Quantifying peripheral and foveal perceived differences in natural image patches to predict visual search performance

Anna E. Hughes*

Department of Physiology, Development & Neuroscience, University of Cambridge, Cambridge, UK
Present address: Division of Psychology and Language Sciences, University College London, London, UK

Rosy V. Southwell*

Department of Physiology, Development & Neuroscience, University of Cambridge, Cambridge, UK
Present address: Ear Institute, University College London, London, UK

Iain D. Gilchrist

School of Experimental Psychology, University of Bristol, Bristol, UK

David J. Tolhurst

Department of Physiology, Development & Neuroscience, University of Cambridge, Cambridge, UK

Duncan and Humphreys (1989) identified two key factors that affected performance in a visual search task for a target among distractors. The first was the similarity of the target to distractors (TD), and the second was the similarity of distractors to each other (DD). Here we investigate if it is the perceived similarity in foveal or peripheral vision that determines performance. We studied search using stimuli made from patches cut from colored images of natural objects; differences between targets and their modified distractors were estimated using a ratings task peripherally and foveally. We used search conditions in which the targets and distractors were easy to distinguish both foveally and peripherally ("high" stimuli), in which they were difficult to distinguish both foveally and peripherally ("low"), and in which they were easy to distinguish foveally but difficult to distinguish peripherally ("metamers"). In the critical metameric condition, search slopes (change of search time with number of distractors) were similar to the "low" condition, indicating a key role for peripheral information in visual search as both conditions have low perceived similarity peripherally. Furthermore, in all conditions, search slope was well described quantitatively from peripheral TD and DD but not foveal.

However, some features of search, such as error rates, do indicate roles for foveal vision too.

Introduction

Finding a target among distractor stimuli is an important skill in many everyday tasks. Laboratory studies of visual search have traditionally used arrays of simple, geometrical objects on a plain background, for example, Roman capital letters or Gabor patches (Treisman & Gormican, 1988; Wolfe & Horowitz, 2004). In such studies, the target might be a particular letter or color with the distractors differing from the target in one or more ways. Alternatively, the target might be a Gabor patch of one orientation and spatial frequency presented in a field of Gabors with different parameters (Gilchrist, Heywood, & Findlay, 1999; Joseph, Chun, & Nakayama, 1997; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012). Other researchers have used arrays composed of images based on natural objects (Alexander & Zelinsky, 2011, 2012).

One seminal framework for understanding search difficulty is given by Duncan and Humphreys (1989).

Within this framework, search time depends on three factors. First, search time increases as the target is made more similar to the distractors, i.e., as the target–distractor difference (TD) decreases. Second, search time increases as the distractors are made more different from one another, i.e., as the distractor–distractor difference (DD) increases. In fact, these two factors interact: Heterogeneity among the distractors becomes less important when the TD becomes greater (i.e., the importance of DD decreases as TD increases). Finally, the third factor determining search time is the number of items in the search array. Search time increases proportionately with set size so that it is customary to summarize search efficiency as the *search slope*, i.e., the rate at which search time increases with set size. This powerful framework for understanding search times or search efficiency has been replicated many times and with many kinds of stimuli (Alexander & Zelinsky, 2012; Bauer, Jolicoeur, & Cowan, 1996; D'Zmura, 1991; Foster & Ward, 1991; Macquistan, 1994; Nagy & Sanchez, 1990; Phillips, Takeda, & Kumada, 2006; Treisman & Gelade, 1980; Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992). One key issue in the application of this framework is defining how the differences between items in the display should be independently measured. After all, a measurement of these differences lies at the heart of the predictive power of the model.

Our own particular interest is to study search stimuli when the targets and distractors are naturalistic (rather than letters or Gabor patches) with the stimuli being colored photographs of natural scenes or objects (Asher, Tolhurst, Troscianko, & Gilchrist, 2013; Lovell, Gilchrist, Tolhurst, & Troscianko, 2009) in order to see how far the straightforward rules for search among discrete geometrical items can be extended to search in more natural scenes. In this context, it is potentially even more challenging to extract a single measure of the difference between two stimuli (Tolhurst et al., 2010).

The focus of Duncan and Humphreys' (1989) work was in describing the factors that determined search time rather than in formally quantifying the key predictive parameters (TD and DD) and in showing precisely how search time depends upon their magnitudes. However, such a quantification would provide a strong test of the framework. To understand how we might quantify differences between items in search, we need to consider the search process in more detail.

At the start of the search task, the subject is presumably holding a "template" of the target's appearance in his or her memory. In almost all situations, this template will have the detail consistent with foveating the target. At this point, the search array is presented. The first task for the visual system is to identify an item in the display that is a good candidate to be the target. This judgment, by definition, will be carried out using peripheral vision. This is our first candidate for how similarity has an impact on search: It is the similarity between the stored template and the *peripheral* information in vision that drives search performance. Following this process, the subject then allocates attention to that location and will often move his or her eyes to foveate it. This candidate item is now compared to the stored template using foveal vision. This comparison process will take time and depends on the foveal similarity between the fixated item and the stored template; this is our second candidate for how similarity has an impact on search performance. If the currently foveated item does not match the target, then other items in the search array need to be investigated, and the next item will again be selected using peripheral vision. If the target is different enough from the distractors, it might be possible to identify it even with peripheral vision. In the extreme case, the target will be easily detectable in peripheral vision, resulting in the "pop out" effect (Treisman & Gelade, 1980). However, in many search tasks, the target does not immediately pop out; then, if the presently foveated item is rejected, the subject must make eye movements (saccades) during search to foveate on further items that might potentially be the target. Those items that appear to be most similar to the target *as perceived with peripheral vision* are likely to be foveated next (Zelinsky, 2008).

Thus, both pop out and any decisions about where to make the next saccade are based on a comparatively degraded signal from the peripheral parts of the visual field, and peripheral information has indeed been shown to play an important role in visual search in a number of paradigms. In experiments using gaze-contingent displays, it has been shown that reducing the peripheral information available to the observer increases the length of search times and the number of fixations seen (Geisler, Perry, & Najemnik, 2006; Loschky & McConkie, 2002). In addition, the detectability of a sine wave grating in noise at different peripheral locations predicts the number and pattern of eye movements in a search task (Najemnik & Geisler, 2005, 2008, 2009). There is also evidence that set size effects become greater with increased eccentricity (Carrasco, Evert, Chang, & Katz, 1995; Carrasco & Yeshurun, 1998), further suggesting that peripheral vision acts as a constraint on the difficulty of visual search. Despite these findings, relatively few models of visual search consider the importance of peripheral vision (Zelinsky, 2008). In addition, there has been little consideration of whether the type of peripheral information available to the viewer affects search.

