 50 ms (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005; Fei-Fei, Iyer, Koch, & Perona, 2007; Loschky et al., 2007).[1] Scene gist recognition is important because it activates scene schemas which affect later critical cognitive processes, such as directing attention within a scene (Eckstein, Drescher, & Shimozaki, 2006; Gordon, 2004; Torralba, Oliva, Castelhano, & Henderson, 2006), facilitating object recognition (Boyce & Pollatsek, 1992; Davenport & Potter, 2004; but see Hollingworth & Henderson, 1998), and long-term memory for objects within the scene (Brewer & Treyens, 1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). Since scene gist recognition is important to so many other cognitive processes and gist is acquired extremely rapidly, the key question examined here is, how do differing fields of view contribute to scene gist recognition?

Looking at a real world scene takes up the entire visual field, although when displayed on a computer screen they tend to be much smaller. Of particular interest to gist recognition is the fact that visual resolution drops off dramatically with distance from the center of vision. As shown in Figure 1, this drop-off of visual resolution is strongly related to several anatomically distinct regions of the visual field, each defined relative to the center of vision: the fovea, parafovea, and the periphery (Polyak, 1941; Rodieck, 1998). The fovea is a tiny region at the center of vision with the highest visual resolution, which extends to the boundary of the rod-free area of the retina (~1° eccentricity) (Polyak, 1941; Yamada, 1969). The parafovea extends to the foveal rim, where the highest density of rods is found (~4–5° eccentricity) (Beard & Ahumada, 1999; Coletta & Williams, 1987; Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981). Vision scientists most commonly refer to all vision beyond the parafovea as peripheral vision (e.g., Hollingworth, Schrock, & Henderson, 2001; e.g., Holmes, Cohen, Haith, & Morrison, 1977; Rayner et al., 1981; Shimozaki, Chen, Abbey, & Eckstein, 2007; van Diepen & Wampers, 1998), which we will do as well.[2] Thus, there is a strong inverse relationship between the size of the visual region and the quality of visual information it provides as one moves from foveal to parafoveal to peripheral vision.

A key question that has received little attention is what is the relative importance for recognizing scene gist of central vision (foveal and parafoveal), versus peripheral vision? What is the relative efficiency of gist information from central versus peripheral vision? Finally, how well is gist recognition performance predicted by mathematically
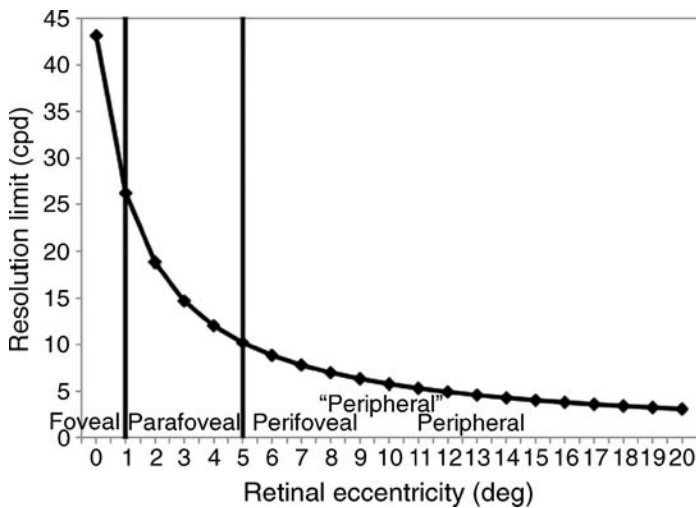
Figure 1. The limits of visual resolution as a function of retinal eccentricity, with the visual field divided into three regions: Foveal, Parafoveal, and Peripheral. Resolution limits are in terms of spatial frequency cut-offs in cycles/degree (cpd), based on the results of Loschky et al. (2005).

defined relationships between retinal eccentricity and vision, such as cortical magnification functions (described below)?

Central versus peripheral vision and scene gist recognition

A great deal is known about the neurobiology of central versus peripheral vision, and we will only briefly touch on a few highly salient points of contrast here (for an excellent review, see Wilson, Levi, Maffei, Rovamo, & DeValois, 1990). The drop-off in visual resolution with retinal eccentricity, shown in Figure 1, can be explained in several ways. First, the density of retinal receptors, particularly cones, is far greater in foveal and parafoveal vision than in the visual periphery, allowing central vision to encode higher spatial frequencies. Then, the pooling of information from retinal receptors by retinal ganglion cells is far greater in the visual periphery than in foveal or parafoveal vision, leading to a loss of visual resolution (i.e., higher spatial frequencies) in the periphery. Consequently, both the lateral geniculate nucleus (LGN) and the primary visual cortex (V1) devote many more cells to processing central vision than to peripheral vision, a fact known as *cortical magnification*. Because of this, small detailed pattern information, encoded by higher spatial frequencies, can be processed in central vision, whereas the same pattern information must be enlarged, and encoded by lower spatial frequencies, to be resolved in the visual periphery (Virsu, Näsänen, & Osmoviita, 1987; Virsu & Rovamo, 1979). It is for this reason that people *foveate* (or fixate) objects in scenes in order to

recognize them. Studies have shown that perception of an object is best when viewers fixate within 1–2° of it, with performance dropping off rapidly with increasing distance from the closest fixation to the object (Henderson & Hollingworth, 1999; Hollingworth et al., 2001; Nelson & Loftus, 1980; O'Regan, Deubel, Clark, & Rensink, 2000; Pringle, 2000).[3] If gist recognition requires the use of detailed information (specifically, spatial frequencies > 10 cpd), then central information will be important, since high spatial frequencies are only processed centrally (as seen in Figure 1). Interestingly, a study by Oliva and Schyns (1997) showed that higher spatial frequency information can be highly useful for scene gist recognition when it provides diagnostic information, suggesting a possible key role for central vision in scene gist recognition.

Nevertheless, the issue of which spatial frequency band, or spatial scale, of scene information is most useful for scene gist recognition suggests a reason to argue for the importance of peripheral vision for gist. Figure 1 shows that to the extent that lower spatial frequencies (≤10 cpd) are important for scene gist recognition, peripheral vision will be important, because it can provide them. Several studies bear on this issue. Schyns and Oliva (1994) showed a bias for viewers to recognize scenes encoded by lower spatial frequencies early in processing, with recognition based on higher frequencies occurring only later in processing.[4] Using very different methods, McCotter, Gosselin, Sowden, and Schyns (2005) showed that the information most important for recognizing scene gist (specifically, structure encoded in the phase spectrum) was contained in the lower spatial frequencies. Similarly, studies of scene gist masking have shown that visual masks containing predominantly lower spatial frequencies are the most effective at disrupting scene gist recognition (Harvey, Roberts, & Gervais, 1983; Loschky et al., 2007). Furthermore, the fact that people can recognize the gist of scenes containing only lower spatial frequency information, yet cannot recognize individual objects in those scenes, has been used to argue that object recognition may not be needed for scene gist recognition (Schyns & Oliva, 1994), and instead the global layout of the scene may be more important (Sanocki, 2003; Sanocki & Epstein, 1997). The above suggests that peripheral vision may be important for recognizing scene gist, since it can resolve the lower spatial frequencies so useful for gist. However, one could go further and argue that peripheral vision might be better at recognizing scene gist than central vision because the periphery contains a larger expanse of the visual field from which to gather such information.

