Capabilities and Limitations of Peripheral Vision

Ruth Rosenholtz

Department of Brain and Cognitive Sciences, CSAIL, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; email: rruth@mit.edu

Annu. Rev. Vis. Sci. 2016. 2:437–57

The *Annual Review of Vision Science* is online at vision.annualreviews.org

This article's doi:
10.1146/annurev-vision-082114-035733

Keywords

encoding, crowding, attention, acuity

Abstract

This review discusses several pervasive myths about peripheral vision, as well as what is actually true: Peripheral vision underlies a broad range of visual tasks, in spite of its significant loss of information. New understanding of peripheral vision, including likely mechanisms, has deep implications for our understanding of vision. From peripheral recognition to visual search, from change blindness to getting the gist of a scene, a lossy but relatively fixed peripheral encoding may determine the difficulty of many tasks. This finding suggests that the visual system may be more stable, and less dynamically changing as a function of attention, than previously assumed.

INTRODUCTION

The rod-free fovea (sometimes referred to as the foveola but here simply called the fovea) covers approximately the central 1.7° of the visual field. Vision outside the fovea, which this review will colloquially refer to as peripheral vision, covers as much as 99.9% of the visual field. To understand our visual capabilities and limitations, we must understand what we can see at a glance. To understand vision at a glance, we must understand peripheral vision.

Researchers have learned much about peripheral vision. However, one can find a number of misconceptions within the broader field of vision science; within related fields of computer graphics, human–computer interaction, and human factors; and in popular science accounts in the media. In particular, there exist two common and essentially conflicting accounts. In the first, peripheral vision is impoverished and useful for very little; the fovea is where the interesting action occurs. In the second view, peripheral vision is merely foveal vision with slightly lower resolution—often incorrectly described as "slightly blurry."

The fact is that we rely on peripheral vision for much of visual processing, but it is degraded relative to foveal vision. The degradation, however, mostly involves not the loss of resolution in the traditional sense (e.g., number of photoreceptors per degree of visual angle), but rather peripheral vision's particular vulnerability to clutter. Recent progress in understanding this vulnerability, and what it tells us about the underlying mechanisms, has important implications for understanding a wide array of visual phenomena and vision more generally. In fact, understanding the strengths and limitations of peripheral vision is crucial for answering one of the most fundamental questions of vision science: To what extent should we think of vision as having a static, stable, underlying encoding that is usable for many tasks, as opposed to being dynamic and ever-changing to adapt to the task at hand? This review attempts to clear up the common misconceptions and to illuminate a more recent understanding of peripheral vision. It frames the big-picture question of a static versus dynamic encoding in terms of peripheral vision and discusses recent progress toward answering that question.

View 1: Peripheral Vision Is Impoverished and All but Useless

> *For a long time the prejudice was prevailing that indirect, compared to direct vision, is imperfect and irrelevant, and only very slowly the insight of the fundamental importance of seeing sidelong has prevailed.* —Korte (1923)

In the first popular view, vision outside the fovea is severely impoverished. Common wisdom holds that peripheral vision has poor acuity (the ability to resolve fine details) and poor color perception. Because of the lack of acuity, according to this view, one cannot imagine using peripheral vision for much of anything. To perform tasks such as object recognition, one first must point one's "high-resolution fovea" at the informative parts of the scene.

Both acuity and color perception are worse in the periphery than in the fovea, but the effect is not as extreme as is generally assumed. One can find many unintentionally misleading (if perhaps factually correct) demonstrations of the lack of acuity, as well as flat-out exaggerations of the poverty of peripheral vision. Peripheral vision provides us with a great deal of useful information, and if we actually had to rely on multiple foveal views, our vision would be far worse. In addition, many visual tasks require neither high acuity nor eye movements, belying the need to point one's fovea at the important bits in order to recognize objects.

Color vision does deteriorate in the periphery. Many readers have no doubt seen a depiction of how the density of cones—the cells in the retina responsible for color vision—drops off with eccentricity (i.e., the distance to the point of fixation). **Figure 1*a*** shows a simplified version of

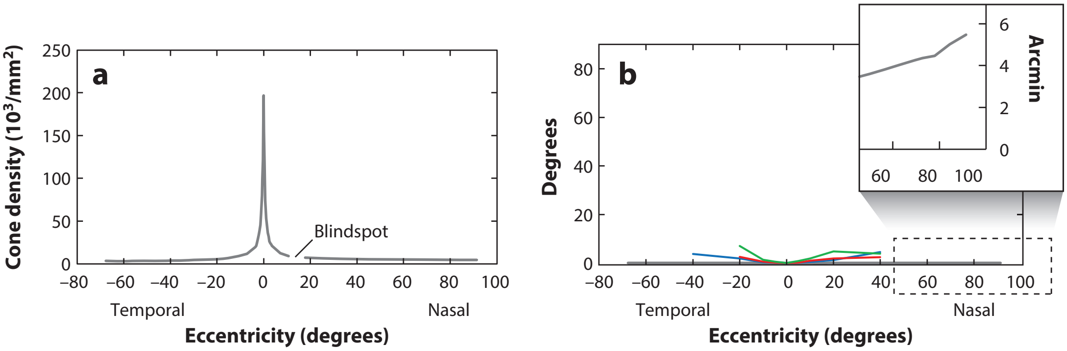

Figure 1

(*a*) Density of color-sensitive photoreceptors as a function of eccentricity. Plots modeled after Østerberg (1935) and data from Curcio et al. (1990). (*b*) Cone spacing estimated from cone density in panel *a*, as described in the text (*gray curve*). Inset shows detail to distinguish this curve from the *x*-axis. Red, green, and blue curves show size of stimuli necessary to induce similar color judgments to those made for a small foveal patch (Abramov et al. 1991). Figure adapted from Abramov et al. (1991).

this plot. The reduced density of color-sensitive photoreceptors leads to reduced ability to resolve high–spatial frequency color. However, plots such as **Figure 1*a***, though factually correct, can lead to misunderstandings. The *y*-axis is cone density, in thousands of cones per square millimeter. It is clear that the density drops off rather sharply outside the fovea. However, in elucidating human capabilities, we are interested in the size of the details one can just resolve, which depends not upon cone density, but rather upon the distance between neighboring cones. Because **Figure 1*a*** plots the square of the more relevant cones per millimeter, the apparent decline with eccentricity is exaggerated. This, in turn, makes it difficult to represent both the peak and the trough of the graph with high fidelity. In many such plots, the cone density erroneously appears to go to zero at an eccentricity of approximately 20°. But in fact, the curve asymptotes at approximately 4,000 cones per mm^2—still a sizeable number, if much smaller than in the fovea.

We would really like to know the spacing between neighboring cones—in other words, millimeters per cone, or $1/\sqrt{\text{cone density}}$. **Figure 1*b*** shows cone spacing in degrees of visual angle (millimeters to degrees conversion from Watson 2014). The spacing increases roughly linearly with eccentricity over a wide range. This plot appears far less dire than **Figure 1*a***.

The full story for peripheral color vision is more complicated, involving differences among the three cone types and how many cones connect to each ganglion cell, but the above picture, in broad strokes, gets us in the ballpark. Practically speaking, how bad is color perception in the periphery? One can find many popular science claims that the periphery has virtually no color vision, and that we merely “fill in” with colors we know either from a previous glance or from prior knowledge (“The trees in my periphery must be green”). Clearly, this is not the case. First, 4,000 cones/mm^2 is a significant density. Second, researchers studying peripheral color vision carefully control for fixation and prior knowledge, so we know that human vision is not just filling in known colors. In fact, humans are quite reasonable at peripheral color judgments, so long as the patches are sufficiently large [see Hansen et al. (2009) for a review, and **Figure 1*b*** for example data].