Perhaps if the perception of all kinds of stimuli were equally degraded by peripheral vision compared to foveal vision, it would not matter whether search times depended upon peripheral or on foveal vision. How-

ever, the fall-off in performance with eccentricity is not uniform for different tasks (Levi, Klein, & Aitsebaomo, 1985; Hess & Field, 1993), and therefore, the degree to which peripheral vision informs search may depend on the particular task. For example, color discrimination with naturalistic stimuli is fairly well preserved peripherally whereas changes in shape become much more difficult to discern (To, Gilchrist, Troscianko, & Tolhurst, 2011). The cortical representation of the periphery is not just a low-scale copy of the fovea (Tolhurst & Ling, 1988), and the peripheral degradation of visual information is more complex than just a simple "blurring" of the retinal inputs. Recent research has suggested that the loss of information at increasing eccentricities may be due to the representation of the visual input as a set of summary statistics (Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011; Levi, 2008; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Pelli & Tillman, 2008; Rosenholtz et al., 2012). These peripheral summary statistics may list what features are present but not necessarily where they are exactly or how they relate to each other; they might thus stay the same if an object's shape is changed but not if the color is changed. The same idea may explain the phenomenon of "crowding" whereby peripheral identification performance is reduced when there are other items nearby (Bouma, 1970; Levi, 2008; Pelli & Tillman, 2008).

That visual performance is degraded differently for different stimulus dimensions allows us to address the extent to which visual search times are determined by foveal or peripheral processes. Specifically, we can ask whether it is foveally or peripherally measured perceived differences among the targets and distractors that predict behavior (Zelinsky, 2008). It is possible to construct targets and distractors from natural images that are "metamers" (Freeman & Simoncelli, 2011). These are stimuli that are physically different (e.g., in shape) as can be seen foveally but that are perceived peripherally as being identical. In this way, targets and distractors could be constructed for a search task that might be easy to distinguish foveally but that might be difficult (or even impossible) to distinguish peripherally. Rosenholtz and colleagues have quantified this process using computer graphics techniques that allow them to synthesize stimuli with approximately the same image statistics as the original stimulus, creating images that they call "mongrels" (Balas et al., 2009; Rosenholtz et al., 2012). A key question therefore is how visual search is affected when using targets and distractors that are metamers or mongrels of each other (Rosenholtz et al., 2012)

In this paper, we study search with target and distractor stimuli constructed from digitized photographs of everyday objects, i.e., "natural scenes." We have previously shown (To et al., 2011) that changes in

the color of such stimuli can be identified as well peripherally as foveally, but changes in the spatial structure ("shape") within the test stimuli are hard to identify peripherally even when they are easy foveally; the latter, therefore, form a convenient opportunity to make "metamer" distractors. Thus, we have the basis for constructing homogeneous and heterogeneous search arrays that, in some instances, will be made of stimulus components that are near "metamers." Such search stimuli yield two clearly different predictions of search performance, depending on whether it is foveal or peripheral difference that is the predictor of search performance. If a search array uses distractors that are metamers of the target, then the TD will be high foveally and low peripherally. Search for such a target should be fast if foveal image difference is more important but slow if peripheral difference is more important. Rosenholtz et al. (2012) showed indirectly that it is peripheral discriminability that determines search slope for arrays of simple geometric items with homogeneous "mongrel" distractors.

When search is among simple stimuli, there are clear metrics available for measuring the difference among target and distractors, such as line length. Duncan and Humphreys' (1989) rules are successful in predicting search efficiency given these intrinsic measures of stimulus difference. However, for more complex stimuli, such as images of natural objects, such straightforward metrics of difference are often not available. There are countless dimensions along which images of real-world objects might differ perceptually, and these differences may not be directly relatable to any physical scale (Tolhurst et al., 2010). Importantly, any measure based simply on a physical scale is unlikely to highlight the differential effects of foveal and peripheral vision. In order to begin to apply Duncan and Humphreys' framework to search scenes involving naturalistic stimuli, we need a single measure of the perceived differences among these complex target and distractor items. In fact, we have already developed just such a direct measure of perceptual difference: We ask observers to give numerical ratings to represent the perceived magnitude of the difference between pairs of images as compared to a reference pair. Although the changes in the stimulus may be multidimensional, the numerical rating gives a single-valued measure of "visual difference" (To, Gilchrist, Troscianko, Kho, & Tolhurst, 2009; To et al., 2011; To, Lovell, Troscianko, & Tolhurst, 2010).

Thus, in this paper, we study search in arrays in which the targets and distractors are constructed from colored images of natural objects. The distractors are formed by manipulating the color or spatial form of the original targets, and we use a ratings methodology to estimate perceived differences between these distractors and the respective targets (TD) both foveally and
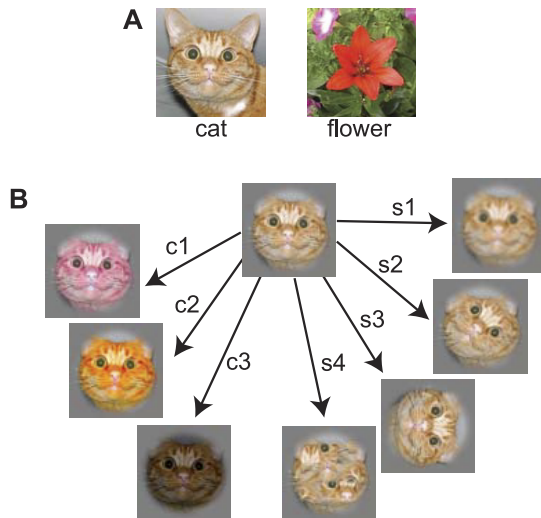
Figure 1. (A) Bitmap images of the two objects that were used to construct variant images. In a preliminary rating experiment, there were five such objects (Supplementary Figure 1A). (B) Some examples of variant image patches made from the "cat" parent image. The patches had a circular outline gradually blending into the gray background. "Color" changes: c1 is a change in hue, c2 is a change in chroma/saturation, and c3 is a change in overall brightness. "Shape" changes: s1 is blurred; s2 has the center of the image rotated and then blended into the image rim; in s3, the whole patch is rotated; and in s4, the central part of the image was broken up into nine squares, which were then shuffled and blended into each other.
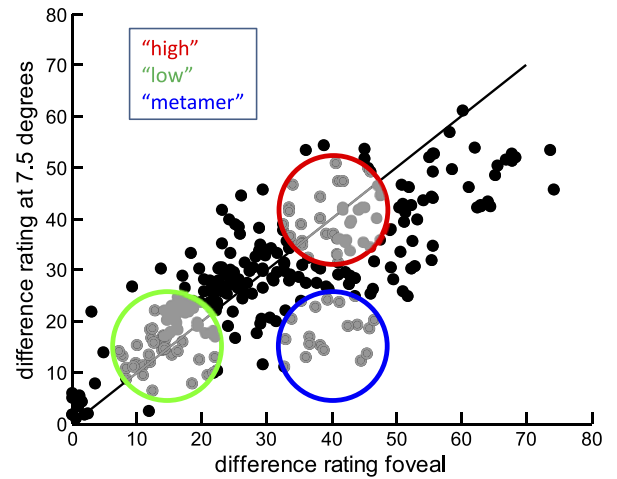