Using windows and scotomas to study scene gist recognition

A number of studies have investigated the roles of central versus peripheral vision in scene perception using what are
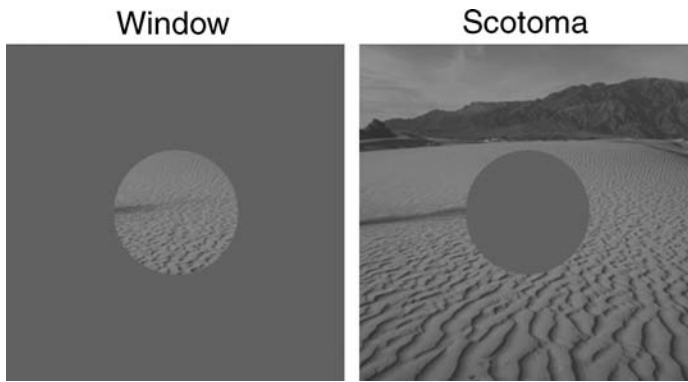
Figure 2. Examples of a *Window* and a *Scotoma* image.

known as the "Window" and "Scotoma" paradigms (for review, see van Diepen, Wampers, & d'Ydewalle, 1998). Examples of each are shown in Figure 2. The *Window* paradigm gets its name by analogy to viewing a scene through a Window, for example a porthole. Unaltered imagery is presented within the Window, which is centered on the viewers' fovea, while outside the Window, imagery is either absent (Saida & Ikeda, 1979) or degraded by adding noise, image filtering, or other means (Kortum & Geisler, 1996; Loschky & McConkie, 2002; Parkhurst, Culurciello, & Neiburm, 2000; Shioiri & Ikeda, 1989; van Diepen & Wampers, 1998). The *Scotoma* is an inverted Window, where central information is blocked from view, while information outside the Scotoma is unaltered (Henderson, McClure, Pierce, & Schrock, 1997; van Diepen, Ruelens, & d'Ydewalle, 1999). This term is taken from an analogous medical condition where a specific region of the visual field is degraded or blocked from view. Window and Scotoma paradigms were first used to study reading processes (McConkie & Rayner, 1975; Rayner & Bertera, 1979), and have been later used to study scene perception (Geisler & Perry, 1999; Henderson et al., 1997; Loschky & McConkie, 2002; Loschky, McConkie, Yang, & Miller, 2005; Parkhurst et al., 2000; Parkhurst & Niebur, 2002; Reingold, Loschky, McConkie, & Stampe, 2003; Shioiri & Ikeda, 1989; van Diepen & Wampers, 1998; van Diepen et al., 1998). They allow one to use real-world scenes in meaningful tasks, while varying which regions of the visual field provide information. The logic of both paradigms is that processing will be disrupted to the extent that the missing information is needed for the task, whereas processing will be normal to the extent that the missing information is not needed.

Window and Scotoma paradigms have typically been used dynamically to investigate scene perception over the course of multiple fixations. However, they can also be used statically to study perception within the first fixation on a scene, when gist is recognized. However, to our knowledge, only one previous study has taken such an approach to study scene gist recognition (van Diepen, De Graef, Lamote, & van Wijnendaele, 1994). That study used simple line drawings of scenes (measuring 16° × 12°), which were flashed for 160 ms, and compared a control condition (whole image) with a Window condition and a Scotoma condition, both of which were of fixed size (a 3.5° × 3° radius oval). Scene gist recognition was measured implicitly by asking participants whether a pre-cued object name would fit in the briefly flashed scene. Scene gist recognition in the control condition was superior to the Window condition, but not the Scotoma condition, and there was no significant difference between Window and Scotoma conditions, though there was a trend favoring the Scotoma condition. The results are unclear but suggest that peripheral vision may be more important for scene gist recognition than central vision. However, van Diepen et al.'s (1994) study leaves important unanswered questions. At a methodological level, line drawings consist primarily of higher spatial frequencies unlike photographic scenes, which generally include a wide range of frequencies. Thus, van Diepen et al.'s (1994) study cannot address arguments regarding the importance for gist of central versus peripheral vision that hinge on the usefulness of high frequency versus low frequency information. In addition, the study used only a single Window and Scotoma radius. Thus, it would be useful to include a range of Window and Scotoma radii that more clearly differentiate between foveal, parafoveal, and peripheral vision to determine the relative importance of each for gist recognition. Varying the Window and Scotoma sizes would also allow one to plot psychometric functions for each, from which one could then estimate parameters of interest such as the Window and Scotoma radii that produce equal performance to a whole image control condition, or the cross-over point where both Window and Scotoma produce equal scene gist recognition. As we will see, doing so would allow one to ask and answer more detailed questions about the roles of central versus peripheral vision in scene gist recognition.

Experiment 1

The current experiment investigated the relative value of central versus peripheral vision for scene gist recognition, using a Window and Scotoma paradigm, statically within a single fixation, like van Diepen et al. (1994), as illustrated in Figure 2. However, the current study used photographic images as stimuli, rather than line drawings, in order to provide viewers with a wide range of spatial frequencies to use in categorizing scenes. Scene images were briefly flashed with varying sizes of Window or Scotoma, which matched our previous definitions of "foveal," "parafoveal," and "peripheral" vision (as illustrated in Figure 1). Then, after each scene presentation, participants were queried as to the basic level category of the scene (e.g., "Beach," "Forest," or "Street"). We used these methods to determine whether central or peripheral vision was more useful for recognizing scene gist, and the

extent to which each was needed to achieve maximum performance, i.e., equal to the control condition in which the entire image was presented.

In addition, the current experiment investigated whether any differences in performance between the Window and Scotoma conditions might be explainable simply in terms of the viewable scene area inside the Window, or outside of the Scotoma. Specifically, we used large enough images that in Window conditions of 1° radius (foveal) or 5° radius (foveal + parafoveal), the viewable area would be a relatively small proportion of the original image area, whereas in the corresponding 1° or 5° radius Scotoma conditions, most of the original image content would still be viewable. We therefore included two other Window and Scotoma radii that would either 1) have equal viewable areas in both the Window and Scotoma conditions, or 2) have more viewable area in the Window condition than in the Scotoma condition. In this way, we could determine whether simply increasing the viewable area always produced better performance. As a corollary test of this hypothesis, we could determine whether having equal viewable area produced equal performance, or whether it produced superior performance for image content presented either centrally or peripherally.

Method

Participants

100 undergraduates from Kansas State University (67 Females) participated in the experiment for course credit. Participants had a mean age of 19.14 ($SD = 1.49$), and all had a visual acuity of 20/30 or better.

Materials

Ten scene image categories (5 Natural: Beach, Desert, Forest, Mountain, and River; 5 Man-made: Farm, Home, Market, Pool, and Street) were drawn from the Correl database. There were 32 scene images per category, for a total of 320 scene images. All scene images were equalized for mean luminance and RMS contrast (for details, see Appendix B of Loschky et al., 2007). All scene images were 674 × 674 pixels and viewed at a distance of 53.34 cm, subtended an area of 27° × 27° of visual angle (extending to 13.5° eccentricity on the vertical and horizontal axes), and were presented on 85 Hz SyncMaster 957 MBS monitors.

Scene images were manipulated to place either a Window or a Scotoma at the center of the scene image (see Figure 3). The radii of the Windows and Scotomas were 1°, 5°, 10.8°, or 13.6°. The 1° radius conditions manipulated the presence or absence of foveal vision; the 5° radius conditions manipulated the presence or absence of central (foveal + parafoveal) vision versus peripheral vision; the 10.8° radius conditions presented equal viewable area inside the Window and outside the Scotoma; and the 13.6° radius conditions presented more viewable area in the Window condition than the Scotoma condition.

Design and procedures

In order to avoid fatiguing our subject pool participants, we separated the 2(+1) (Window/Scotoma (+ Control)) × 4 (Radius: 1°, 5°, 10.8°, 13.6°) × 10 (basic level categories) × 2 (cue validity) design in the following
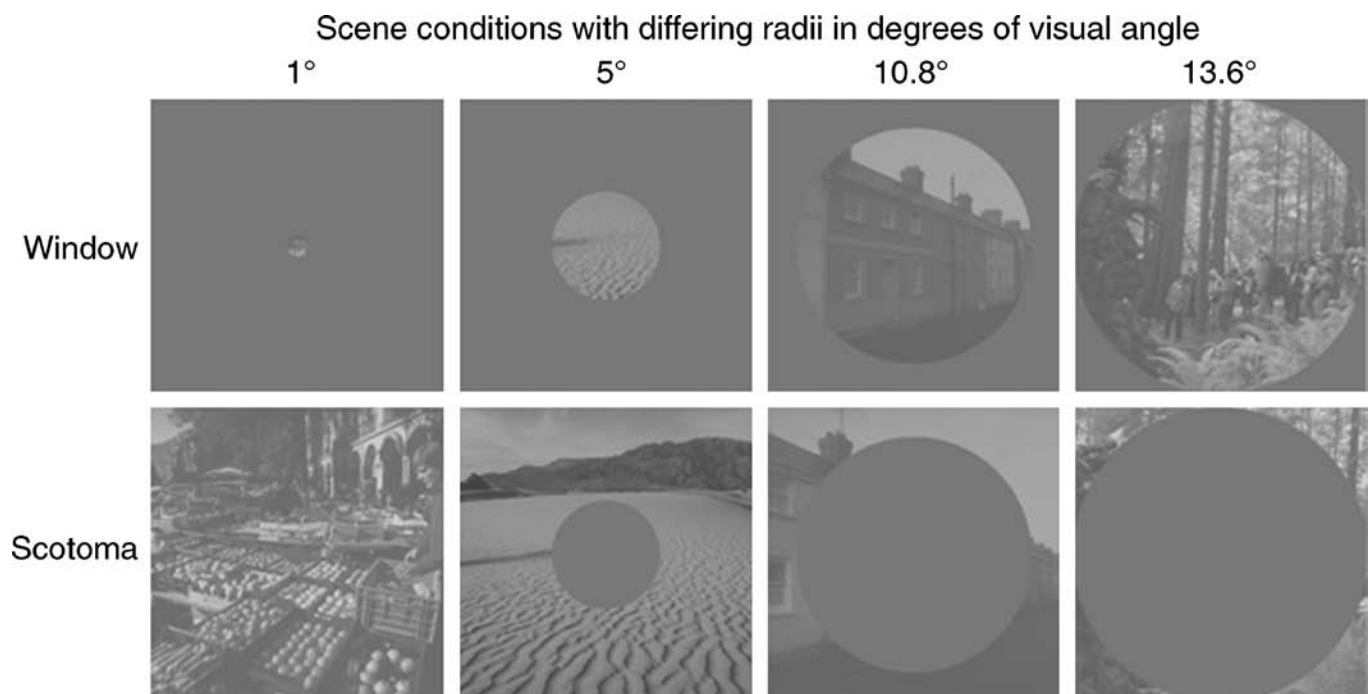