One can also ask about peripheral acuity more generally. Again, degradation with eccentricity has been misunderstood to be greater than it actually is. A common statement one finds is that we cannot read without the high-resolution fovea, and as a result, we must make eye movements. Even this modest-sounding statement is fundamentally incorrect, although certainly reading does

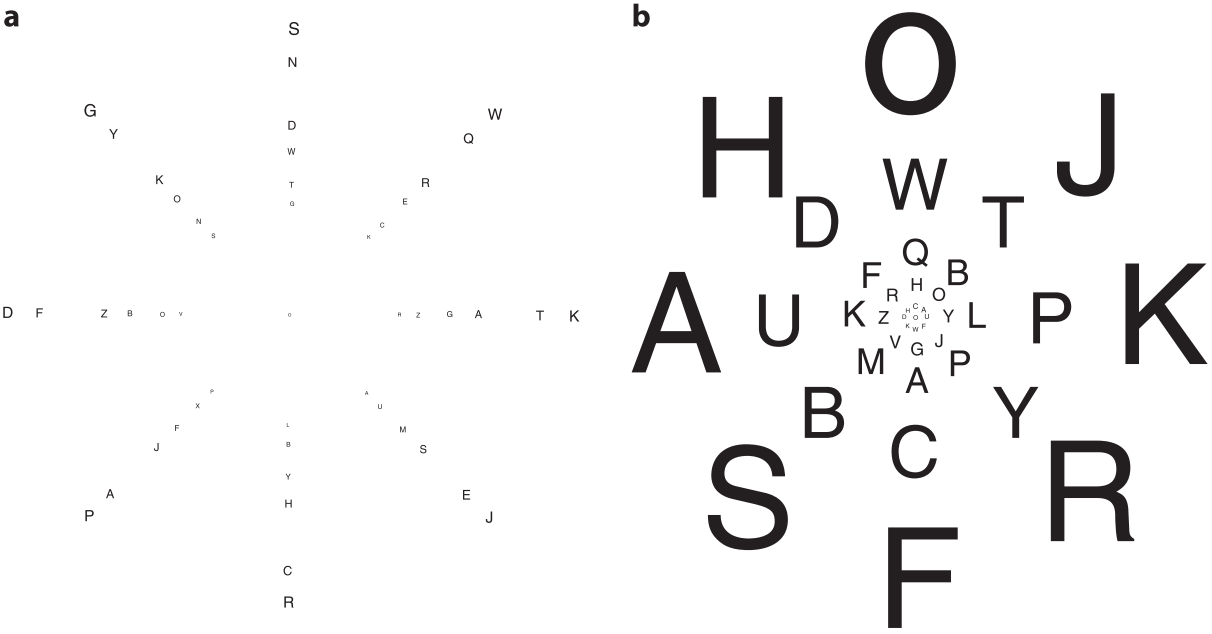

Figure 2

(*a*) Letter size at threshold for identification, as a function of eccentricity. (*b*) Ten times threshold size. Figure adapted from Anstis (1974) with permission.

require quite a bit of acuity. Here again, the misunderstanding may arise from confusion over a visualization that, although correct, has misled the casual observer. Anstis (1974) measured the threshold letter size necessary for recognition as a function of eccentricity. He plotted two useful visualizations. The first, reproduced in **Figure 2*a***, shows the threshold letter size at each eccentricity. One can see that this threshold size is quite small, even at high eccentricities. If acuity were the only concern, one would be able to read a page of text in a reasonable font size, and at normal reading distance, without moving one's eyes. (Of course we cannot, actually. As discussed later, acuity is not the main issue.)

From where, then, derives the idea that acuity drops off much more rapidly? **Figure 2*a*** is a bit unsatisfying. Because it shows letter size at threshold (i.e., at the minimum size for recognition), none of the letters seem easily readable when one fixates the center. Anstis (1974) also plotted, at each eccentricity, letters at 10 times the threshold size (**Figure 2*b***). This makes a much more satisfying demo. All the letters seem to have about the same degree of readability. However, **Figure 2*b*** makes the falloff of peripheral vision seem much more dramatic. But of course, that is because it exaggerates that falloff by a factor of 10! As with cone spacing, threshold letter size increases roughly linearly with eccentricity. The slope of this linear function is quite modest. If we plot 10 times threshold as a function of eccentricity, we get a curve with 10 times the slope. Unfortunately, this subtle point requires some thought. To make matters worse, the figure showing 10 times the threshold often appears with no reference to the scaling; for example, Johnson (2010) captioned this figure with merely, "the resolution of our visual field is high in the center, but much lower at the edges."

This demonstration likely brings to mind for some readers the notions of the cortical magnification factor and *M* scaling. For a foveated visual system, the area of the brain that is dedicated to processing 1° of visual angle is less in the periphery than in the fovea. The cortical magnification factor, M, represents this change in area as a function of eccentricity. The reciprocal often proves more useful. One can use this function to derive the factor by which we need to scale a stimulus in the periphery so that it is processed by approximately the same amount of cortex as the original foveal stimulus for a given anatomical region of visual processing. For primary visual area, V1, one approximation yields the recommendation that one scale stimuli at eccentricity E (in degrees of visual angle) by a factor of $M^{-1}(E)/M^{-1}(0) = 0.27E + 1$, relative to the size of the foveal stimuli (Horton & Hoyt 1991; see, for example, Rousselet et al. 2005a). Experimentalists sometimes scale their stimuli in this way to control for the cortical magnification factor when comparing perception at different eccentricities. Plugging in to the above equation, we see that one would need to "*M* scale" a stimulus at 10° eccentricity by a factor of approximately four compared with a foveal stimulus. This *M* scaling might again make peripheral vision sound quite bad—four times worse and only 10° out. But how bad is it? If one starts with a large foveal stimulus (as in **Figure 2*b***), then the *M*-scaled version at 10° looks huge, suggesting highly impoverished peripheral vision. If one started with a stimulus near threshold, then *M* scaling seems to imply that peripheral vision is not so bad (as in **Figure 2*a***). Acuity in the fovea is excessively high—higher than we probably need for virtually all tasks. We can resolve extremely small details. At 10° eccentricity, we can only resolve details four times that size. However, that means that the details we can resolve remain very small.

Other tasks fall off at different rates, but the overall story remains the same. The *M*-scaling equations above give a reasonable approximation of the falloff of grating acuity with eccentricity. Vernier acuity falls off three to four times faster, but it also starts off higher in the fovea; Vernier acuity is a hyperacuity (Levi et al. 1985). Detecting unreferenced motion starts out relatively difficult even in the fovea (it is difficult to detect a motion in the absence of a static reference to compare against) and then hardly falls off at all by 10° eccentricity (Levi et al. 1984).

Anstis demonstrated a clever trick for visualizing the reduced acuity in peripheral vision. Anstis (1998) applied Photoshop's radial blur with parameters of spin = 1 and zoom = 4, producing an eccentricity-dependent blur so as to mimic the loss of high–spatial frequency information in the periphery. One should note that this is merely a visualization. A loss of acuity should not necessarily lead to a percept of a blurry scene. The high spatial frequencies are lost, but our visual system should not infer that there are no high spatial frequencies in the scene, and thus, we should not necessarily perceive blur. As Anstis (1998) pointed out, we lack the acuity to throw away our microscopes, and yet this lack does not make our foveal vision appear blurry. **Figure 3*b*** shows the result of this process applied to the photograph in **Figure 3*a***. The reader will notice that the blur is quite modest. It is actually exaggerated. By using this procedure to blur both

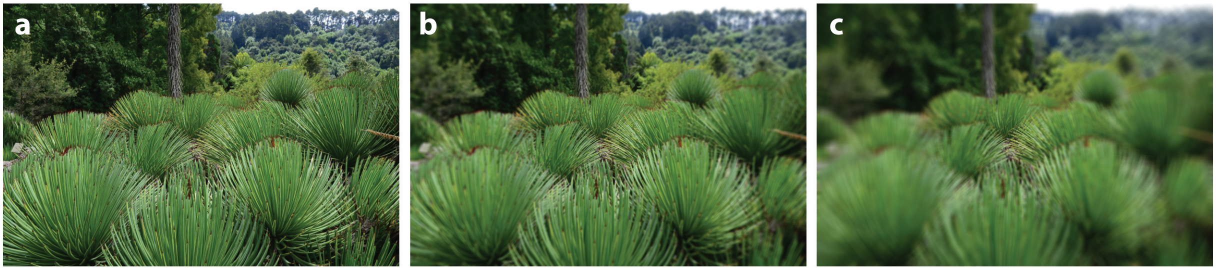

Figure 3

(*a*) Original image (photograph by Benjamin Wolfe). (*b*) Blurred using Anstis's Photoshop technique for mimicking loss of peripheral acuity (spin = 1, zoom = 4). Note that this already exaggerates the loss of acuity. The more typical demonstration looks more like that in panel *c*, which exaggerates the loss even further (spin = 3, zoom = 12).

at-threshold letter acuity stimuli and line-pair acuity stimuli, I estimate that the blur is approximately four times that needed to mimic peripheral loss of acuity. Anstis may have exaggerated the blur "for purposes of exposition" or perhaps because Photoshop cannot perform a spin of less than 1. In practice, imitations of Anstis's demonstration often exaggerate the blur even more, as in **Figure 3*c***. After all, one can then better see the blur. Unfortunately, such demonstrations often occur with no reference to the exaggeration. Instead, the caption may read something like, "When fixated at their respective centers, both pictures look equally sharp because the progressive blurring in the right hand picture just matches the progressive loss of acuity with eccentricity caused by the increasingly coarse grain of the peripheral retina" (Peripheral Acuity 2012). But in fact, when fixating the center, the two images do not look equally sharp. Having perhaps already been told not to trust his peripheral percept of color, the viewer also ceases to trust his peripheral judgments of image sharpness.