Figure 2. The results of a preliminary difference ratings experiment. The graphs plot the averages of nine observers' ratings for 286 image pairs. A peripherally viewed rating (7.5° eccentricity) is given on the ordinate plotted against the rating for the same pair seen foveally; the ratings are averaged across observers. The large colored circles show the kinds of foveal/peripheral perceived differences that we would ideally use to make the distractor patches in search arrays: Red have "high" perceived differences both foveally and peripherally, green have "low" perceived differences both foveally and peripherally, and blue have "metamer" perceived differences (high foveally, but low peripherally).

peripherally. We also estimate the perceived differences between different distractor variants (DD) foveally and peripherally. For those search arrays that dissociate the foveal and peripheral measures of perceived difference ("metamers"), we show that search is slow despite the large foveal perceived differences among the search items. The slow search for such arrays is quantitatively compatible with the small peripherally perceived differences among the items. We therefore provide the first direct evidence that peripheral information is critical in determining search slopes. However, we show a different pattern of results for search intercepts and error rates with these appearing to depend more upon foveal information, highlighting the need for both peripheral and foveal information in successful visual search.

Methods

The stimuli were presented on a 40 × 30 cm Sony CRT with a pixel resolution of 800 × 600. The CRT was viewed binocularly by the observers in a dimly lit room from a distance of 80 cm so that the display subtended 28.5° × 21.4° and each square pixel was 2.14

min. Stimuli were displayed under the control of a ViSaGe system (Cambridge Research Systems, Rochester, UK), which allowed precise control of stimulus timing and, crucially, of observers' reaction times during search (see below). The display was primarily a uniform mid-gray (60 cd.m^{-2}) except for a fixation spot and grid lines when appropriate and the actual rating or search stimuli. The stimuli were constructed from separate circular patches derived from colored photographs of natural scenes. The patches had a diameter of 68 pixels (2.4°), and their edges were blended into the gray of the background with a Gaussian window (Figures 1B and 3). For the rating experiments, only one such patch would be displayed at any one time; for the search experiment, there could be five, 10, or 15 nontouching patches presented concurrently in a search array.

Stimulus patch construction

Stimulus patches were constructed from square-cropped photographs of natural objects. In the preliminary ratings experiment (Supplementary Materials, section 1), we used five photographs, but for the main search and ratings experiments, we used only the cat and flower photographs depicted in Figure 1A.
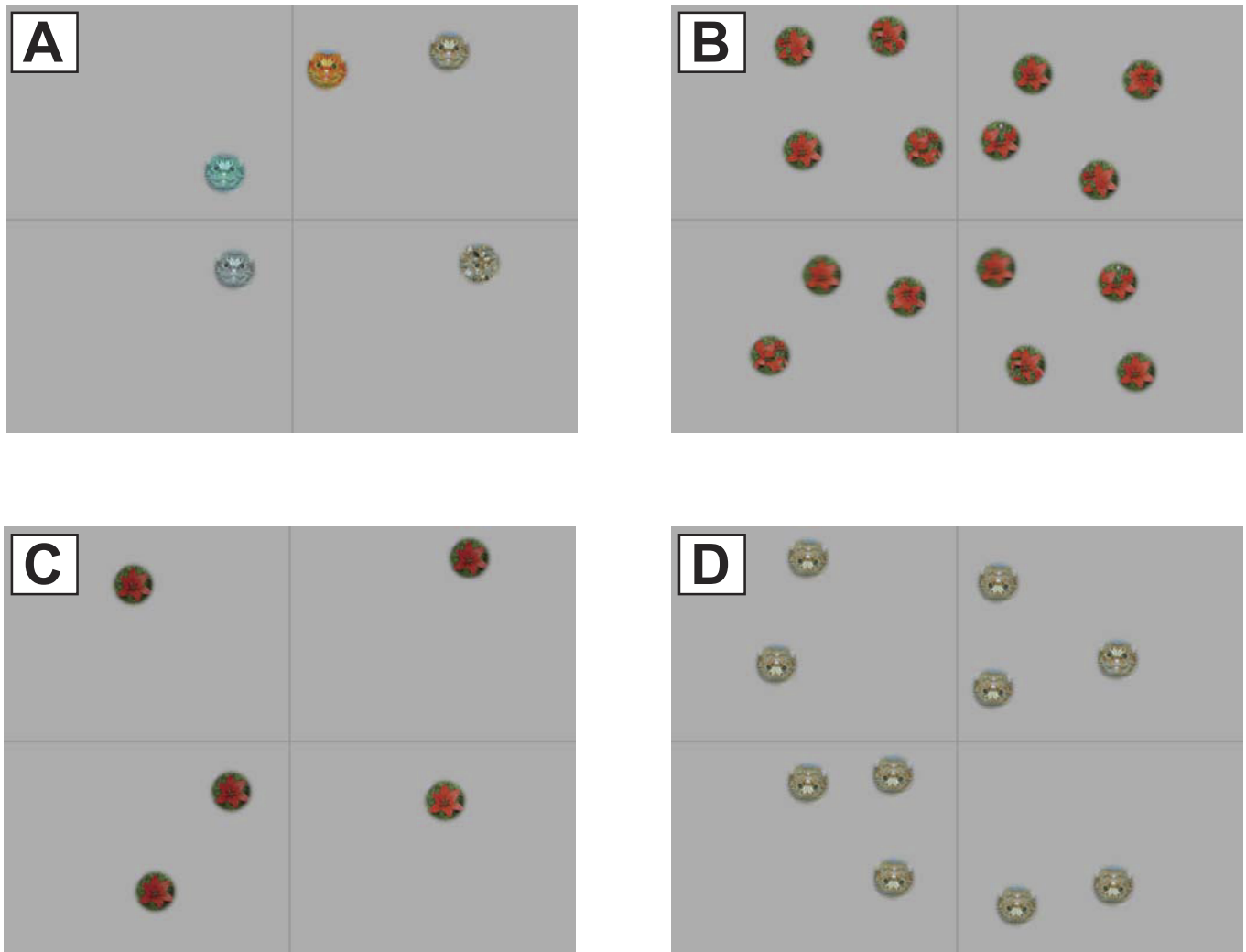
Figure 3. Four examples of search arrays, representing four of the 36 stimulus classes. The gridlines were to aid the observer in deciding in which quadrant the target was presented. The patches were arranged so as to keep the numbers in each quadrant as similar as possible. For stimulus construction (see Supplementary Material), each of the delineated major quadrants was considered to consist of four subquadrants. For each of the 36 stimulus classes, 10 different arrays were produced with the same constituent patches scattered in different quadrants or subquadrants. (A) Cat, five items, high TD and DD discriminability, heterogeneous distractors; the target patch is in the top right quadrant. (B) Flower, 15 items, metamer discriminability, heterogeneous; the target patch is in the top left quadrant. (C) Flower, five items, low discriminability, homogeneous; the target patch is in the lower right quadrant. (D) Cat, 10 items, metamer discriminability, homogeneous; the target patch is in the top right quadrant.