Figure 3. Examples of the *Windows* and *Scotomas* of differing radii, in degrees visual angle, in Experiment 1.

way. There were two between-subjects factors: Image condition (2 levels: Experimental [Window/Scotoma] vs. Control) × Natural/Man-made (2 levels). Participants were randomly assigned to one of the four conditions. Splitting of the Natural/Man-made factor between subject groups reduced the number of trials for each subject by half. Each of the two experimental conditions (Natural or Man-made) contained two within-subjects factors: Window/Scotoma (2 levels) × Radius (4 levels: 1°, 5°, 10.8°, 13.6°), with each subject having an equal number of trials for each of the 8 levels. The two control conditions (Natural or Man-made) only differed from the experimental conditions in that they always displayed the entire scene image. Each scene image was presented twice, once each in the Window and Scotoma conditions (twice each in the control conditions), with the order of trials randomized.

There were 320 self-paced trials in the experiment. Figure 4 shows a schematic of the events in a trial. Participants were instructed to look at a fixation cross at the center of the screen, and press a "Next" button to initiate the trial. After 750 ms, a scene image was flashed for 106 ms, a duration long enough to generally achieve asymptotic scene gist recognition (Bacon-Mace et al., 2005; Loschky et al., 2007), but too short to move their eyes. After another 750 ms, a scene category post-cue (e.g., "Beach," "Forest," or "Street") was presented, which matched the preceding scene image category 50% of the time. Participants were instructed to press the "Yes" button if the cue matched the scene image, or the "No" button if it did not, and they were encouraged to guess if they were unsure. Before starting the experiment, participants

were first familiarized with the scene image categories by presenting each category label together with two example scene images, and they then carried out 30 practice trials, both using images not included in the experimental image sets. After completing the experiment, participants were debriefed and thanked for their participation.

Results

We first carried out a 2 × 2 between-subjects ANOVA to determine whether the Natural/Man-made factor had a significant main effect, or interacted with the Experimental/Control condition factor. This analysis showed that, overall, viewers were slightly (2%) more accurate at determining the gist of Man-made scenes than Natural scenes ($F(1, 94) = 9.85$, $p = .002$), but that this effect did not significantly differ between the Experimental and Control conditions (Interaction: $F(1, 94) < 1$, $p = .496$). However, follow-up post-hoc comparisons (Bonferroni corrected alpha level of $p < .025$) revealed that the difference between Natural and Man-made conditions was not significant for the Experimental condition ($t(49) = 1.430$, $p = .159$), but was significant for the Control condition ($t(45) = 4.107$, $p < .001$), although both group means were greater than 94 percent accurate with a mean difference of 2 percent indicating a ceiling effect for both levels. Therefore, for all further analyses, we have combined data across the Natural and Man-made groups.
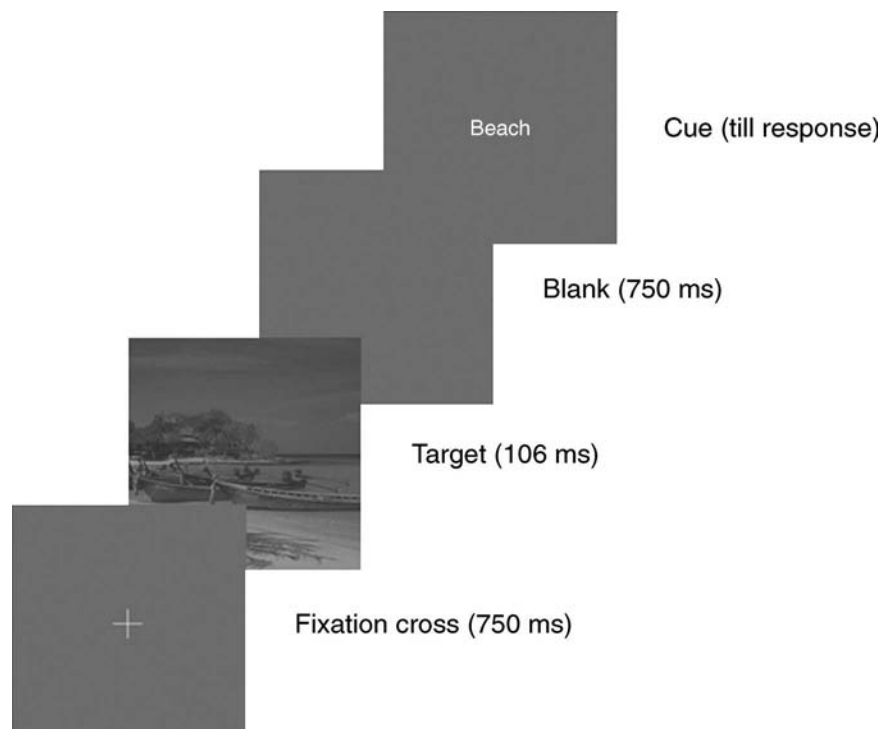

Figure 4. Trial schematic for Experiment 1.

As can be seen in Figure 5, the Window and Scotoma conditions produced very different results ($F(1, 150)$ = 145.458, $p < .001$), as did the different radii of the Window and Scotoma conditions ($F(3, 150)$ = 204.354, $p < .001$), and these two factors strongly interacted ($F(3, 150)$ = 530.152, $p < .001$). Specifically, Figure 5 shows that as the radius of a Window increases, performance increases monotonically, while the opposite is observed with Scotoma images—as a Scotoma radius increases performance *decreases* monotonically. We probed the exact nature of this interaction using Bonferroni corrected t-tests (using the corrected critical p value of .004). First, we compared performance between the Window and Scotoma conditions at each radius, which are shown in Table 1. All comparisons were significantly different ($ts(50) > 4.453$, $ps < .001$, Cohen's $ds > 0.62$).

Next, we compared performance between each Window (or Scotoma) condition at a given radius with the control condition. These comparisons showed that the Window condition was significantly different from the control condition at 1°, 5°, and 10.8° radius ($ts(96) > 3.41$, $ps \leq$ .001, Cohen's $ds > 0.69$) but not at 13° radius ($t(96)$ = 0.38, $p = .70$, Cohen's $d = 0.08$). Conversely, the Scotoma condition differed from the control condition only at 11° and 13° radius ($ts(96) > 7.61$, $ps < .001$, Cohen's $ds >$ 1.53), but not at 1° and 5° radius ($ts(96) < 1.02$, $ps > .31$, Cohen's $ds < 0.18$). The importance of this pattern of results is discussed in detail in the Discussion.