In addition to misunderstandings about peripheral color vision and acuity, behavioral evidence contradicts the idea that perception requires moving the eyes to bring the high-resolution fovea to bear on informative parts of the scene. Unlike the gratings often used to measure acuity, most of our visual world is broadband (with an amplitude spectrum that falls off as approximately $1/f$, in which f is spatial frequency), meaning that it contains energy, and likely also useful information, at a wide range of frequencies. As a consequence, we can recognize objects and scenes at resolutions far below that available in peripheral vision. Many readers can demonstrate this for themselves simply by removing their glasses. Alternatively, one can zoom in on a low-resolution thumbnail of a scene. The image, in most cases, proves quite recognizable. Torralba (2009) has studied this more formally. He downsampled scenes to 32 × 32 pixels and showed that observers could nonetheless recognize both scene category and objects within the scene with high accuracy.

In fact, little recognition truly involves only the fovea. One can get a sense of this by holding one's thumb out at arm's length. The nail subtends an area approximately the size of the fovea. Look around the room at the objects one can recognize. Likely, most are far larger than the fovea, implying that recognition involves significant peripheral processing.

Do we, then, recognize objects by executing multiple fixations? Imagine looking at the world through a fovea-sized aperture. Although moving one's eyes certainly provides useful information, several lines of research indicate that recognition does not proceed merely from piecing together multiple foveal views. First, the impairment or lack of peripheral vision (e.g., from glaucoma or retinitis pigmentosa) greatly impairs mobility (Lévy-Schoen 1976). If it were normal to recognize objects predominantly by moving one's fovea, these patients should not be so impaired. Similar results have been found for normal vision when the observer is forced to perform tasks by looking through a small and movable aperture (e.g., Hochberg 1968, Young & Hulleman 2013). By comparison, patients with age-related macular degeneration, who lack foveal vision, do not perform significantly worse than controls on a standard mobility task (Popescu et al. 2011). In part, this is simple math: The fovea subtends so little of the visual field and therefore captures very little of the available information. Nonetheless, it belies claims of the necessity of foveal vision for recognition.

Second, extensive work in the past 40 years has shown that observers can perform many recognition tasks with short presentation times that preclude multiple fixations. This includes object recognition (e.g., Mace et al. 2009), scene perception (e.g., Potter 1975, Loftus & Ginn 1984, Rousselet et al. 2005b, Loschky et al. 2007, Greene & Oliva 2009, Potter & Fox 2009), and face perception (e.g., Esteves & Öhman 1993, Costen et al. 1994, Loffler et al. 2005). In a single fixation, very little of the object or scene lands in the fovea. Recognition must heavily rely upon peripheral vision.

In summary, peripheral acuity and color vision are better than is often implied. The available acuity, in particular, is considerably higher than our visual systems need for many tasks, belying the need to "bring the high-resolution fovea to bear" in order to perform tasks such as recognition.

View 2: Peripheral Vision Is Like Foveal Vision, with a Bit Lower Resolution

The second common view regards peripheral vision as nearly indistinguishable from foveal vision, rather than as greatly impoverished because of reduced acuity. Although one can find many explicit statements of the first view, View 2 tends to crop up in an implicit rather than an explicit way.

For instance, suppose a researcher runs a visual search experiment in which observers must find a target item among a number of other distractor items. As in many search experiments, the researcher picks the target and distractors such that individual items seem easy to distinguish. Nonetheless, the researcher finds that search is inefficient—in other words, that it becomes significantly slower as one adds more distractors. Why is search difficult? One can easily discriminate the target from distractors when looking directly at them. The poor search performance implies that vision is not the same everywhere, or, as Julian Hochberg put it, "vision is not everywhere dense" (Hochberg 1968). If vision were the same throughout the visual field, search would be easy.

What is our default reason that vision is not the same everywhere, when considering this and other phenomena? Clearly foveal and peripheral vision differ, owing to our foveated visual systems. And some of the stimuli in the search experiment lie in the periphery. However, in many cases, researchers have not examined this straightforward explanation. [As summarized by Rosenholtz et al. (2012b), there have been notable exceptions, including examinations by Geisler & Chou (1995), Carrasco et al. (1995), Carrasco & Frieder (1997), Vlaskamp et al. (2005), Geisler et al. (2006), Wertheim et al. (2006), Gheri et al. (2007), Najemnik & Geisler (2008), and Michel & Geisler (2011).] Perhaps researchers look elsewhere because discriminating the target and distractor, by construction, does not require high acuity, and vision research has, until recently, poorly understood other differences between foveal and peripheral vision. Instead, in many cases, researchers have suggested that the way in which vision is not the same everywhere has to do with attention: Attention is a limited resource, and vision is better where the observer attends than where he or she does not. That researchers prefer an explanation based on attention—ill defined, difficult to measure, and often not explicitly manipulated by the experiment—demonstrates a profound lack of faith in peripheral vision as an explanation. This implicitly assumes that no interesting or task-relevant differences exist between foveal and peripheral vision. In other words, it assumes that peripheral vision is like foveal vision, albeit with a bit lower resolution (View 2).

If View 2 were correct, then human vision would be full of puzzling results. Why is it so hard to find your keys? They may be quite visible once found and fixated. Why are observers so bad at spotting the difference between two images in a change blindness task (Rensink et al. 1997, Simons & Levin 1997)? When you look at the difference, it is easy to see. Why is it so hard to tell that an impossible figure, such as a devil's pitchfork, is impossible? When we look at one end, it clearly has two tines, whereas when we look at the other end, it has three. Why can it be so hard to trace a path through a maze? Near fixation, it is clear in which directions the path extends.

This implicit view has held sway for a long time. However, it ignores an important way in which peripheral vision differs from foveal: Peripheral vision is susceptible to clutter, as evidenced by the phenomena of visual crowding. Crowding points to significant and interesting qualitative differences between foveal and peripheral vision. These differences are likely task relevant for a wide variety of tasks. My lab has argued that one must control for, or account for, these differences before considering explanations based upon visual attention. Otherwise, one risks fundamental misunderstandings of both vision and attention.

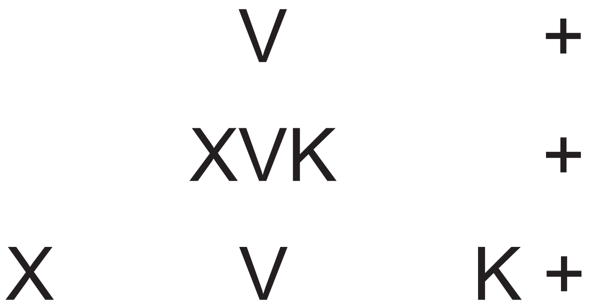

Figure 4

Classic demonstration of visual crowding. One can more easily recognize an isolated letter than one flanked by other symbols if the flankers are sufficiently close by. The effect occurs for stimuli in general—not just for letters or for arrays of items.

Crowding Is the Most Important Factor in Peripheral Vision

The loss of information in the periphery relative to the fovea goes beyond issues of acuity. Reduced peripheral acuity has a modest effect when compared with visual crowding. An example of this phenomenon appears in **Figure 4**. A reader fixating the cross in the top row will likely have no difficulty identifying the isolated letter on the left. However, the same letter can be difficult to recognize when flanked by additional letters (row 2). An observer might see these crowded letters in the wrong order. They might not see a V at all or might see strange letter-like shapes made up of a mixture of parts from several letters (Lettvin 1976). This effect does not result from a lack of resolution (Lettvin 1976). Move the flankers farther from the target and at some critical spacing, recognition is restored (row 3). Behavioral work suggests that the critical spacing is approximately 0.4 to 0.5 times the eccentricity for a fairly wide range of stimuli (Bouma 1970, Pelli et al. 2004). The critical spacing also shows some interesting anisotropies. First, there is the inward-outward anisotropy, in which for a given target eccentricity and spacing between target and flanker, a more eccentric (outward) flanker interferes with target recognition more than a more central (inward) flanker does [Banks et al. (1979) have one of the clearest demonstrations of this effect]. Second is the radial-tangential anisotropy, in which the critical spacing is larger by a factor of approximately 2:1 for flankers aligned radially with the target (as in **Figure 4**), compared with that for flankers positioned in the orthogonal direction (Toet & Levi 1992).

Crowding is likely to be a big effect when compared with acuity, as shown by plotting the critical spacing as a function of eccentricity along with functions relating various kinds of acuity to eccentricity (**Figure 5**). The slope for this crowding function is considerably higher than that for acuity, meaning that in some sense, peripheral vision degrades because of crowding faster than it does because of loss of resolution. These linear functions of eccentricity likely serve a useful purpose, as they make the information encoded about a stimulus relatively invariant to viewing distance (Van Essen & Anderson 1995). They are what allow us to view many of the figures in this review with little regard to viewing distance.