From an original photograph, a number of variants could be made (e.g., Figure 1B). For color changes, the RGB values of the whole or the central part of the photograph were converted approximately into L*c*h space (via L*a*b) using built-in routines in MATLAB (The MathWorks, Inc.), in which the hue (Figure 1B *c1*), chroma (saturation, *c2*), or luminance (*c3*) could be changed by varying amounts either alone or in combination. The modified L*c*h matrices were transformed back to RGB. Shape changes were intended to spatially reorganize the photograph without any changes in overall color. Figure 1B shows some examples. The photograph could be blurred (Figure 1B *s1*) by convolution with a 2-D Gaussian of varied size.

Alternatively, the whole or the central part of the image could be rotated to varying degrees (*s2*, *s3*). Finally, the whole or the central part of the image could be broken up into a number of squares (3 × 3 up to 10 × 10), and these squares could be randomly shuffled to generate a new image; the squares were blended with Gaussian edges to conceal the segment boundaries (*s4*). The aim was to produce a number of color or shape changes from each photograph that could range perceptually in magnitude from near-undetectable to extremely obvious.

The square originals and their variants were then resized (shrunk) to be 68 pixels square, and a flat-topped circular mask with Gaussian edges was applied.

This blended the edges of the circular image patches into the uniform mid-gray of the rest of the display without any visible hard edges (Figure 1B).

Participants

Different groups of observers participated in the various experiments (see details below). All observers were naïve to the purpose of the experiments and gave their informed consent to take part. The research was carried out in accordance with the Declaration of Helsinki.

Preliminary rating experiment

For each of the five original images (Supplementary Figure 1A), five color and five shape variants were constructed. The perceived difference between an original and each of its variants both peripherally and foveally was determined in a ratings experiment (To et al., 2011) whose details are given in the Supplementary Materials, section 1. Figure 2 plots the average of the ratings obtained by nine observers for each stimulus pair peripherally (ordinate) against the average rating for the same pair viewed foveally (abscissa). As found by To et al. (2011) and as discussed in the Supplementary Materials (Supplementary Figure 3A), the perceived magnitude ratings were generally lower peripherally than foveally with the shape changes being more affected by peripheral vision than the color changes.

The colored rings in Figure 2 identify three areas of interest on the graph. The red ring ("high") shows the kinds of stimulus pairings that evoked a large perceived difference rating both foveally and peripherally. The green ring ("low") shows stimuli that were hard to discern both foveally and peripherally. The blue ring shows "metamers" or "mongrels" (Freeman & Simoncelli, 2011; Rosenholtz et al., 2012). These pairs are perceptually clearly different when viewed foveally but are perceived as near-identical when viewed peripherally; these are key for the search experiment. It turned out that most of the successful metamer pairs in this preliminary experiment had come from the cat and flower families of image pairs, and so the experiments described below were carried out only with stimuli in those two families. In order to make search arrays, we required seven variants each of the cat and the flower in each of the three areas ("high," "low," and "metamer"). As there were cases in which we did not have seven variants for a condition, we had to construct additional variants judging the types and magnitudes of change from the stimuli that had been appropriately located in this preliminary experiment.

Search experiment

In each trial of the search experiment, the initial presentation was the target for the next search array (either the untransformed cat or flower image patch), presented in the center of the display. When observers indicated that they were ready by pushing a button on a CB6 response box (Cambridge Research Systems), this image was removed and was replaced briefly for 500 ms by a fixation spot in the center of the mid-gray display. The spot disappeared, and a search array was presented, at which point a timer was started on the ViSaGe system. Observers were instructed to fixate the central spot until the search array was presented; then they were free to move their gaze if they wished. The array consisted of five, 10, or 15 nonoverlapping patches, all drawn from the same family as the target that had just been displayed. One of the patches was the original untransformed image of the cat or flower (the target), and the other patches (the distractors) were all variants of the target (e.g., Figure 3 and Supplementary Figure 5). As well as the stimulus patches, the display included faint lines to split the screen into four obvious quadrants. The observer's task was to identify the quadrant in which the target was located and to respond quickly but accurately. Observers responded by pushing one of four buttons on the CB6 response box to indicate their choice of quadrant, and we recorded whether that choice was correct. The button push also halted the ViSaGe timer, and the observer's reaction time (search time) was recorded to a precision of 1 ms.

We generated 36 families of search array, comprising 18 families based on the cat image and 18 based on the flower image. For each of the 18 families based on one image, six were made with distractors drawn from image variants with "high" foveal and peripheral perceived difference; six with "low" perceived difference distractors; and six with "metamer" distractors. Within a group of six families, three were "homogeneous" distractor arrays (all the distractor patches were identical to each other) and three were "heterogeneous" in which the distractors were drawn nonrandomly from a pool of seven variants (see Supplementary Materials, section 2). Finally, each of the heterogeneous or homogeneous categories was constructed with three different set sizes (five, 10, and 15, i.e., with four, nine, and 14 distractor patches). See Supplementary Materials (section 2) for details of how homogeneous and heterogeneous distractors were selected and how the patches were placed pseudorandomly on the display.

For each of the 36 families, 10 different arrays were constructed. The locations of the target and distractor patches were chosen differently for each, and the particular distractor patches were chosen pseudoran-

domly from the appropriate pool of seven. Five observers viewed and responded to each of the 360 search arrays, and the order of presentation was different for each observer.