Discussion

The first purpose of this study was to determine whether central (i.e., foveal, or parafoveal) or peripheral vision are more useful for scene gist recognition. First, consider foveal vision, which has the greatest visual resolution, and
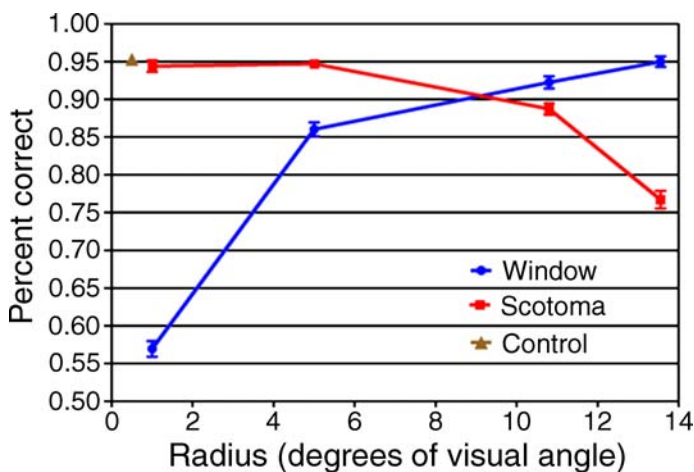
Figure 5. Scene gist accuracy as a function of viewing condition (Window vs. Scotoma vs. Control condition) and radius (of Window or Scotoma) for Experiment 1. Error bars = Standard Error of the Mean (*SEM*).

Radius	Window		Scotoma	
	Mean	St. Dev	Mean	St. Dev
1°	0.570[a]	0.074	0.944	0.056
5°	0.860[a]	0.066	0.947	0.040
10.8°	0.922[a]	0.057	0.887[a]	0.055
13.6°	0.950	0.050	0.767[a]	0.083

Table 1. Descriptive Statistics of Window and Scotoma scene images at each radius level. *Note*: All differences between Window and Scotoma conditions at each radius are significant at $p < .001$. [a]Denotes a significant difference from the control condition ($M = 0.953$, $SD = 0.023$).

is particularly useful for pattern and object recognition. As shown in Figure 5, when participants were presented with only foveal information (the 1° Window condition) of a scene, they were barely above chance at recognizing its gist, which is meaningfully and significantly worse than the control condition. Thus, high resolution foveal vision, which is extremely useful for object recognition, is clearly *not useful* for recognizing gist. Conversely, when participants had information from everything but foveal vision (i.e., the 1° Scotoma condition), they had no problem at all recognizing scene gist, with performance being identical to that of the control condition—i.e., blocking out the central 1° of vision resulted in no loss of scene gist recognition. Therefore, we conclude that although foveal vision has the highest resolution and most acute pattern vision, it is *not useful* for scene gist recognition.

Next consider central versus peripheral vision. Figure 5 shows that although adding information from parafoveal vision to that from foveal vision (i.e., the 5° Window condition) increased scene gist recognition well above chance, it was still meaningfully and significantly worse than normal gist recognition (i.e., the control condition, which shows a very large Cohen's d [1.88] for the difference from the 5° radius Window condition). Thus, central (i.e., foveal + parafoveal) vision is not enough for maximal scene gist recognition. Now consider the results in terms of peripheral vision (i.e., >5° eccentricity). When participants only viewed scene content in peripheral vision (the 5° Scotoma), performance was still identical to the control condition—i.e., peripheral vision is *more useful* for attaining maximal gist recognition. Therefore, despite the very limited visual resolution of peripheral vision, which in this study extended from 5° to 13.6° eccentricity and therefore was capable of resolving information between roughly 10 cpd to 5 cpd (as shown in Figure 1), we conclude that it is very useful for scene gist recognition. Conversely, despite its much greater visual resolution and its extreme usefulness for object recognition, central vision is less useful for recognizing the gist of a scene. Thus, low-resolution peripheral vision is more useful than high-resolution central vision for recognizing the gist of a scene.

Having established that peripheral vision is more important for scene gist recognition than central vision,

we can dig deeper and try to explain why this is so. Earlier, we discussed a number of reasons why central or peripheral vision could be more important for scene gist recognition than central vision. Several hypotheses based on that earlier discussion are directly testable based on the current results.

Hypothesis 1. More area is better. More image area (in pixels) produces better performance, less area produces worse performance, and equal area produces equal performance.

Hypothesis 1 is not particularly theoretically motivated, but instead is based on the methodology of the current experiment. Using Occam's Razor, Hypothesis 1 would seem to be the simplest explanation for the advantage of peripheral vision over central vision. It is therefore a default hypothesis, which can be compared with the following more theoretically motivated competing hypotheses.

Hypothesis 2. Peripheral vision has an advantage over central vision *only because* peripheral vision has more area.

Hypothesis 2 draws on the earlier stated ideas that a) low frequencies are important for scene gist recognition, and b) peripheral vision has much more area than central vision from which to gather such low frequency information. An important corollary hypothesis is if peripheral vision did not have more area, it would not have an advantage over central vision.

Hypothesis 3. Central vision is privileged, and more efficient relative to peripheral vision, in that less area is needed in central vision to attain a given level of gist recognition.

Hypothesis 3 draws on the large body of well-established research, discussed earlier, showing that a greater proportion of both the LGN and V1 are devoted to processing central vision than to processing peripheral vision.

We first consider the default hypothesis, Hypothesis 1. As noted above, Figure 5 clearly shows gist performance monotonically increased as Window radius increased, and decreased as Scotoma radius increased, with both results being consistent with Hypothesis 1 that "more area is better." However, Hypothesis 1 cannot provide a complete explanation of our data. A stronger test of all three Hypotheses can be accomplished by asking if we eliminate the gist advantage for peripheral vision over central vision if we ensure that *equal* image areas are presented centrally and peripherally. As discussed earlier, when both Window and Scotoma conditions had equal viewable areas (in the 10.8° radius conditions), the Window condition showed a moderate difference that was significantly better than the Scotoma condition. This result is inconsistent with Hypothesis 1, that "more area is better," since it predicts that equal areas should produce equal performance. On the other hand, the result is consistent with Hypothesis 2, that the peripheral gist advantage is due *only* to having greater area in the

periphery. The result is also consistent with Hypothesis 3, in that a given equal area produced better performance in the centrally presented condition. The latter point becomes even clearer by noting in Figure 5 that the two monotonic gist accuracy curves for the Window and Scotoma conditions cross at an eccentricity of roughly 9°, which is *less than* the 10.8° radius equal area conditions. A 95% confidence interval was calculated for the mean of the cross-over point (8.73°), to determine if it contains the 10.8° radius of the equal area condition. The confidence interval ranged from 8.22° to 9.24°, which suggests that the 10.8° radius is significantly larger than the estimated cross-over point. The crossing point implies that there should be *equal* performance in Window and Scotoma conditions in which there is *more* viewable area presented peripherally than centrally. This also contradicts Hypothesis 1 that "more area is better" since more area in the periphery would be predicted to produce better, not equal performance. Instead, that result is consistent with Hypothesis 3 that centrally presented image content is privileged, in that less area is needed to achieve a given level of gist recognition. Such a conclusion would be generally consistent with the idea that cortical magnification plays a role in scene gist recognition.

Nevertheless, we must be cautious in drawing the above conclusions since they are primarily based on data in the 10.8° radius conditions. We therefore ran a control experiment, in which we included another pair of Window and Scotoma conditions having equal area. The viewable area in the 10.8° radius Window and Scotoma conditions of Experiment 1 were equalized, where 50% of the scene was viewable either inside the Window or outside the Scotoma. For the control experiment, we wanted to use a different percentage of viewable area, and therefore decided to use less, in this case 40% (Window = 8.53°; Scotoma = 10.54°), in order to avoid a ceiling effect. In addition, all images were made circular, with a maximum outer radius of 13.55° in the Scotoma condition (see Figure 6) so that, like the Window condition, the retinal eccentricity was constant along the outer edge of the Scotoma condition. We used the same scene categories and images as in Experiment 1, and participants were again divided between Natural and Man-made image categories, with 19 participants in the Man-made condition and 23 in the Natural condition. A 2-way ANOVA for Natural/Man-made × Window/Scotoma was conducted to determine if different categories of scenes produced significantly different responses. The analysis found no main effect for the Natural vs. Man-made factor. There was a significant interaction ($p = .048$) with the Window vs. Scotoma factor, but when probed with t-tests, no significant differences were observed between Natural versus Man-made Window conditions ($p = .18$, Cohen's $d = 0.42$) or Natural versus Man-made Scotoma conditions ($p = .29$, Cohen's $d = 0.34$) (see Table 2 for descriptive statistics). Because these non-significant differences were also relatively small in terms of their effect sizes, we

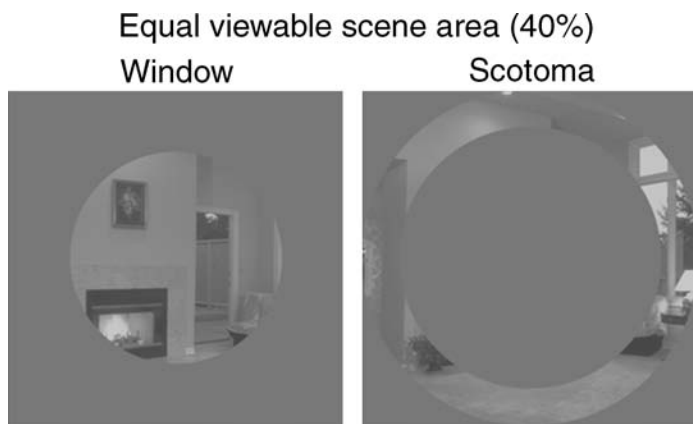## Equal viewable scene area (40%)

Figure 6. Examples of *Window* and *Scotoma* versions of an image, with both having an identical percent of viewable area (40%), used in the Control Experiment.

collapsed across the Natural and Man-made scene categories in our critical comparison between the 40% Window and 40% Scotoma conditions.