Crowding is a suprathreshold set of phenomena. Under conditions of crowding, one does not generally have difficulty detecting a target (Pelli et al. 2004), although "perceptual shortening" may occur, in which observers shown a crowded stimulus containing three letters report only two instead (Korte 1923). Furthermore, one can see effects of crowding for a wide range of stimuli: arrays of faces, oriented bars, letters, or colored circles. Pelli & Tillman (2008) gave a review and demonstrated crowding involving graphic symbols, real-world objects, etc. Furthermore, crowding need not involve discrete target and flanker items; Martelli et al. (2005) demonstrated

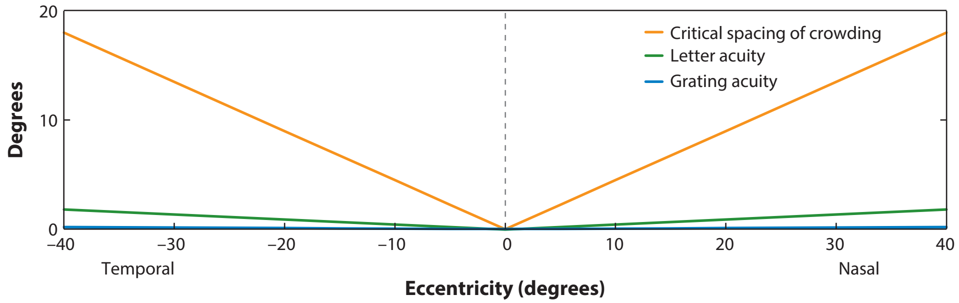

Figure 5

Scaling with eccentricity for grating acuity, letter acuity, and the critical spacing of crowding. Grating acuity shows a linear fit (Levi et al. 1985) to data from McKee & Nakayama (1984). Letter acuity shows Anstis's (1974) linear fit to threshold letter height. The critical spacing of crowding is often cited as being between 0.4 and 0.5 times the eccentricity. Here, we split the difference and show 0.45. The scaling for crowding is considerably worse than that for either kind of acuity.

that "self-crowding" occurs in peripheral perception of complex objects and scenes. One can easily see this by attending to peripheral vision and noting what details are difficult to discern or appear jumbled. By attending to the periphery in this way, one can confirm what crowding research has demonstrated: Crowding is likely to be task relevant for a very broad range of stimuli and tasks.

PERIPHERAL VISION, ATTENTION, AND A FUNDAMENTAL QUESTION IN VISION SCIENCE

Given that peripheral and foveal vision do differ in interesting ways (contrary to View 2), we need to re-evaluate many demonstrations that vision is not the same everywhere. Is the nature of our foveated visual system sufficient explanation for many of the observed phenomena? This investigation often pits peripheral vision against selective attention. As I discuss in this section, a fundamental question in vision science hangs in the balance.

What is attention? Researchers discuss many different kinds of attention (top-down versus bottom-up, selective versus diffuse, etc.). We can broadly define attention as a set of mechanisms that the brain uses to deal with limited capacity, or with limited resources more generally (although one typically excludes limited memory capacity). Because of the brain's limited capacity, our visual systems cannot automatically perform every possible visual task at the same time. Instead, the brain may concentrate resources on one task in a given moment, then switch to focus on another task. Attention is fundamentally about change, and in fact, we might redefine it as the short-term ways in which the brain changes its processing to adapt to the task(s) at hand. (Throughout, I use "task" in a broad way, not referring to the overall goal, but rather to the particular steps the visual system takes at a given moment to accomplish that goal.)

To what extent should we think of the brain's computations as dynamic, constantly changing to perform a new task, versus fixed and stable? Is it reasonable to think of the visual system as having a fixed encoding—in other words, that a given visual input will lead to relatively fixed, automatic processing, regardless of task? This is a fundamental question in vision science. The question is not whether the encoding is literally fixed, but rather, to the extent that one can approximate it as such, how much predictive power derives from knowing that encoding?

Much of the study of vision relies to some extent on the assumption of a fixed encoding. We would not bother trying to characterize the computations of various brain areas (V1, V2, etc.) if we thought those computations would change radically with task. Behaviorally, we often assume that performance at (for example) a Vernier task at a given location in the periphery allows us to predict the acuity available for other tasks, such as getting the gist of a scene.

At the fixed encoding end of the spectrum lie the mythical feed-forward models of vision. I say "mythical" because to my knowledge, no one believes they are correct. Rather, they serve precisely to test how much predictive power one gets from postulating a particular fixed encoding. By examining what a fixed encoding cannot predict, we can uncover the dynamic mechanisms of vision. At the other end of the spectrum, theoretically the brain could be totally dynamic, changing its computations to suit each task. But if it were, we would have a tough time understanding the underlying computations and architecture.

Many possibilities lie between. For instance, attention appears to modestly enhance the responses of V1 and V4 neurons (McAdams & Maunsell 1999). Although this encoding changes dynamically, so long as the modulations remain small, approximating the encoding as fixed might still have a fair amount of predictive power.

However, Chelazzi et al. (2001) found larger attentional effects in V4 when a monkey attended to one of two stimuli within a single receptive field. Depending upon which stimulus the monkey attended to, the cell responded as if only that stimulus was present in the receptive field. Such larger effects point to limitations in the predictive power of a fixed encoding, as the encoding changes dramatically with attention. Similarly, many theories of selective attention posit fairly radical changes in the information available with and without attention (i.e., as one changes the task). Even if one could think of the preattentive computations as the fixed encoding—for example, individual feature bands, as in feature integration theory (Treisman & Gelade 1980)—the predictive power of knowing this encoding remains fairly minimal, because much more information becomes available once an observer attends. Such theories have also posited "diffuse attention" to an entire scene or set of items (Treisman 2006), leading to the availability of yet again different information. In such theories, one really needs to know the task to know what information the brain encodes.

Let us return to the question: In what way is vision not the same everywhere? The selective attention explanation for search difficulty, change blindness, and other effects points to a highly dynamic visual system. The peripheral vision account, instead, asks to what extent differences in foveal versus peripheral encoding underlie the phenomena. If the peripheral vision account proves correct, a far more stable visual system is indicated. This remains an open possibility, given that View 2 is incorrect and that peripheral vision is interesting and likely task-relevant for a broad set of phenomena.

A STABLE, FOVEATED ENCODING OR DYNAMIC ATTENTIONAL MECHANISMS?

To test whether a fixed encoding is a good approximation, one needs a candidate model for that encoding—what information is preserved, if not the exact form in which it is available. If a single model can explain many results, that means there is value to thinking in terms of a fixed encoding. The model does not need to be perfect to test the hypothesis; but of course, the better it performs, the better we can estimate the predictive power achieved by making the approximation of a stable encoding. Occam's razor insists that all else being equal, we favor a simple fixed-encoding explanation over a more complex explanation in terms of a constantly changing visual system, even if the current instantiation of the simple explanation leaves some of the variance unexplained.

We can also use behavioral methods alone to re-examine phenomena that suggest vision is not the same everywhere and that have previously been attributed to dynamic attentional mechanisms. If we can demonstrate that the strengths and limitations of peripheral vision can (also) explain those phenomena, then we call into question the more dynamic account in favor of an explanation in terms of a more stable visual system.

A short example might help clarify the level of explanation we seek. In the case of attributing, say, search to attentional mechanisms, researchers did not mean to suggest that the conclusion "attention is required for feature binding" (Treisman & Gelade 1980) provides a full model of search. One would require additional mechanisms to direct attention, remember the target, identify each attended item, and decide when to stop. Similarly, a relatively low-level peripheral encoding does not even provide a full model of crowded object recognition, not to mention search, change blindness, or scene perception. Rather, the goal is to pinpoint a critical determinant of difficulty on those tasks. Both selective attention and peripheral encoding are hypothesized to lose certain details of the stimulus and preserve others. What performance do we expect, given the available information?

Here, I first discuss recent behavioral evidence in favor of a peripheral vision explanation. Then, I review our candidate model for the stable encoding and discuss evidence that our model predicts performance on a number of tasks.

Behavioral Evidence

In the traditional explanation of many not-the-same-everywhere phenomena, task difficulty arises from an inability to attend to more than one item at a time. To search for a target, for example, we must attend to one item in order to compare it with the target, then attend to another, and so on, in series, making performance slow. This explanation attributes the phenomena to differences in perception between attended and unattended locations. If this explanation is correct, then examining which tasks were easy versus which were difficult would allow us to infer the differences between attended and unattended vision.