Correlation coefficients ($n = 360$ search arrays) were calculated between the search times of each observer and those of each other observer for the same stimuli. The correlation coefficients were positive but not high (range 0.32 to 0.58, $M = 0.42$). Low values may not be surprising. In the easiest searches, the target may visually "pop out" in peripheral vision and may be spotted immediately. However, for the harder search arrays, a wide range of search times is to be expected. In the absence of sufficient peripheral information to guide the gaze straight to the target, the first saccade might chance upon the correct target location but, in a more "unlucky" trial, a series of saccades over all of the rest of the visual field might have to be made before the target is finally found. The 360 arrays presented to each observer consisted of 10 different instances of each of 36 conditions. After discarding the trials in which the observer chose the wrong quadrant (see Results), we looked at the distributions of search time for the remaining (up to 10) instances of each condition and found that they were typically non-Gaussian but skewed to longer search times (cf. Reddi, Asrress, & Carpenter, 2003). As a measure of central tendency, we therefore took the median (rather than the mean) after discarding the times for those trials that resulted in a quadrant error. Now, the correlation coefficients ($n = 36$) between each observer's median search times and those of each other observer were much increased (range 0.70 to 0.90, $M = 0.83$).

Full rating experiment

We are interested in whether the measured search times and calculated search slopes (see Results) can be explained quantitatively as implied by Duncan and Humphreys (1989) from knowledge of the perceived TD and the perceived DD. The key issue of the present paper is whether TD and DD should be based on measures of differences in foveal or in peripheral vision. Because some new image patches had been generated to bring the total number of images in each family up to seven, not all TD and DD ratings were known from the preliminary ratings experiment. We therefore measured TD and DD for the actual stimulus patches used in constructing the search arrays, using a perceived magnitude ratings protocol (To et al., 2009; To et al., 2011; To et al., 2010) similar to that used in the preliminary ratings experiment and described in detail in the Supplementary Materials, section 1. We made measurements of TD and DD foveally and at two eccentricities: 5° and 12°. In the search arrays, the

nearest neighbor distances between patches was commonly about 5°, and the average distance between any two patches was about 12°.

We chose 128 pairs of stimulus patches for measurement. For both the cat and flower families, we compared the original (target) patch with all seven distractor patches in each of the "high," "low," and "metamer" classes, giving 42 pairs contributing to measures of TD. Given seven distractor patches in each family and class, we could potentially have measured six times 21 distractor–distractor pairings for the heterogeneous search arrays; however, we made measurements for only half of them (63). We also included 23 pairs in which a distractor patch was actually paired against itself, i.e., there was no change in the stimulus. The observers were told that there would be such "identity" pairs. They were included in order to formally measure the perceived differences between distractor patches in the homogeneous arrays. In addition, we hoped this would encourage observers to give rating magnitudes of zero if they perceived no difference in a pair of patches (even when there might actually be a difference). The rating values given to the identity pairs are shown in Figure 6.

Observers viewed the 128 image pairs at each of the three eccentricities. In a given trial, one of the pair was chosen randomly and was presented for 833 ms. After an interval of 83 ms while the fixation spot alone was visible, the other patch was presented (833 ms). Finally, after another brief interval, the first patch was presented a second time. The reasons for having three presentation intervals in a trial are given in To et al. (2010). Unlike in our previous experiments (To et al., 2009; To et al., 2011), the foveal and peripheral ratings were not collected in separate blocks, but foveal and peripheral presentations were randomly interleaved. The random interleaving of trials at different eccentricities was intended to ensure that the observers maintained a single rating scale for all stimuli irrespective of eccentricity. The observer fixated a spot to the left of screen center, and in any trial, the stimulus patches might appear at the spot or 5° or 12° to the right of the spot but slightly above the horizontal. For foveal stimulus presentations, the fixation spot was removed while the stimulus patches were actually present. The standard pair (Supplementary Figure 2) was presented *foveally* every 12 trials, i.e., at the fixation spot. At the end of the trial, the observer gave a numerical rating of perceived difference in proportion to the perceived difference of the foveally viewed standard pair, which was deemed to have a magnitude of "20." This approach was taken to provide participants with a single "anchor" for their all their ratings, foveal or peripheral. The 384 image pairs were presented in four blocks of 96 in a different random order for each of 11 observers.

The ratings given by the 11 observers to each stimulus pair were averaged. Standard errors of the means increased with increasing mean, being 15.8% of the averaged rating foveally (on average) and 21.4% at 12° peripherally. Comparing each observer's ratings for the 128 foveal stimuli with those of the other 10 observers gave correlation coefficients in the range 0.55–0.87 ($M = 0.76$); the correlations for peripheral viewing were slightly lower (e.g., at 12°, mean $r = 0.67$).

Results

Search in arrays of naturalistic image patches

Five observers each responded to 360 search arrays comprising five, 10, or 15 patches constructed from photographs of either a cat or a flower. Half the arrays were homogeneous (all distractor patches identical), and half were heterogeneous (distractors drawn from a pool of seven different variants of the target). There were three conditions of array, distinguished by the TD and, for the heterogeneous arrays, the DD. TD and DD could be "high" both for foveal and for peripheral vision, they could be "low" both foveally and peripherally, or they could be "metamers" (high perceived difference foveally but low difference peripherally).

We recorded search time (time between array onset and the observer's button push) and whether the observer had correctly identified the screen quadrant containing the target. In the 1,800 total trials, there were 77 errors (4.3%) in which the observers had chosen the wrong quadrant (see section 3A in Supplementary Materials for details). Most (69) of the errors were evoked by arrays in the "low" discriminability condition, especially for the heterogeneous arrays. The feature of the "low" condition distinguishing it from the other two conditions is in having low TD and DD discriminability *foveally*.