The results showed that when the centrally and peripherally presented information had equal area, performance was significantly better for centrally presented information (Window: $M = 0.89$, $SD = 0.06$; Scotoma: $M = 0.86$, $SD = 0.07$), $t(41) = 2.89$, $p = .006$, Cohen's $d = 0.42$. This replicated and extended our previous results in the 10.8° radius conditions. Again, results were inconsistent with Hypothesis 1, that "more is better," since equal area did not produce equal performance. However, the results were consistent with both Hypothesis 2, that the peripheral advantage is only due to having more area, and Hypothesis 3 that centrally presented information is privileged.

Before moving on, our above discussion of scene gist recognition in terms of viewable area in central and peripheral vision suggests an alternative approach to graphing the data presented in Figure 5. Specifically, we can graph scene gist accuracy as a function of the viewable area in each Window and Scotoma size condition.[5] This is shown in Figure 7. The Window condition presents the smallest Window radius (1° Window = 0.4% viewable area) at the far left, while each subsequent data point is the next larger radius Window (i.e., 5°, 10.8°, and 13.6°; which show 10.6%, 50%, and 78.5% of the viewable area, respectively). Conversely, the Scotoma condition is plotted

from the largest to the smallest Scotoma, showing from less to more viewable area, where the 13.6° Scotoma (showing 21.5% of the viewable area) is the left-most data point, and each subsequent data point is the next smaller Scotoma radius showing more viewable area (i.e., 10.8°, 5°, and 1°; show 50%, 89.3%, and 99.5% of the viewable area, respectively). Figure 7 is especially helpful in evaluating Hypotheses 1 and 3. As with Figure 5, inspection of Figure 7 shows evidence generally consistent with Hypothesis 1 ("more area is better") in that as the viewable area increases in both Window and Scotoma conditions, gist recognition performance increases. However, Figure 7 also supports Hypothesis 3 ("central vision is privileged, and more efficient"), in that the addition of viewable pixels in the Window condition results in a more rapid increase in gist recognition, which is represented by the generally steeper slopes in the Window condition (e.g., between the 1° and 5° radii Windows) compared to the Scotoma condition. In addition, the curve for the Window condition is above that for the Scotoma condition at every point, and particularly so when the percentage of viewable area is relatively small (i.e., 50% of viewable area or less).

Thus far, our results suggest that central information is more efficient for scene gist recognition, because when compared with an equal amount of peripheral area, the central portion of a scene conveys more information on a per-pixel basis for gist recognition, as shown by the steeper curve in the Window condition, above that of the Scotoma condition, in Figure 7. The latter conclusion is generally consistent with the idea of
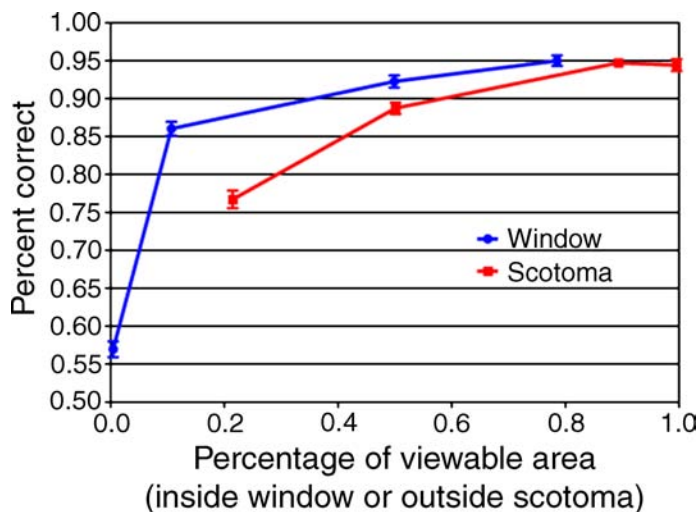

Figure 7. Scene gist accuracy plotted as a function of the percentage of viewable area presented comparing Window and Scotoma conditions for Experiment 1. Please note that the Scotoma condition proceeds from larger Scotoma radii (i.e., less viewable area) to smaller radii (i.e., more viewable area) from left to right. Note also that the smallest Window, 1° shows 0.4% of the viewable area, while the smallest Scotoma, 1°, shows 99.5% of the viewable area. Error bars = Standard Error of the Mean (*SEM*).

	Natural		Man-made	
Condition	Mean	St. Dev	Mean	St. Dev
Window	0.883	0.056	0.908	0.065
Scotoma	0.868	0.056	0.843	0.091

Table 2. Descriptive Statistics of Natural and Man-made scenes for Window and Scotoma images in the Control Experiment.

cortical magnification. One way to further test this idea is to more carefully investigate the crossing point of the performance curves for the Window and Scotoma conditions shown in Figure 5. As noted above, this crossing point suggests that there should be a particular radius of Window and Scotoma conditions that produces equal performance and that this should contain less area inside the Window than outside of the Scotoma. If so, one could compare the critical radius with predictions derived from cortical magnification functions in order to determine if this explains the results. Furthermore, we would want to use circular images to carefully control eccentricity in both Window and Scotoma conditions. In addition, in the current study, the natural/man-made distinction is a nuisance variable. Therefore, instead of splitting it across groups and then collapsing across those groups, we could simply present both natural and man-made scenes to all participants. These issues were explored in Experiment 2.

Experiment 2

The present study will estimate the *critical radius*—the radius that produces equal gist performance between Scotoma and Window scene images. Based on the results of Experiment 1, we predict that less centrally presented image content is needed than peripherally presented image content to achieve the same level of gist recognition performance. Such a result would be consistent with the idea of cortical magnification of central vision, because more cortical area is devoted to processing centrally located information than peripherally presented information. Thus, less image area would be needed in central vision than peripheral vision to achieve the same level of performance.

Once we have found the critical radius, we can compare that to the predicted critical radii based on published cortical magnification equations. For our purposes, the key idea of cortical magnification is that V1 devotes a greater proportion of cells to process foveal information than to peripheral visual information. Thus, to achieve equal performance in the periphery to that in central vision, a peripheral stimulus will need to be larger, such that *it stimulates the same number of V1 cells* as would be stimulated by a smaller stimulus in the fovea (Virsu et al., 1987; Virsu & Rovamo, 1979). Taking this idea, if V1 cortical magnification explains the empirically derived critical radius for gist recognition, we would predict that there should be *an equal area of V1 cortex* devoted to processing scene imagery at eccentricities less than and greater than the empirically derived critical radius.

One way to test the V1 cortical magnification hypothesis is to simply use areal cortical magnification functions to predict the critical radius in which half of the total area of V1 stimulated by our images is within the critical radius, and half is beyond the critical radius (with the outer bound being

the outermost radius of the Scotoma condition). We can then compare the predicted critical radii based on cortical magnification functions with our empirically derived critical radius. To the extent that the predicted critical radii differ from the empirically derived one, we can infer differences in the relative importance of V1 central versus peripheral vision for scene gist recognition. Here, a critical assumption is that the smaller the critical radius is, the more important central vision is, whereas the larger the critical radius is, the more important peripheral vision is. Thus, if the predicted and the empirically derived critical radii are essentially the same, V1 cortical magnification would explain the relative importance of central and peripheral vision for scene gist recognition. However, if the predicted critical radii are smaller than the empirical one, it would suggest that V1 cortical magnification functions over-estimate the relative importance of central vision for scene gist. Conversely, to the extent that the predicted critical radii are larger than the empirical one, it would suggest the opposite conclusion, namely that V1 cortical magnification functions under-estimate the importance of central vision for scene gist recognition.

Method

Participants

A total of 18 undergraduates (14 Females) from Kansas State University participated in the experiment for course credit. Participants had a mean age of 18.61 ($SD = 1.46$), and all had a minimum visual acuity of 20/30.