Here, we ask instead whether the phenomena might have arisen from differences in perception between foveal and peripheral vision. Below, we describe some examples of how to test this idea for both search and change blindness.

Search. Suppose that search had nothing to do with selective attention. At a given moment, with a particular fixation, the visual system would simultaneously analyze all probable locations to see which were most likely to contain the target. The visual system would then either decide that it had identified the target or, in normal free-viewing search, execute an eye movement to gather more information. To understand the impact of peripheral encoding on visual search, we need to understand the key subtask of deciding whether a given peripheral location likely contains the target. How hard would search be if attention were not an issue, given losses in peripheral vision? To answer this, we measure performance discriminating a peripheral target at a known location, with full attention. If ease of identifying a peripheral target correlates with ease of finding that target, then search performance likely derives at least in part from limitations of peripheral vision.

We have measured peripheral discriminability of target-present from target-absent patches for both classic search conditions (search for a T among Ls, an O among Qs, a Q among Os, tilted among vertical, and a conjunction of luminance contrast and orientation), as well as for conditions that pose difficulties for selective attention models (cube search versus search for similar polygonal patterns without a three-dimensional interpretation). Each patch contained multiple, crowded search items, as would be the case in a typical search experiment. We found a strong relationship between search difficulty and performance on the peripheral task (Rosenholtz et al. 2012b, Zhang et al. 2015). Differences between foveal and peripheral vision are task relevant for visual search. This is a simpler explanation for why vision is not the same everywhere.

Change blindness. Observers can have difficulty detecting the difference between two similar images if we present them in such a way as to disrupt motion and other transients (Grimes 1996; McConkie & Currie 1996; Rensink et al. 1997; O'Regan et al. 1999, 2000). Typically, one can easily see the change when fixating it, making change blindness another not-the-same-everywhere phenomenon. Again, this effect is typically attributed to attention. Might peripheral vision instead play a role? In fact, researchers have found eccentricity effects. Observers are worse at detecting changes if they have fixated farther from the changed object, either before or after the change (Henderson & Hollingworth 1999, O'Regan et al. 2000, Pringle et al. 2001). However, in each of these cases, the researchers attributed the effects to attention.

For detecting a change, clearly the key subtask consists of deciding how likely a given peripheral location is to contain the change. We asked observers to distinguish between two frames of a standard change blindness stimulus when both the change and its location were known. Presumably, the observers attended to the known location. We found that the threshold eccentricity at which one can reliably discriminate a known change is predictive of difficulty detecting that change (L. Sharan, E. Park & R. Rosenholtz, manuscript under review). Our results suggest that the visual system looks for evidence of a change throughout the visual field, possibly in parallel (see also Parker 1978, Zelinksy 2001), rather than looking for a change merely where one fixates. However, that evidence of a change can be weak because of loss of information in peripheral vision, making change detection difficult.

These results by themselves do not preclude the possibility that attentional limits also play a role in the phenomena of change blindness and search. However, they certainly call into question the attentional explanations of both phenomena. Furthermore, we must question what we learned about encoding in unattended versus attended vision on the basis of these phenomena; rather than attention being the sole way in which vision is not the same everywhere, foveated vision clearly plays a significant role.

Some researchers have suggested that the losses in peripheral vision themselves arise out of limits to visual attention; in particular, that the critical spacing of crowding might represent the limited resolution of selective attention (Intriligator & Cavanagh 2001). Attributing crowding to attentional mechanisms confuses attempts to distinguish between attentional and peripheral vision explanations of various phenomena. We gain clarity by framing the question instead in terms of stable versus dynamic vision. Attentional limits pinpoint ways in which the visual system cannot change with task; they tell us about what is fixed and stable. The critical spacing of crowding may be an attentional limit in the same sense that cone density is—attention cannot make it better, and thus, one would better consider it part of the stable encoding.

Modeling

To go further in assessing the predictive power of a stable, eccentricity-dependent encoding, we need a candidate model for that encoding. Here, the phenomena of crowding and of peripheral vision more generally have provided some insight. Korte (1923) described that firm localization of both the letters and their constituent elements was extremely difficult and error prone when text was viewed in the periphery. More generally, peripheral vision is locally ambiguous in terms of the phase and location of features. Observers have difficulty distinguishing 180° phase differences in compound sine wave gratings in the periphery (Bennett & Banks 1991, Rentschler & Treutwein 1985) and show marked position uncertainty in a bisection task (Levi & Klein 1986). Furthermore, such ambiguities appear to exist in processing of real-world objects and scenes, although we rarely have the opportunity to be aware of them. Peripheral vision tolerates considerable image variation without giving us much sense that something is wrong (Freeman & Simoncelli 2011, Koenderink et al. 2012).

Lettvin (1976) observed that an isolated letter in the periphery seems to have characteristics that the same letter, when flanked, does not, and pointed to a possible explanation in terms of texture processing: "[The crowded letter] only seems to have a 'statistical' existence. . . . The texture of an isolated N specifies an N; the texture of an imbedded N specifies much less about the N as a form." Others have picked up on this suggestion, proposing that crowding phenomena result from "forced texture processing" (Parkes et al. 2001, Pelli et al. 2004) over local pooling regions that grow linearly with eccentricity (Bouma 1970).

To get a first cut of a model of peripheral vision, then, one might examine successful models of texture appearance. Echoing Lettvin (1976), we often think of texture as statistical in nature—in other words, as "stuff" that one can more compactly represent by its summary statistics than by the configuration of its parts (Rosenholtz 2014). The most successful, biologically plausible texture models have described texture using a rich set of image statistics.

For our initial candidate model (Balas et al. 2009), we chose as our set of image statistics those from a state-of-the-art model of texture appearance from Portilla & Simoncelli (2000): the marginal distribution of luminance; luminance auto correlation; correlations of the magnitude of responses of oriented V1-like wavelets across differences in orientation, neighboring positions, and scale; and phase correlation across scale. This seemingly complicated set of parameters is actually fairly intuitive: Computing a given second-order correlation merely requires taking responses of a pair of V1-like filters, point-wise multiplying them, and taking the average over the pooling region. This proposal (Balas et al. 2009, Freeman & Simoncelli 2011) is not so different from models of the hierarchical encoding for object recognition, in which later stages compute more complex features by measuring a sort of co-occurrence of features from the previous layer (Fukushima 1980, Riesenhuber & Poggio 1999, Krizhevsky et al. 2012, Yamins et al. 2014). Second-order correlations are essentially co-occurrences pooled over a substantially larger area.

The rich, high-dimensional set of statistics, pooled over sparse local image regions that grow linearly with eccentricity, provides an efficient, compressed encoding. It captures a great deal of information about the visual input. Nonetheless, the encoding is lossy, meaning one cannot generally reconstruct the original image. We hypothesize that the information maintained and lost by this encoding provides a significant constraint on peripheral processing and constitutes an important and often task-relevant way in which vision is not the same everywhere.

Other models of crowding, per se, do exist and may even explain some of the phenomena (Wolford 1975, Krumhansl & Thomas 1977, Wilkinson et al. 1997, Parkes et al. 2001, Baldassi et al. 2006, May & Hess 2007, Greenwood et al. 2009, van den Berg et al. 2010, Nandy & Tjan 2012, Chaney et al. 2014). If our goal here were to understand merely crowding, these models would warrant more detailed mention. However, none of the models in their current form are viable as a stable, general-purpose model of visual encoding. These models simply cannot make predictions for the range of stimuli and tasks required to test our broader hypothesis and hence are not relevant for the present discussion. We require an image-computable model that is applicable to arbitrary visual inputs. [Models by both Wilkinson et al. (1997) and Chaney et al. (2014) are image computable, but it is unclear how to extend them to a broader range of stimuli.] That said, we may decide that other image statistics work better than the initial candidate set from Portilla & Simoncelli (2000). In particular, as researchers develop other biologically plausible and effective texture models (e.g., perhaps Gatys et al. 2015), their features could provide alternative models of the local encoding.