Figure 4 shows the median search times for the various kinds of search array. Each point shows the average and standard errors of 10 values: the averaged values for five observers for both the cat and flower versions of the particular kind of search array. Consider first the results for the "high" TD and DD (red) and "low" TD and DD conditions (green). These conform well to the Duncan and Humphreys (1989) model stated in the Introduction. Search time is longer when DD is greater, i.e., search time is greater for heterogeneous arrays (Figure 4B) than for homogeneous ones (Figure 4A), and it is longer when TD is low (green) compared to when it is high (red). Search time generally increases with set size, giving positive search slopes, although for the fastest searches (high TD, but
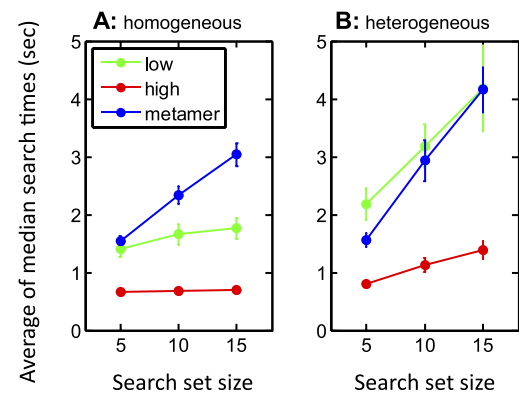

Figure 4. A summary of the search times for homogeneous (A) and heterogeneous (B) distractor arrays. Each graph shows search times as a function of set size for arrays of given distractor difficulty. Low discriminability among constituent patches (green); high discriminability (red); metamer discriminability (blue). There were 10 different arrays within each class. We took each observer's median search time from each class of 10 after discarding the time for any trials where the wrong quadrant was chosen; then we averaged those medians across five observers and across the cat and flower arrays. Averages of 10 items with standard errors are shown.

zero DD, red symbols in Figure 4A), set size seems to have little influence, consistent with fast "parallel" search (Treisman & Gelade, 1980). The key array conditions for our study are those with "metamer" distractors (Figure 4, blue). It is quite clear from the figure that the search times for "metamer" arrays are slow and are much more similar to those for "low" discriminability distractors than for "high," particularly for the heterogeneous condition. Although search times for "metamer" arrays may be similar to those for "low" arrays, the error rates (Supplementary Table 1) for "metamer" arrays are actually similar to those for "high" arrays. In addition, in the homogeneous condition, the search slope for the "metamer" condition appears to actually be steeper than that for the "low" condition.

Figure 4B suggests that the pattern of search times is very similar for the "low" and "metamer" heterogeneous arrays, consistent with the hypothesis that search time is dominated by peripheral perception of TD and DD differences. However, closer inspection of the results shows that this is an oversimplification and that it may not be possible to generalize all aspects of the response to all naturalistic image stimuli. Figure 5 replots the search time medians from Figure 4 and separates the results for the cat and flower target families. The pattern of results still looks the same for the homogeneous arrays (Figure 5A, B), but the pattern differs for heterogeneous distractor arrays. In the cat arrays (Figure 5C), the search times for "metamer" arrays are longer than for "low" whereas for the flower
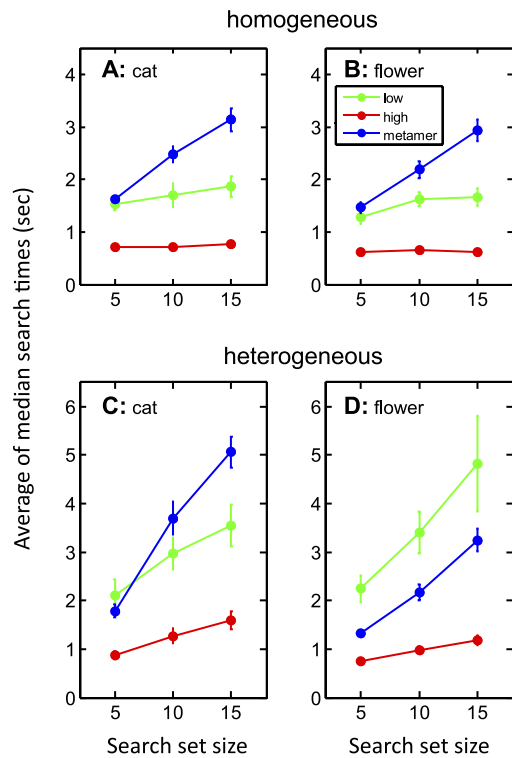
Figure 5. As with Figure 4, the search times for homogeneous distractor arrays (A and B) and heterogeneous arrays (C and D) are shown, but the results for cat (A and C) and flower arrays (B and D) are shown separately. Low discriminability among constituent patches (green); high discriminability (red); metamer discriminability (blue). Each point is the average of five observer medians, and ±1 *SE* is shown.

arrays (Figure 5D) the "metamer" search times are shorter than for "low." We will consider this difference again below (Figure 7).

Search slopes

Above, we have summarized our results on *search time*, but search efficiency is generally summarized as *search slope* (Duncan & Humphreys, 1989): the rate at which search time increases as the set size increases. To obtain a search slope for each observer, we fitted simple least-squares regressions to their median search times plotted against the three set sizes. The average search slopes across all observers are shown in Table 1A; the intercepts of the regressions (extrapolated "search time" for zero items) are also shown in Table 1B. The search times for "low" and "high" discriminability arrays are consistent with the Duncan and Humphreys (1989) rules: Slope is lower for homogeneous arrays than for heterogeneous, and it is higher for the "low" arrays (with low TD). For "high" discriminability homogeneous arrays, the search slope is near zero. The "metamer" arrays generally have

the highest search slopes of all. We ran a linear mixed model on the data (using 60 data points with 12 slopes for each participant) with the *lme4* package (version 1.11-11; Bates, Mächler, Bolker, & Walker, 2014) and the *lmerTest* package (version 2.0-30; Kuznetsova, Brockhoff, & Christensen, 2014) in R (version 3.2.3; Ihaka & Gentleman, 1996). The model contained picture family (cat vs. flower), distractor category (homogeneous vs. heterogeneous), and array type ("low," "metamer," or "high") as fixed factors and subject as a random factor. Overall main effects were generated using the *car* package (version 2.1-1; Fox & Weisberg, 2011). There was no effect of cat versus flower family, $\chi^2(1) = 0.151$, $p > 0.05$. There were effects on slope of whether distractors were homogeneous or heterogeneous, $\chi^2(1) = 21.700$, $p < 0.001$, and on discriminability class, $\chi^2(2) = 36.245$, $p < 0.001$; post hoc comparisons using Tukey tests in package *multcomp* (version 1.4-3; Hothorn, Bretz, & Westfall, 2008) showed that the "metamer" arrays had significantly higher slopes than the "high" arrays ($Z = 6.020$, $p < 0.001$) and the "low" arrays ($Z = 2.976$, $p = 0.008$). The "low" arrays also had significantly higher slopes than the "high" arrays ($Z = 3.044$, $p = 0.007$). The "metamer" arrays therefore had the steepest slopes, followed by the "low" arrays, with the "high" arrays having the flattest slopes. It was not possible to examine all the interactions between the three factors because of the relatively small number of data points in each condition.