Materials

Similarly to the control experiment, we used circular images, as shown in Figure 8. The same scene categories were used as in Experiment 1. Four scene images per category were added to the original set of 32 category images. The Window and Scotoma radii were selected based on predicted values from pilot testing. The middle radius (7.64°) was predicted to produce equal gist accuracy in the Window and Scotoma conditions, and we then selected one larger radius (9.21°) and one smaller radius (6.03°), that were predicted to produce ±1% gist accuracy in the Scotoma condition, which had a shallower slope than that of the Window condition. The percentages of viewable area between Window and Scotoma scenes for each radius are 20%:80%, 32%:68%, and 46%:54%, respectively (see Figure 8). All scene categories and Scotoma/Window sizes were presented to each participant.

Procedures

Participants received the same familiarization and practice sessions as in Experiment 1. The experiment consisted of 360 trials (36 scene images per category) with
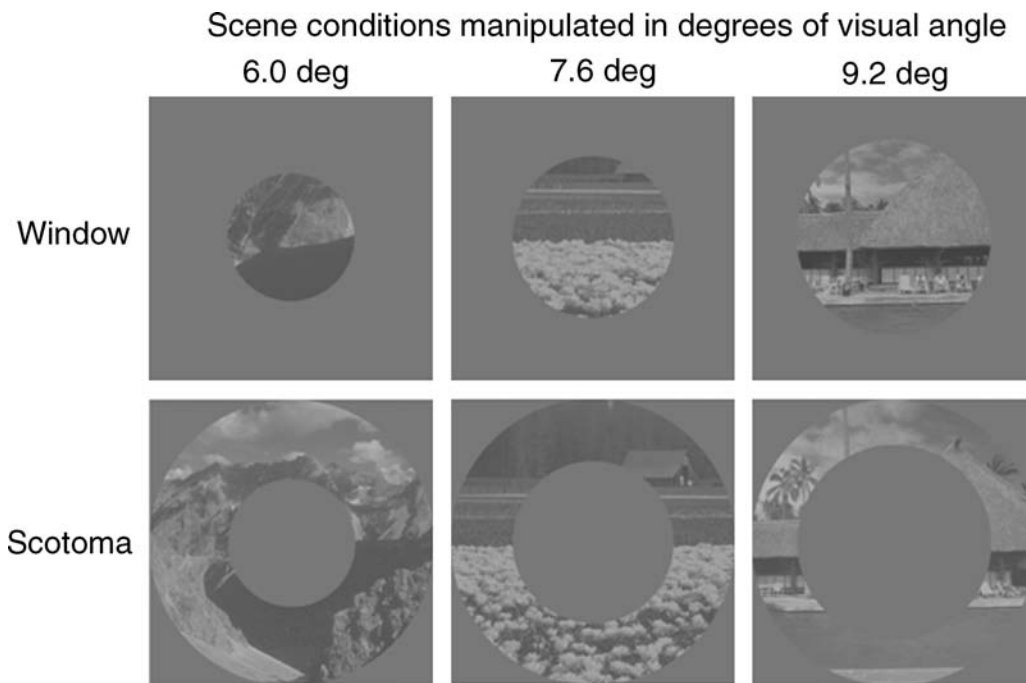
Figure 8. Examples of *Windows* and *Scotomas* of differing radii, in degrees of visual angle, used in Experiment 2.

each scene image appearing once. The radius sizes of Window and Scotoma images were equally likely to occur and were presented randomly throughout the 360 trials. The events that occurred in a trial were the same as in Experiment 1, except that the target was only presented for 24 ms (see Figure 4).

Results and discussion

In order to determine the critical radius that would produce equal gist performance between central and peripheral information, two linear equations, one for the Window condition and the other for the Scotoma condition, were used to predict gist accuracy at each pixel between the 6.0° and 7.6° radius conditions presented in Figure 9. The critical radius was then determined by calculating the crossing point for the two equations. As shown in Figure 9, the crossing point between Window and Scotoma conditions (the critical radius) is 7.40°.[6] The area within the critical radius (i.e., the Window condition) contains 29.83% of the total scene, with the remaining 70.17% of the scene being presented beyond the critical radius (i.e., the Scotoma condition). Therefore, centrally presented image content provides more information for recognizing the gist of a scene, on a per-pixel basis, than peripherally presented content.

In order to test this explanation in terms of cortical magnification, we compared the empirical critical radius we found with the predicted critical radius of two different areal cortical magnification functions from the literature (Florack, 2007; Van Essen, Newsome, & Maunsell, 1984).

We used the following equation from Florack's geometric model of the fovea (2007, p. 922, Equation 10),

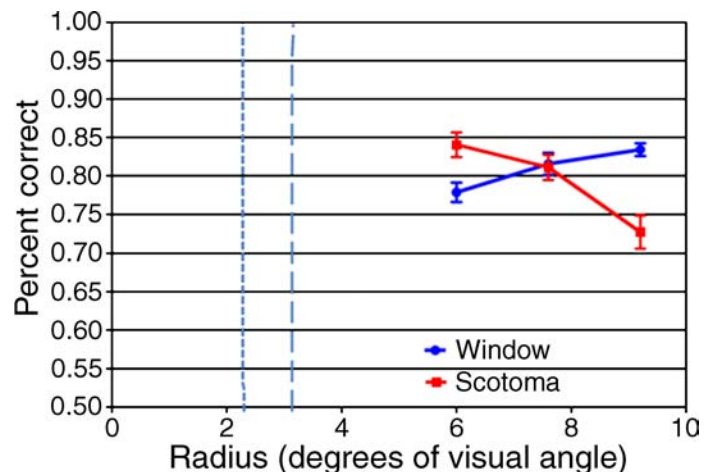$$v(t, T) = \ln(1 + t^2)/\ln(1 + T^2), \tag{1}$$

Figure 9. Scene gist accuracy between Window and Scotoma radii from 6° to 9.2°, showing a crossing at 7.4°, representing the critical radius producing equal gist performance in Experiment 2. The dotted vertical line represents the predicted critical radius for the cortical magnification function of Van Essen et al. (1984), and the dashed vertical line represents the same for Florack (2007). These predictions rest on the assumption of equal performance in Window and Scotoma conditions, if each activates equal cortical area in V1. Error bars = Standard Error of the Mean (*SEM*).

in which v, the integrated retino-cortical magnification, which ranges from 0 to 1, represents the relative area within V1 subsumed by a circular region, centered at the fovea, of a given radius. The parameter $t = \rho/a$, with ρ being the radius of a disk centered at the fovea, and a being the outer edge of the geometric foveola, or 0.22 mm (Florack, 2007, p. 923). We can convert degrees of visual angle to mm based on the fact that 1° is roughly 288 μm on a normal retina (Florack, 2007, p. 923). The parameter $T = R/a$, with R representing the maximum radius of the retina, or roughly 21 mm (Florack, 2007, p. 923). Given the above, Florack's (2007) equation becomes:

$$v(t, T) = \ln\left(1 + [\{r^*288/1000\}/0.22]^2\right) \qquad (2)$$
$$/\ln\left(1 + [21/0.22]^2\right).$$

Using Florack's (2007) equation we calculated the proportion of area of V1 that would be processing the entire image, that is, the outer radius in the Scotoma condition (13.55°), which was 0.631. On the assumption that equal gist performance results from equal areas in V1, half that area of V1 (0.316) would be predicted to be within the critical radius and the other half would be beyond it. We then calculated the critical radius that would contain 0.316 of the area of V1 (i.e., half of the total area of V1 devoted to the entire image), which was 3.13° radius.

We did a similar analysis based on Van Essen et al.'s (1984) areal cortical magnification function derived from physiological recordings from the macaque (1984, p. 438, Equation 4),

$$M_a = a(b + E)^{-x}$$
$$= 103(0.82 + E)^{-2.28} \text{mm}^2/\text{deg}^2, \qquad (3)$$

in which M_a is the areal cortical magnification, a and b are parameters fit to their data, and E is retinal eccentricity. To do this, we found the integral from 0° to 13.55° to predict the total area of V1 in mm²/deg² of the entire image, which was 183.03 mm²/deg². We then found the critical radius whose integral contained half that area (91.51 mm²/deg²), which was 2.38° radius.