To test whether our model can predict performance on a given visual task, one can incorporate techniques to generate, for a given input and fixation, images with approximately the same summary statistics (Portilla & Simoncelli 2000, Balas et al. 2009, Rosenholtz 2011, Rosenholtz et al. 2012a, and most elegantly, Freeman & Simoncelli 2011). Observers look at these synthesized images (**Figure 6** shows some examples) and perform essentially the original task, whether that be object

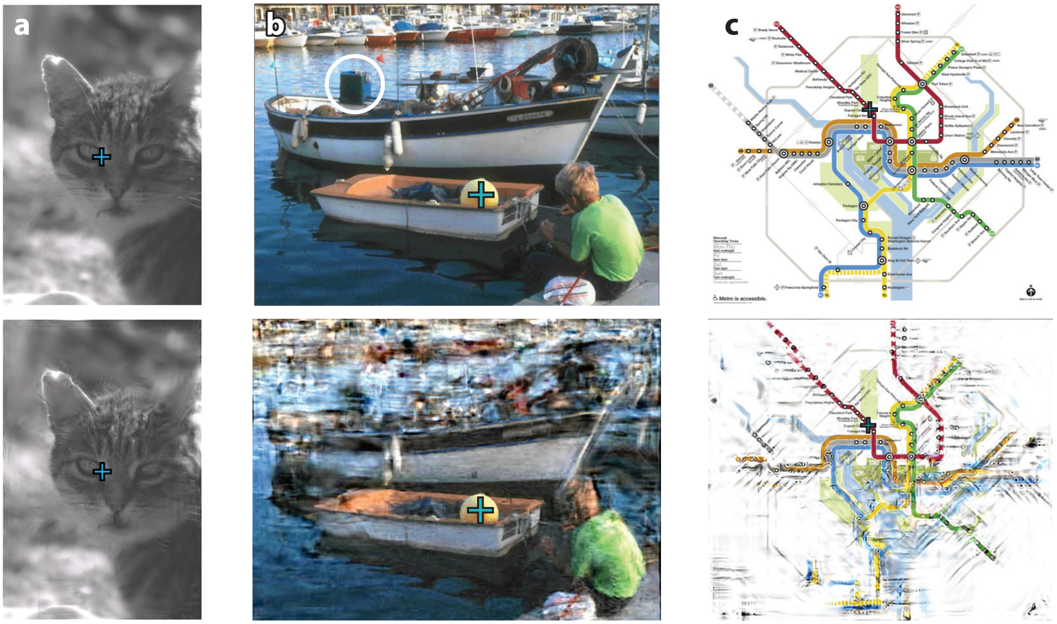

Figure 6

Images synthesized (*bottom*) to have approximately the same local summary statistics as the originals (*top*). Fixation indicated by the cross. (*a*) The hypothesized encoding seems adequate to support high human performance at object recognition. This material was originally published in Rosenholtz (2014) and has been reproduced by permission of Oxford University Press (**http://ukcatalogue.oup.com**). For permission to reuse this material, please visit **http://www.oup.co.uk/academic/rights/permissions**. (*b*) The encoding captures much information about the gist of a scene but is vague on the details. This is one frame from a change blindness demo. In the changed image, the blue box disappears. (The white circle indicates the location of the change but was not present in the original display.) From the information available when fixating the yellow buoy, one could easily imagine that detecting that change would be difficult. (*c*) This synthesis procedure can also lend insights into design, for example of transit maps. Map information that appears clear in such syntheses may be available at a glance.

recognition, scene perception, or some other task. This allows one to determine how inherently easy or difficult each task is, given the information lost and maintained by the proposed encoding. One could easily imagine this venture failing, even if a static encoding in vision is a reasonable approximation. Our current model of that encoding might simply be wrong. There might be significant losses of information later in the visual system, which would mask the effects of an earlier fixed encoding.

Crowding. Can our hypothesized encoding predict recognition of crowded symbols? Balas et al. (2009) flanked a peripheral letter with a variety of different flankers within the critical spacing of crowding (similar letters, dissimilar letters, bars, real-world objects, etc.). We showed that the inherent confusions and ambiguities of the hypothesized image statistics, pooled over the region of critical spacing, can predict difficulty recognizing the target. More recently (Keshvari & Rosenholtz 2016), we have also shown that this local representation can explain the results of three sets of crowding experiments, involving letter identification tasks (Freeman et al. 2012), classification of the orientation and position of a crossbar on t-like stimuli (Greenwood et al. 2012), and identification of the orientation, color, and spatial frequency of crowded Gabors (Põder & Wagemans 2007). Furthermore, Freeman & Simoncelli (2011) set the pooling region sizes and arrangement so as to ensure that observers have difficulty distinguishing between two synthesized images with the same local statistics. They find that with these pooling regions, they can predict the critical spacing of crowding for letter triplets.

Search. Rosenholtz et al. (2012b) and Zhang et al. (2015) further tested this model on identification of crowded symbols derived from visual search stimuli on the basis of the notion that an important subtask in visual search consists of examining peripheral regions for evidence of the target. Because of the large critical spacing of crowding, we should think of these regions as often containing multiple items—the key discrimination is between multi-item target-present and target-absent patches. The model predicts difficulty identifying crowded peripheral symbols from a variety of classic search tasks and in turn also predicts search difficulty. For example, when target and distractor bars differ significantly in orientation, the model statistics are sufficient to identify a crowded peripheral target, predicting easy popout search. Peripheral stimuli with white horizontal bars and black vertical bars can produce, according to the model, illusory color-orientation conjunctions, thus predicting both illusory conjunction phenomena and the difficulty of conjunction search. More recently, we have demonstrated that, with model in hand, we can subtly change classic search displays and correctly predict whether these changes make search easier or more difficult (H. Chang & R. Rosenholtz, manuscript under review). Characterizing visual search as limited by peripheral processing represents a significant departure from earlier interpretations that attributed performance to the limits of processing in the absence of covert attention (Treisman 1985).

Object recognition and scene perception. In order to be a general-purpose encoding, the proposed mechanisms must be operating not only during not-the-same-everywhere tasks, but also during normal object and scene processing. For object recognition, consider the image in **Figure 6*a***, synthesized to have the same local summary statistics as the original (Rosenholtz 2011, Rosenholtz et al. 2012a; see also Freeman & Simoncelli 2011). The fixated object is clearly recognizable even though substantial parts of it fall outside of the central 1.7°; it is quite well encoded by this representation. Glancing at a scene (**Figure 6*b***), much information is available to deduce the gist and guide eye movements. We have demonstrated that the information available is predictive of difficulty performing a number of navigation and other scene tasks at a glance versus with free viewing (K.A. Ehinger & R. Rosenholtz, manuscript under review). However, precise details are lost, which may be why change blindness occurs (Oliva & Torralba 2006, Freeman & Simoncelli 2011, Rosenholtz et al. 2012a).

Finally, Freeman & Simoncelli (2011) adjusted the size of the pooling regions until observers could not tell apart two synthesized images with the same local encoding. They demonstrated that observers have trouble distinguishing between such metamers regardless of whether the observers attend to regions with large differences. One can reinterpret this result as showing that they can predict performance telling apart two distorted real-world images on the basis of the information available in the local summary statistics. They used this technique to attempt to pinpoint where in the brain this encoding might occur; given the change in pooling region size with eccentricity required to produce metamers, they concluded that the mechanism resides in V2. These results have led to further physiological experiments which, for the first time, clearly distinguish between the responses of V1 and V2 (Freeman et al. 2013). Any attempt to localize the mechanisms of crowding within a particular brain area must meet the additional challenge of showing that the physiological receptive fields (RFs) agree with behavioral data regarding the critical spacing of crowding (Levi 2008, Nandy & Tjan 2012). However, we should not expect that the critical spacing directly specifies the size and shape of individual RFs. In recognizing a crowded peripheral object, the observer has access to multiple RFs/pooling regions, which combine to produce the behavioral effects. Multiple pooling regions that increase in size with eccentricity may suffice to produce the inward-outward anisotropy (and perhaps account for part of the radial-tangential anisotropy). The spatial layout of the RFs will also matter. Anisotropically increasing the density of the RFs provides additional information that could be used to better disambiguate the target when flanked tangentially.

In summary, both behavioral evidence and modeling suggest that peripheral vision plays a greater role than previously thought in not-the-same-everywhere phenomena. We must re-evaluate what those phenomena seemed to teach us about attention and revisit the fundamental question of to what degree vision is fixed versus dynamic. Strikingly, it seems that a single encoding may lose just the right sort of information to explain both the tasks human vision is bad at, such as inefficient search, and the tasks human vision is good at, such as getting the gist of a scene. One could imagine that a number of other phenomena might lend themselves to study using similar techniques: impossible figures, errors in judging shape or lighting direction, visual cognition tasks such as determining whether two points line on the same line, and so on. Beyond the study of vision, these techniques could enable better design by providing intuitions about the information available at a glance at something such as a map (**Figure 6*c***), the dashboard of a car, or a web page.

CONCLUSIONS

Peripheral vision is more interesting than commonly believed. Rather than being merely a second-class citizen to foveal vision or hopelessly low resolution, peripheral vision is useful and most likely supports a wide variety of visual tasks. However, peripheral vision is not just like foveal vision. Rather, mechanisms underlying the phenomena of crowding lead to eccentricity-dependent degradation that is task relevant for many tasks and stimuli.