Table 1B shows the intercepts of the search time regressions. A linear mixed model of the data (set up in exactly the same way as described for the slope model above) showed that there was no effect of cat versus flower family, $\chi^2(1) = 1.205$, $p > 0.05$. Nor was there an effect on intercept of whether distractors were homogeneous or heterogeneous, $\chi^2(1) = 3.350$, $p > 0.05$. However, there was a significant effect of target–distractor class, $\chi^2(2) = 21.741$, $p < 0.001$; post hoc testing using Tukey tests showed the "low" category had significantly higher intercepts (1215 ms) than the "metamer" ($Z = 4.147$, $p < 0.001$) or "high" ($Z = 3.919$, $p < 0.001$) categories (which were nearly the same, 550–590 ms on average; $Z = 0.228$, $p = 0.972$).

Overall, these results suggest that there are differences between all target types in both search slopes and search intercepts. "High" stimuli had flat search slopes and low intercepts, and "low" stimuli had steeper search slopes and relatively high intercepts. Like "low" stimuli, "metamer" stimuli showed a steep search slope but also showed a low search intercept more similar to "high" stimuli.

Discriminability among target and distractor patches

In order to explain the search times or search slopes quantitatively in terms of Duncan and Humphreys'
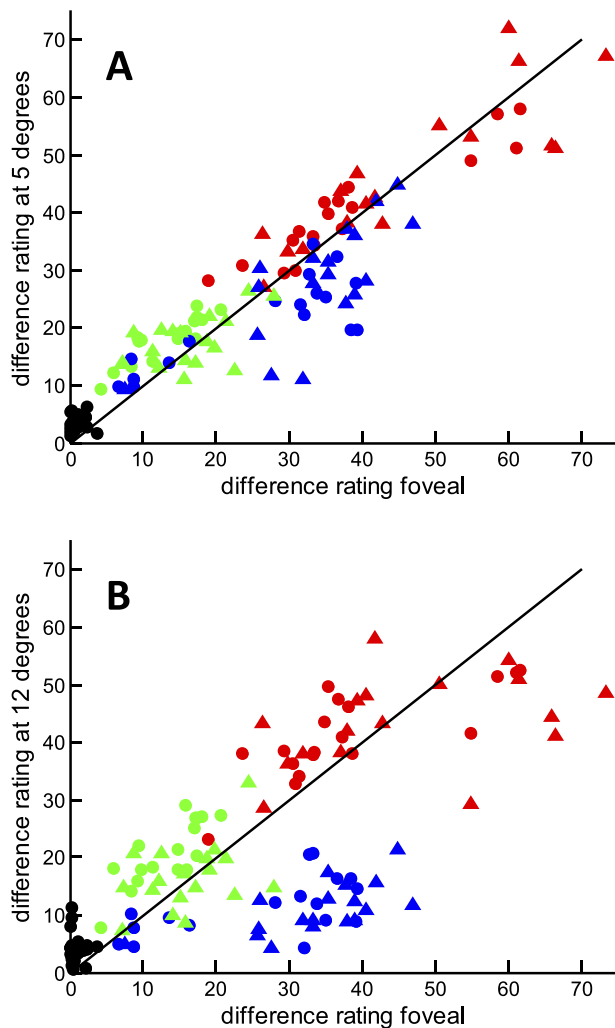
Figure 6. The graphs plot the averages of 11 observers' ratings for 128 image pairs. A peripherally viewed rating is given on the ordinate plotted against the rating for the same pair seen foveally; the ratings are averaged across observers. (A) Ratings for 5° eccentricity are plotted against foveal rating. Circles are "cat"; triangles are "flower." The line of identity is shown. As in Figures 4 and 5, low discriminability among constituent patches is green; high discriminability is red; metamer discriminability is blue. Black symbols are for 23 image pairs in which there was actually no change. (B) As with part A, but for ratings for 12° eccentricity plotted against foveal rating.

(1989) rules, one must measure the perceived differences among the targets and distractors actually used in constructing the search arrays. We obtained direct measures of perceived difference using a ratings magnitude protocol (see Methods and Supplementary Materials, section 1, for details). Figure 6 plots the difference ratings obtained with peripheral viewing of image pairs (ordinate) against foveal viewing (abscissa) of the same pairs (Figure 6A, 5° peripheral; Figure 6B, 12°). As in our previous studies (To et al., 2009; To et al., 2011), we found the perceived differences of most

image pairs to be reduced in peripheral vision compared to foveal, especially so for image pairs that differed in shape rather than color. However, Figure 6 (especially panel B) shows that many image pairs evoked *higher* perceived difference ratings peripherally (the line of equality is shown). We discuss this finding further in Supplementary Materials, section 3B. The color coding of the symbols follows Figures 2, 4, and 5, showing the discriminability between target and distractors and among distractors for the three classes of search array. As planned, the "high" patch pairs (red, mostly color changes) had high perceived differences both foveally and peripherally, and the "low" pairs (green, mostly shape changes) had low perceived differences foveally and peripherally. The blue symbols show that the "metamer" pairs had low perceived differences peripherally and moderately high perceived differences foveally. The split into three distinct classes is more obvious at the greater eccentricity (12°). This figure confirms that our design of the "metamer" search arrays used patches that were as discriminable foveally as the "high" stimuli but were as poorly discriminable peripherally as the "low" stimuli.

The black symbols in Figure 6 show the ratings given for those image pairs for which there was, in fact, no difference; ideally, the observers should have given a rating of zero to these stimuli. However, it can be seen that the averaged ratings of the 11 observers for these 23 image pairs are not zero, but more interestingly, they tend to be greater peripherally than foveally, particularly at 12°. These results are analyzed and discussed further in Supplementary Materials, section 3C.