Interestingly, the predicted critical radii from Florack (2007) and Van Essen et al. (1984) were relatively similar and both were considerably smaller than the empirical critical radius. To test for the significance of these differences, we carried out two one-sample t-tests between the empirical critical radius and both predicted critical radii. For this, we found the critical radius for each of 14 subjects (4 of 18 subjects were excluded for having multiple crossings or out of bounds crossings) and compared the mean of those critical radii with the predicted critical radii. The critical radius ($M = 7.48°$, $SD = 0.68$) was significantly and meaningfully larger than that

predicted by either V1 cortical magnification function ($ts(13) > 23.94$, $ps < .001$, Cohen's $ds > 6.64$). Thus, our empirical critical radius suggests that central vision contributes less to gist recognition than predicted by cortical magnification functions for V1. In summary, although centrally presented information is more valuable for scene gist recognition on a per-pixel basis than peripherally presented information, the degree of cortical magnification is considerably less than that predicted for V1. One possibility is that the weaker cortical magnification shown in our study occurs at a higher level of cortical processing where peripheral vision plays a relatively greater role (e.g., the Parahippocampal Place Area (Epstein, 2005)).

Conclusions

Scene gist recognition is a critically important early stage of scene perception, influencing more complex cognitive processes such as directing our attention within a scene, facilitating object recognition, and influencing long term memory. However, until now the relative contributions of central versus peripheral vision in this process have been unclear. The current study shows that peripheral vision (i.e., vision beyond 5° eccentricity) is more useful for recognizing the gist of a scene than central vision (i.e., foveal + parafoveal vision). Our study also shows that the advantage for peripheral vision is largely explained by the fact that it covers a wider area of the visual field. Interestingly, this obscures the fact that, as we have also shown, central vision is more efficient at processing gist. This is shown by the fact that on a per-pixel basis, less information is needed in central vision to attain the same level of gist recognition performance as in peripheral vision. This result is consistent with the general idea of cortical magnification of central vision. Specifically, many more cells in area V1 are devoted to central vision than to the periphery, so a relatively small central image and a larger peripheral image would potentially activate the same amount of cortical area in V1. However, our results showed that cortical magnification factors derived from the topography of cortical area V1 overestimated the degree of cortical magnification, and therefore the importance of central vision, for scene gist recognition. We speculate that the reason for this may be that the critical brain area involved in scene gist recognition is at a higher cortical level than V1, such as the parahippocampal place area (PPA) (Epstein, 2005), in which cortical magnification of the fovea is relatively attenuated compared to area V1 (Hasson, Levy, Behrmann, Hendler, & Malach, 2002).

Indeed, the current results suggest the following hypothesis: the equal performance produced by unequal areas in the Window and Scotoma conditions is due to

equal activity aroused by Window and Scotoma stimuli within a higher visual region specific to scene processing, such as the PPA, but unequal activity in area V1. This hypothesis would be testable through a brain imaging (e.g., fMRI) study. Specifically, given Window and Scotoma stimuli both sharing a critical radius that produces equal scene gist recognition, one may ask what the ratio of V1 activation would be for the central versus peripheral regions. Using the cortical magnification functions described earlier, the Florack (2007) equations predict that our empirical critical radius of 7.4° would have 3.78 times more area of V1 within the critical radius than beyond it, and the Van Essen et al. (1984) equations predict an even larger ratio of 5.65 times more area within than beyond the critical radius. Thus, we would predict that activation in area V1 should be several times greater for the Window stimuli than the Scotoma stimuli, showing the predicted degree of cortical magnification. Conversely, the area of activation in higher cortical areas specific to scene processing, such as the PPA, should have a ratio of approximately unity for both the Window and Scotoma stimuli, showing an attenuated degree of cortical magnification, due to a somewhat greater emphasis on processing stimuli in the visual periphery (Hasson et al., 2002).

The fact that our study has shown the importance of peripheral vision for scene gist recognition is consistent with the conclusions of other recent studies (Rousselet, Thorpe, & Fabre-Thorpe, 2004; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bulthoff, 2001). However, in those studies, entire scene images were presented extrafoveally. Thus, a contribution of the current study is that it directly investigated the relative contributions of both central and peripheral vision in recognizing the gist of images that extend from the fovea into the periphery. An advantage of this approach is that it simulates the normal situation in scene perception—in normal vision a scene does extend from the fovea into the periphery. Conversely, a limitation of the current approach is that it involves blocking out information from a given portion of the field of view, which is only "normal" for viewers with visual defects (e.g., Scotomas or retinitis pigmentosa). Nevertheless, participants in the current study did not know from trial to trial whether they would see a Window, Scotoma, or normal image, and so could not prepare for one or the other. Thus, their performance should have shown the independent values of each portion of the visual field for scene gist recognition.

Because the current results suggest that peripheral vision plays a key role in recognizing scene gist, they raise the question of what role attention plays in gist recognition. If one assumes that attention is usually not focused on the visual periphery, then one might argue on this basis that the current results are consistent with the claim of several recent studies that scene gist recognition does not require focal attention (Fei-Fei, VanRullen, Koch, & Perona, 2005; Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002; Rousselet et al., 2004;

Thorpe et al., 2001). However, given that the current study did not manipulate attention or the time course of processing, it is silent in this regard. Follow-up studies in our laboratory are investigating this issue.

An important debate in the scene gist recognition literature is centered on whether object recognition plays an important role in gist recognition. Specifically, it has been argued that recognizing one or more "obligatory" diagnostic objects may enable one to automatically categorize the entire scene; for example recognizing a "football player" could automatically activate the mental category of "stadium" (Bar & Ullman, 1996; Davenport & Potter, 2004; Friedman, 1979). Conversely, others have argued that rather than depending on recognizing objects, scene gist recognition may instead depend on processing of low spatial frequency layout, texture, or other low-level visual features (Oliva & Torralba, 2001; Renninger & Malik, 2004; Schyns & Oliva, 1994). Interestingly, to the extent that diagnostic objects play an important role in scene gist recognition, one would also expect that central vision would too, given the importance of central vision for object recognition (Henderson & Hollingworth, 1999; Hollingworth et al., 2001; Nelson & Loftus, 1980; O'Regan et al., 2000; Pringle, 2000; c.f., Thorpe et al., 2001). However, the current results may be partially explained by the relative presence of diagnostic features in central versus peripheral vision. For example, eye movements tend to cluster in the center of images (Mannan, Ruddock, & Wooding, 1995; Parkhurst & Neibur, 2003) and it has been argued that this may be due to photographers pointing their cameras at interesting objects, which end up in the center of pictures (though see Tatler, 2007). Based on this logic, there may be more diagnostic features for scene gist in central vision than peripheral vision. However, we have shown that blocking out the central 5° of a scene nevertheless results in asymptotic gist recognition performance. Therefore, if we still wanted to try to explain our results in terms of the distribution of diagnostic features, we would need to argue that the periphery contains more diagnostic features, perhaps because it contains a large proportion of the visual field (in Figure 7, 89.3% of the viewable area). By this logic, the 5° Window may not have been able to achieve asymptotic performance because it contained too little of the visual field (in Figure 7, 10.7% of the viewable area), thus having a lower probability of reaching a threshold amount of diagnostic features. In sum, our results may not be completely due to the cortical specialization of processing in central versus peripheral vision, but may also be influenced by the presence or absence of diagnostic features in differing regions of the visual field. The same factor may have influenced the size of the critical radius estimated in Experiment 2. This suggests an interesting hypothesis to test, namely is the equivalent performance observed with the critical radius due to an equivalent amount of diagnostic information in both central and peripheral vision? Further studies are needed

to examine the influence of diagnostic features at differing retinal eccentricities to evaluate their impact on scene gist recognition.

In summary, we have found that for fairly large images (e.g., extending out to roughly 14° eccentricity), peripheral vision is all that is needed for recognizing the gist of a scene. This seems to be due to the fact that the periphery contains so much more area from which lower spatial frequency information, which is useful for recognizing gist, can be acquired. However, this does not imply that central vision is of no value to recognizing the gist of a scene. On the contrary, information presented to central vision is privileged and more efficient, producing greater gist recognition than peripheral vision on a per-pixel basis. While this finding is generally consistent with the idea of cortical magnification, it seems that the value of centrally presented information is considerably less than that predicted by cortical magnification functions for V1. The reason for this relatively attenuated bias towards central vision may be because the critical information for recognizing scene gist is processed at a higher cortical area, for example the PPA.

Acknowledgments

Commercial relationships: none.
Corresponding author: Lester Loschky.
Email: loschky@ksu.edu.
Address: Department of Psychology, Kansas State University, 471 Bluemont Hall, Manhattan, KS 66506-5302, USA.