Because of this relevance of peripheral information loss for many tasks, we as a field need to re-evaluate whether the underlying mechanisms might be behind a number of not-the-same-everywhere phenomena. Results of testing this hypothesis have so far been promising. If correct, the implications are deep and not merely relevant to search, change blindness, and so on; a fixed-encoding model may have more predictive power than previously thought. This is not to say that attention plays no role, as clearly there are effects of attention throughout visual processing. However, Occam’s razor essentially says that one should go with the more “boring” explanation (here, a fixed encoding) until it is proven that one needs a more interesting explanation (here, attentional mechanisms that change with task). Vision might theoretically be very “interesting” (in an Occam’s razor sense) and complex. But instead, it appears that that vision might in fact be more boring than previously thought. One might say that peripheral encoding mechanisms suggest that vision is boring in a really interesting way.

SUMMARY POINTS

1. The most important limit on peripheral vision comes from crowding, peripheral vision’s vulnerability to clutter. Crowding has a big influence on many real-world tasks.
2. Although peripheral vision also has poorer acuity and color vision compared with foveal vision, these effects usually have a minor impact compared with crowding.

3. We rely on peripheral vision for many visual tasks. Vision does not proceed predominantly through piecing together of multiple foveal views.
4. Crowding may be explained by a visual encoding in terms of a rich set of image statistics, pooled over sparse local regions that grow linearly in size with eccentricity. These regions overlap and tile the visual field.
5. Both behavioral experiments and modeling indicate that a number of visual phenomena previously attributed to dynamic attentional mechanisms may instead arise, at least in part, from a foveated visual system, with a largely fixed-encoding scheme that has lower fidelity in the periphery.

FUTURE ISSUES

1. Why might peripheral vision pool image statistics over such large regions, and are the pooled features in some sense optimal? Are there particular tasks that peripheral vision is optimized for, whereas performance can be sacrificed on other tasks?
2. What other phenomena might arise from lossy encoding in peripheral vision?
3. What does attention do? In what way does it affect the information available to perform a task? Also, physiological effects of attention have been observed even early in vision; what purpose do they serve?
4. How do we use this peripheral encoding to obtain a percept of a stable visual world and to maintain continuity of object representation across saccades?
5. If early visual encoding consists of image statistics pooled over sizeable regions, how does this affect our thinking about the visual system's algorithms for guiding eye movements, finding perceptual groupings, estimating shape, and invariant object recognition?
6. In developing version 2.0 of the model of visual encoding, what features are pooled? What happens with motion and stereo? Are there later losses of information due to pooling at other levels of visual processing, and if so, what is their nature?
7. How do we get from visual encoding to behavior? How do higher level decision-making processes use the information provided by peripheral vision?
8. Does this idea of pooling apply to other sensory systems? What can we learn about sensory encoding in general by modeling and studying peripheral vision?

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Thanks to Alexis Wynne Mogul for editing the manuscript, and Benjamin Wolfe for useful comments.

LITERATURE CITED

Abramov I, Gordon J, Chan H. 1991. Color appearance in the peripheral retina: effects of stimulus size. *J. Opt. Soc. Am. A* 8(2):404–14

Anstis SM. 1974. A chart demonstrating variations in acuity with retinal position. *Vis. Res.* 14:589–92

Anstis SM. 1998. Picturing peripheral acuity. *Perception* 27(7):817–25

Balas BJ, Nakano L, Rosenholtz R. 2009. A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9(12):13. doi: 10.1167/9.12.13

Baldassi S, Megna N, Burr DC. 2006. Visual clutter causes high-magnitude errors. *PLOS Biol.* 4(3):e56. doi: 10.1371/journal.pbio.0040056

Banks WP, Larson DW, Prinzmetal W. 1979. Asymmetry of visual interference. *Percept. Psychophys.* 25:447–56

Bennett PJ, Banks MS. 1991. The effects of contrast, spatial scale, and orientation on foveal and peripheral phase discrimination. *Vis. Res.* 31(10):1759–86

Bouma H. 1970. Interaction effects in parafoveal letter recognition. *Nature* 226:177–78

Carrasco M, Evert DL, Chang I, Katz SM. 1995. The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Percept. Psychophys.* 57:1241–61

Carrasco M, Frieder KS. 1997. Cortical magnification neutralizes the eccentricity effect in visual search. *Vis. Res.* 37:63–82

Chaney W, Fischer J, Whitney D. 2014. The hierarchical sparse selection model of visual crowding. *Front. Integr. Neurosci.* 8:73. doi: 10.3389/fnint.2014.00073

Chelazzi L, Miller EK, Duncan J, Desimone R. 2001. Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb. Cortex* 11(8):761–72

Costen NP, Craw I, Ellis HD, Shepherd JW. 1994. Masking of faces by facial and non-facial stimuli. *Vis. Cogn.* 1:227–51

Curcio CA, Sloan KR, Kalina RE, Hendrickson AE. 1990. Human photoreceptor topography. *J. Comp. Neurol.* 292:497–523

Esteves F, Öhman A. 1993. Masking the face: recognition of emotional facial expressions as a function of the parameters of backward masking. *Scand. J. Psychol.* 34(1):1–18

Freeman J, Chakravarthi R, Pelli DG. 2012. Substitution and pooling in crowding. *Atten. Percept. Psychophys.* 74(2):379–96. doi: 10.3758/s13414-011-0229-0

Freeman J, Simoncelli EP. 2011. Metamers of the ventral stream. *Nat. Neurosci.* 14(9):1195–201

Freeman J, Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA. 2013. A functional and perceptual signature of the second visual area in primates. *Nat. Neurosci.* 16(7):974–81

Fukushima K. 1980. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* 36:193–202

Gatys LA, Ecker AS, Bethge M. 2015. Texture synthesis using convolutional neural networks. *Proc. Neural Inf. Process. Syst. 2015, Montreal, Can., Dec. 7–12.* **http://arxiv.org/pdf/1505.07376**

Geisler WS, Chou K-L. 1995. Separation of low-level and high-level factors in complex tasks: visual search. *Psychol. Rev.* 102:356–78

Geisler WS, Perry JS, Najemnik J. 2006. Visual search: the role of peripheral information measured using gaze-contingent displays. *J. Vis.* 6(9):1. doi: 10.1167/6.9.1

Gheri C, Morgan MJ, Solomon JA. 2007. The relationship between search efficiency and crowding. *Perception* 36:1779–87

Greene MR, Oliva A. 2009. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.* 58:137–76

Greenwood JA, Bex PJ, Dakin SC. 2009. Positional averaging explains crowding with letter-like stimuli. *PNAS* 106:13130–35. doi: 10.1073/pnas.0901352106

Greenwood JA, Bex PJ, Dakin SC. 2012. Crowding follows the binding of relative position and orientation. *J. Vis.* 12(3):18. doi: 10.1167/12.3.18

Grimes J. 1996. On the failure to detect changes in scenes across saccades. In *Perception*, Vol. 2, ed. K Akins, pp. 89–110. New York: Oxford Univ. Press

Hansen T, Pracejus L, Gegenfurtner KR. 2009. Color perception in the intermediate periphery of the visual field. *J. Vis.* 9(4):26

Henderson JM, Hollingworth A. 1999. The role of fixation position in detecting scene changes across saccades. *Psychol. Sci.* 10:438–43

Hochberg J. 1968. In the mind's eye. In *Contemporary Theory and Research in Visual Perception*, ed. RN Haber, pp. 309–31. New York: Holt, Rinehart, and Winston

Horton JC, Hoyt WF. 1991. The representation of the visual field in human striate cortex. *Arch. Ophthalmol.* 109:816–24

Intriligator J, Cavanagh P. 2001. The spatial resolution of visual attention. *Cogn. Psychol.* 43:171–216

Johnson J. 2010. *Designing with the Mind in Mind: Simple Guide to Understanding User Interface Design Rules.* San Francisco: Morgan Kaufmann

Keshvari S, Rosenholtz R. 2016. Pooling of continuous features provides a unifying account of crowding. *J. Vis.* 16(3):39

Koenderink J, Richards W, van Doorn AJ. 2012. Space-time disarray and visual awareness. *i-Perception* 3(3):159–65

Korte W. 1923. Über die Gestaltauffassung im indirekten Sehen. *Z. Psych.* 93:17–82

Krizhevsky A, Sutskever I, Hinton GE. 2012. ImageNet classification with deep convolutional neural networks. *Proc. Neural Inf. Process. Syst. 2012*, *Lake Tahoe*, *NV*, *Dec. 3–8*, pp. 1097–105

Krumhansl CL, Thomas EAC. 1977. Effect of level of confusability on reporting letters from briefly presented visual displays. *Percept. Psychophys.* 21(3):269–79

Lettvin JY. 1976. On seeing sidelong. *Sciences* 16(4):10–20

Levi DM. 2008. Crowding—an essential bottleneck for object recognition: a mini-review. *Vis. Res.* 48:635–54

Levi DM, Klein SA. 1986. Sampling in spatial vision. *Nature* 320:360–62

Levi DM, Klein SA, Aitsebaomo AP. 1984. Detection and discrimination of the direction of motion in central and peripheral vision of normal and amblyopic observers. *Vis. Res.* 24(8):789–800

Levi DM, Klein SA, Aitsebaomo AP. 1985. Vernier acuity, crowding, and cortical magnification. *Vis. Res.* 25(7):963–77

Lévy-Schoen A. 1976. Exploration et connaissance de l'espace visual sans vision périphérique. *Trav. Hum.* 39:63–72

Loffler G, Gordon GE, Wilkinson F, Goren D, Wilson HR. 2005. Configural masking of faces: evidence for high-level interactions in face perception. *Vis. Res.* 45(17):2287–97

Loftus GR, Ginn M. 1984. Perceptual and conceptual masking of pictures. *J. Exp. Psychol. Learn. Mem. Cogn.* 10(3):435–41

Loschky LC, Sethi A, Simons DJ, Pydimarri TN, Ochs D, Corbeille JL. 2007. The importance of information localization in scene gist recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 33(6):1431–50

Mace MJ-M, Joubert OR, Nespoulous J, Fabre-Thorp M. 2009. The time-course of visual categorizations: You spot the animal faster than the bird. *PLOS ONE* 4(6):e5927

Martelli M, Majaj NJ, Pelli DG. 2005. Are faces processed like words? A diagnostic test for recognition by parts. *J. Vis.* 5(1):6

May KA, Hess RF. 2007. Ladder contours are undetectable in the periphery: a crowding effect? *J. Vis.* 7(13):9. doi: 10.1167/7.13.9

McAdams CJ, Maunsell JHR. 1999. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* 19(1):431–41

McConkie GW, Currie CB. 1996. Visual stability across saccades while viewing complex pictures. *J. Exp. Psychol. Hum. Percept. Perform.* 22(3):563–81

McKee SP, Nakayama K. 1984. The detection of motion in the peripheral visual field. *Vis. Res.* 24(1):25–32

Michel M, Geisler WS. 2011. Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *J. Vis.* 11(1):18. doi: 10.1167/11.1.18

Najemnik J, Geisler WS. 2008. Eye movement statistics in humans are consistent with an optimal search strategy. *J. Vis.* 8(3):4. doi: 10.1167/8.3.4

Nandy AS, Tjan BS. 2012. Saccade-confounded image statistics explain visual crowding. *Nat. Neurosci.* 15:463–69. doi: 10.1038/nn.3021

Oliva A, Torralba A. 2006. Building the gist of a scene: the role of global image features in recognition. *Prog. Brain Res.* 155:23–36

O'Regan JK, Deubel H, Clark JJ, Rensink RA. 2000. Picture changes during blinks: looking without seeing and seeing without looking. *Vis. Cogn.* 7(1–3):191–211

O'Regan JK, Rensink RA, Clark JJ. 1999. Change-blindness as a result of 'mudsplashes.' *Nature* 398(4):34

Østerberg G. 1935. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmol. Suppl.* 6–10:11–96

Parker RE. 1978. Picture processing during recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 4(2):284–93

Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M. 2001. Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* 4:739–44

Pelli DG, Palomares M, Majaj N. 2004. Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J. Vis.* 4(12):12

Pelli DG, Tillman KA. 2008. The uncrowded window for object recognition. *Nat. Neurosci.* 11(10):1129–35

Peripheral acuity. 2012. *Illusions*, November 20. **http://anstislab.ucsd.edu/2012/11/20/peripheral-acuity/**

Põder E, Wagemans J. 2007. Crowding with conjunctions of simple features. *J. Vis.* 7(2):23. doi: 10.1167/7.2.23

Popescu ML, Boisjoly H, Schmaltz H, Kergoat M-J, Rousseau J, et al. 2011. Age-related eye disease and mobility limitations in older adults. *Investig. Ophthal. Vis. Sci.* 52:7168–74

Portilla J, Simoncelli EP. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* 40(1):49–71. doi: 10.1023/A:1026553619983

Potter MC. 1975. Meaning in visual search. *Science* 187:965–66

Potter MC, Fox LF. 2009. Detecting and remembering simultaneous pictures in a rapid serial visual presentation. *J. Exp. Psychol. Hum. Percept. Perform.* 35:28–38

Pringle HL, Irwin DE, Kramer AF, Atchley P. 2001. The role of attentional breadth in perceptual change detection. *Psychon. Bull. Rev.* 8(1):89–95

Rensink RA, O'Regan JK, Clark JJ. 1997. To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* 8:368–73

Rentschler I, Treutwein B. 1985. Loss of spatial phase relationships in extrafoveal vision. *Nature* 313:308–10

Riesenhuber M, Poggio T. 1999. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2(11):1019–25

Rosenholtz R. 2011. What your visual system sees where you are not looking. *Proc. SPIE 7865*, *Hum. Vis. Electron. Imaging*, *XVI*, *San Francisco*, *Feb. 2.*

Rosenholtz R. 2014. Texture perception. In *Oxford Handbook of Perceptual Organization*, ed. J Wagemans, pp. 167–86. Oxford, UK: Oxford Univ. Press. doi: 10.1093/oxfordhb/9780199686858.013.058

Rosenholtz R, Huang J, Ehinger KA. 2012a. Rethinking the role of top-down attention in vision: effects attributable to a lossy representation in peripheral vision. *Front. Psychol.* 3:13. doi: 10.3389/fpsyg.2012.00013

Rosenholtz R, Huang J, Raj A, Balas BJ, Ilie L. 2012b. A summary statistic representation in peripheral vision explains visual search. *J. Vis.* 12(4):14. doi: 10.1167/12.4.14

Rousselet GA, Husk JS, Bennett PJ, Sekuler AB. 2005a. Spatial scaling factors explain eccentricity effects on face ERPs. *J. Vis.* 5(10):1. doi: 10.1167/5.10.1

Rousselet GA, Joubert O, Fabre-Thorpe M. 2005b. How long to get to the "gist" of real-world natural scenes? *Vis. Cogn.* 12(6):852–77

Simons DJ, Levin DT. 1997. Change blindness. *Trends Cogn. Sci.* 1:261–67

Strasburger H, Rentschler I, Jüttner M. 2011. Peripheral vision and pattern recognition: a review. *J. Vis.* 11(5):13

Toet A, Levi DM. 1992. The two-dimensional shape of spatial interaction zones in the parafovea. *Vis. Res.* 32:1349–57

Torralba A. 2009. How many pixels make an image? *Vis. Neurosci.* 26(1):123–31

Treisman A. 1985. Preattentive processing in vision. *Comput. Vis. Graph. Image Process.* 31:156–77

Treisman A. 2006. How the deployment of attention determines what we see. *Vis. Cogn.* 14:411–43

Treisman A, Gelade G. 1980. A feature-integration theory of attention. *Cogn. Psychol.* 12:97–136

van den Berg R, Roerdink JBTM, Cornelissen FW. 2010. A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLOS Comput. Biol.* 6:e1000646

Van Essen DC, Anderson CH. 1995. Information processing strategies and pathways in the primate visual system. In *An Introduction to Neural and Electronic Networks*, ed. SF Zornetzer, JL Davis, C Lau, T McKenna, pp. 45–76. San Diego, CA: Academic. 2nd ed.

Vlaskamp BNS, Over EAC, Hooge ITC. 2005. Saccadic search performance: the effect of element spacing. *Exp. Brain Res.* 167:246–59

Watson AB. 2014. A formula for human retinal ganglion cell receptive field density as a function of visual field location. *J. Vis.* 14(7):15

Wertheim AH, Hooge ITC, Krikke K, Johnson A. 2006. How important is lateral masking in visual search. *Exp. Brain Res.* 170:387–401

Wilkinson F, Wilson HR, Ellemberg D. 1997. Lateral interactions in peripherally viewed texture arrays. *J. Opt. Soc. Am. A* 14:2057–68

Wolford G. 1975. Perturbation model for letter identification. *Psychol. Rev.* 82(3):184–99

Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, DiCarlo JJ. 2014. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* 111:8619–24. doi: 10.1073/pnas.1403112111

Young AH, Hulleman J. 2013. Eye movements reveal how task difficulty moulds visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 39(1):168–90. doi: 10.1037/a0028679

Zelinsky GJ. 2001. Eye movements during change detection: implications for search constraints, memory limitations, and scanning strategies. *Percept. Psychophys.* 63(2):209–25

Zhang X, Huang J, Yigit-Elliott S, Rosenholtz R. 2015. Cube search, revisited. *J. Vis.* 15(3):9. doi: 10.1167/15.3.9