Footnotes

[1]The term "scene gist" is most commonly operationalized as we have here, in terms of applying a category label (Oliva & Schyns, 2000; Rousselet, Joubert, & Fabre-Thorpe, 2005; Schyns & Oliva, 1994; Thorpe et al., 2001).

However, others have conceptualized "scene gist" more broadly to include longer descriptions of a scene, for example, "dog with a frisbee" or "man in a boat" (Fei-Fei et al., 2007; Potter, 1976).

[2]Beyond the parafovea, retinal anatomists distinguish between the perifovea, which extends to the outer boundary of the macula, at approximately 10° eccentricity, and the periphery, which includes the entire visual field beyond that point. However, this latter distinction is not widely used among vision scientists.

[3]Other studies suggest that the effect of retinal eccentricity on object recognition in scenes interacts with both object size and task complexity. Biederman, Mezzanotte, Rabinowitz, Francolini, and Plude (1981) found that object recognition in briefly flashed scenes required foveal or parafoveal vision (0–2° eccentricity) for small objects (1.5° wide) but that peripheral vision (6–8° eccentricity) was useful for recognizing larger objects (3° wide). Thorpe et al. (2001) showed that for very large images (39° × 26°) the simpler task of detecting a single central animal produced above-chance performance at up to 75° eccentricity. Thus, foveal vision seems important for recognizing smaller objects and making finer distinctions, while peripheral vision seems useful for recognizing larger objects and making grosser object categorizations.

[4]It should be noted, however, that Oliva and Schyns (1997) have shown that the particular spatial frequency band most useful in scene categorization can be made to vary with task demands. Specifically, if subjects are trained to use higher frequency information to recognize gist, they will favor those frequencies in categorizing scenes, which would therefore favor central vision.

[5]We would like to thank Tom Sanocki for suggesting this figure.

[6]Follow-up studies using this critical radius have shown that it does indeed produce equal performance in the Window and Scotoma conditions.

References

Bacon-Mace, N., Mace, 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. [PubMed] [Article]

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception, 25,* 343–352. [PubMed]

Beard, B. L., & Ahumada, A. J., Jr. (1999). Detection in fixed and random noise in foveal and parafoveal vision explained by template learning. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 16,* 755–763. [PubMed]

Biederman, I., Mezzanotte, R. J., Rabinowitz, J. C., Francolini, C. M., & Plude, D. (1981). Detecting the

unexpected in photointerpretation. *Human Factors, 23,* 153–164.

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. [PubMed]

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13,* 1207–1230.

Coletta, N. J., & Williams, D. R. (1987). Psychophysical estimate of extrafoveal cone spacing. *Journal of the Optical Society of America A, Optics and Image Science, 4,* 1503–1513. [PubMed]

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

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. [PubMed]

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

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision, 7*(1):10, 1–29, http://journalofvision.org/7/1/10/, doi:10.1167/7.1.10. [PubMed] [Article]

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.

Florack, L. M. J. (2007). Modeling foveal vision. In F. Sgallari, A. Murli, & N. Paragios (Eds.), *First International Conference on Scale Space and Variational Methods in Computer Vision (SSVM 2007)* (vol. 4485, pp. 919–928). Berlin: Springer-Verlag.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316–355. [PubMed]

Geisler, W. S., & Perry, J. S. (1999). Variable-resolution displays for visual communication and simulation. In *The society for information display* (vol. 30, pp. 420–423).

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 760–777. [PubMed]

Harvey, L. O., Roberts, J. O., & Gervais, M. J. (1983). The spatial frequency basis of internal representation. In H. G. Geissler, H. F. J. M. Buffart, E. L. J. Leeuvenberg, & V. Sarris (Eds.), *Modern issues in perception* (pp. 217–226). Berlin: VEB Deutscher Verlag der Wissenschaften.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron, 34,* 479–490. [PubMed] [Article]

Henderson, J. M., & Hollingworth, A. (1999). The role of fixation position in detecting scene changes across saccades. *Psychological Science, 10,* 438–443.

Henderson, J. M., McClure, K. K., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception & Psychophysics, 59,* 323–346. [PubMed]

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127,* 398–415. [PubMed]

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. [PubMed] [Article]

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

Kortum, P. T., & Geisler, W. S. (1996). *Search performance in natural scenes: The role of peripheral vision.* Paper presented at the ARVO.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 9596–9601. [PubMed] [Article]

Loschky, L. C., & McConkie, G. W. (2002). Investigating spatial vision and dynamic attentional selection using a gaze-contingent multiresolutional display. *Journal of Experimental Psychology: Applied, 8,* 99–117. [PubMed]

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.

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. [PubMed]

Mannan, S., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-d images. *Spatial Vision, 9,* 363–386. [PubMed]

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

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

Nelson, W. W., & Loftus, G. R. (1980). The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 391–399. [PubMed]

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. [PubMed]

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210. [PubMed]

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

O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition, 7,* 191–211.

Parkhurst, D., Culurciello, E., & Neibur, E. (2000). Evaluating variable resolution displays with visual search: Task performance and eye movements. In A. T. Duchowski (Ed.), *Proceedings of the eye tracking research & applications symposium 2000* (pp. 105–109). Palm Beach, FL: ACM.

Parkhurst, D. J., & Niebur, E. (2002). Variable-resolution displays: A theoretical, practical, and behavioral evaluation. *Human Factors, 44,* 611–629. [PubMed]

Parkhurst, D. J., & Neibur, E. (2003). Scene content selected by active vision. *Spatial Vision, 16,* 125–154. [PubMed]

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.

Polyak, S. L. (1941). *The retina*. Chicago: University of Chicago Press.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 509–522. [PubMed]

Pringle, H. L. (2000). *The roles of scene characteristics, memory and attentional breadth on the representation of complex real-world scenes*. Unpublished Doctoral, University of Illinois at Urbana-Champaign, Urbana.

Rayner, K., & Bertera, J. H. (1979). Reading without a fovea. *Science, 206,* 468–469. [PubMed]

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. [PubMed]

Reingold, E. M., Loschky, L. C., McConkie, G. W., & Stampe, D. M. (2003). Gaze-contingent multiresolutional displays: An integrative review. *Human Factors, 45,* 307–328. [PubMed]

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

Rodieck, R. W. (1998). *The first steps in seeing*. Sunderland, MA: Sinauer Associates, Inc.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630. [PubMed]

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.

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. [PubMed]

Saida, S., & Ikeda, M. (1979). Useful visual field size for pattern perception. *Perception & Psychophysics, 25,* 119–125. [PubMed]

Sanocki, T. (2003). Representation and perception of scenic layout. *Cognitive Psychology, 47,* 43–86. [PubMed]

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science, 8,* 374–378.

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

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):10, 1–20, http://journalofvision.org/7/12/10/, doi:10.1167/7.12.10. [PubMed] [Article]

Shioiri, S., & Ikeda, M. (1989). Useful resolution for picture perception as a function of eccentricity. *Perception, 18,* 347–361. [PubMed]

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17, http://journalofvision.org/7/14/4/, doi:10.1167/7.14.4. [PubMed] [Article]

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. [PubMed]

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. [PubMed]

van Diepen, P., De Graef, P., Lamote, C., & van Wijnendaele, I. (1994). The role of central and peripheral image cues in scene recognition, *the Seventeenth European Conference on Visual Perception*. Eindhoven, The Netherlands.

van Diepen, P. M. J., Ruelens, L., & d'Ydewalle, G. (1999). Brief foveal masking during scene perception. *Acta Psychologica, 101,* 91–103. [PubMed]

van Diepen, P. M. J., & Wampers, M. (1998). Scene exploration with Fourier-filtered peripheral information. *Perception, 27,* 1141–1151. [PubMed]

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood et al. (Ed.), *Eye guidance in reading and scene perception* (pp. 337–355). Oxford, England UK: Anonima Romana.

Van Essen, D. C., Newsome, W. T., & Maunsell, H. R. (1984). The visual field representation in striate cortex of the macaque monkey: Asymmetries, anisotropies, and individual variability. *Vision Research, 24,* 429–448. [PubMed]

Virsu, V., Näsänen, R., & Osmoviita, K. (1987). Cortical magnification and peripheral vision. *Journal of the Optical Society of America A, Optics and Image Science, 4,* 1568–1578. [PubMed]

Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research, 37,* 475–494. [PubMed]

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 et al. (Eds.), *Visual perception: The neurophysiological foundations* (pp. 231–272). San Diego, CA, USA: Academic Press, Inc.

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