Modeling search slopes and search times

The ratings experiment (Figure 6) determined the perceived difference foveally and at two eccentricities for all possible target–distractor combinations in the search arrays (TD) and also for many of the possible DDs. Table 2A shows averaged TD values for the different kinds of search arrays, both foveally and at 12° peripherally. Each of the TD estimates for homogenous search arrays is the average of the 11 observers' ratings for the two different distractor image patches used to make the search arrays. Each of the TD distances for the heterogeneous search arrays is the average of the 11 observers' ratings for seven stimuli. The bold text shows an example in which the "metamer" foveal and peripheral distances are very similar to the "high" foveal and the "low" peripheral distances, respectively, as planned; the italicized text shows a flower family example. The equivalent pairings for the heterogeneous arrays are not so close.
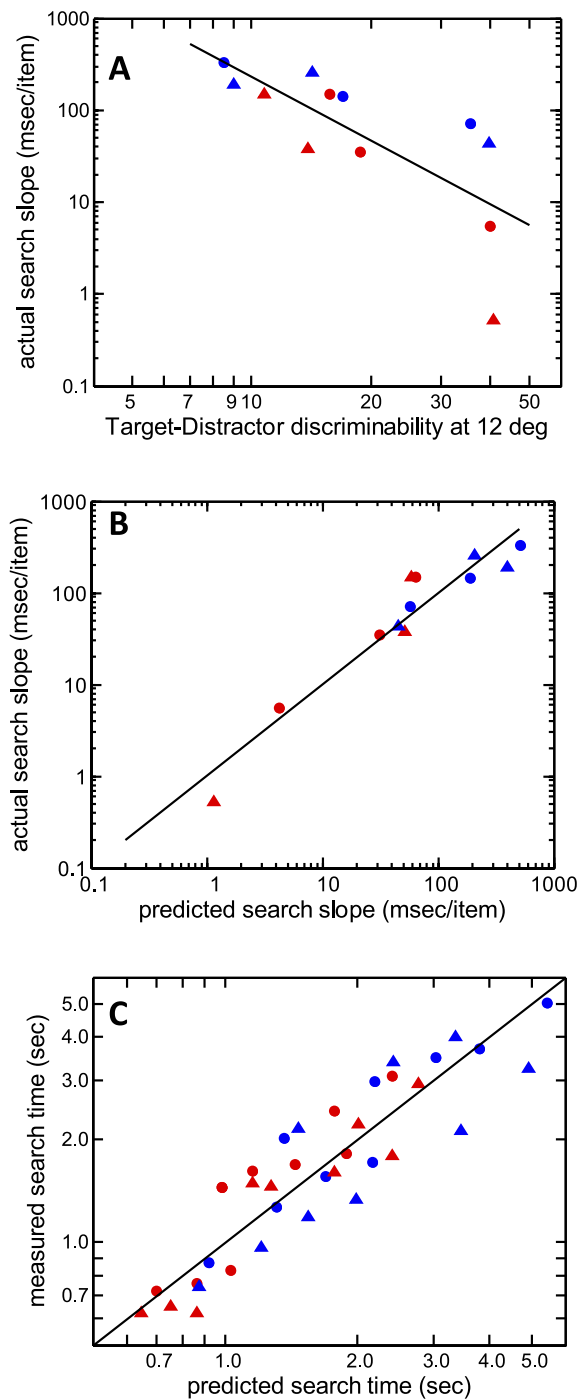
Figure 7. (A) Following Rosenholtz et al. (2012), search slope (*n* = 12) is plotted against the discriminability of the target from the distractors. TD discriminability is the averaged TD rating value for the array at 12° eccentricity. Red symbols for homogeneous distractor arrays; blue for heterogeneous. Circles for cat stimuli; triangles for flower stimuli. The black line is the regression through all 12 data. (B) The experimentally measured search slope (*n* = 12) is plotted against the slope predicted by a multilinear regression on TD <u>and</u> DD at 12°, Equation 2. Symbols as in part A; the black line is the line of equality. (C) The actual search time (*n* = 36) is plotted against the time predicted by the →

Table 2B shows averaged DD estimates. The estimates for homogeneous arrays are based on the particular patches used in those arrays (but see "Notes" to Table 2B); the estimates for heterogeneous arrays are simply averages across the pairs we tested, irrespective that different instances of an array will have sampled differently from the pool of seven distractor patches. For the homogeneous distractor arrays, the DD should, in theory, have been zero. However, in practice, observers did sometimes give nonzero ratings to image pairs when there was, in fact, no change (Figure 6). Not shown are the TD and DD estimates at 5° eccentricity.

We used these averaged estimates of TD and DD to model the search slopes pooled across observers (*n* = 12, Table 1A) and the averaged median search times (*n* = 36, Figure 5) for the various classes of search array. "*Search slope*" is the rate at which search time increases with set size. Duncan and Humphreys (1989) give us the intuition that search time should increase with the number of elements in the array (*size*). The slope should depend inversely on the perceived magnitude of the difference between the target patch and the distractor patches (TD distance) but should have a positive relationship with the perceived differences among the distractor patches (DD).

Rosenholtz et al. (2012) used a simple formulation to model their experiments: Search slope was inversely proportional to peripheral TD when plotted on log–log axes.

$$\log_{10}(\text{search slope}) = k + \beta_1\log_{10}(\text{TD}) \qquad (1)$$

Figure 7A plots the search slopes (Table 1A) from our experiments against the TD distance measured experimentally at 12° eccentricity, averaged over all observers and all TD instances appropriate to a given search array condition (Table 2A). The plot is log–log, following Rosenholtz et al. (2012), and our data have an inverse relationship as in their experiments (*r* = −0.735, black line). For the foveal measures of TD (not plotted), the correlation is only −0.258, and for 5° eccentricity measures, it is −0.525. As stated before, search time increases as the perceived differences among the distractor patches increases (DD). Rosenholtz et al. studied only arrays with homogeneous distractors (DD would be zero in theory), but analysis of our experiment must also account for differences in DD between heterogeneous conditions (Table 2B). Simple regressions of log search slope against log DD (not plotted) gave correlation coefficients of 0.418

←

nonlinear fit to Equation 4. Symbols as in part A. This figure is reproduced in the Supplementary Materials (section 3) with different symbols to highlight the "high," "low," and "metamer" classes of array.

Search array condition	(A) Search slope				(B) Intercept			
	Cat family, slope (ms/item)		Flower family, slope (ms/item)		Cat family, intercept (ms)		Flower family, intercept (ms)	
	Value	*SE*	Value	*SE*	Value	*SE*	Value	*SE*
Homogeneous								
Low	34.8	37.9	37.7	30.8	1347	409	1147	333
Metamer	150.2	31.6	147.6	33.0	915	341	722	357
High	5.6	9.3	0.5	7.6	676	100	628	82
Heterogeneous								
Low	143.4	72.4	257.2	128.6	1443	782	924	1389
Metamer	327.4	57.8	191.4	34.5	243	624	338	372
High	71.8	29.5	43.8	15.7	525	319	534	170

Table 1. Slopes (A) and intercepts (B) of the regressions of search time on set size. *Notes*: For each of the 12 types of search array (cat/flower, homogeneous/heterogenous, low/high/metamer), there were three set sizes (five, 10, 15). Each regression is based on 15 data: three set sizes and the median search times of each of the five observers. The estimated standard errors of the regression parameters are shown.

(foveal), 0.441 (5°), and 0.627 (12°). We performed multilinear fits on TD and DD together, of the form

$$\log_{10}(\text{search slope}) = k + \beta_1 \log_{10}(\text{TD}) + \beta_2 \log_{10}(\text{DD}) \quad (2)$$

For the homogeneous arrays, DD was physically zero, but observers did sometimes report perceived differences especially with peripheral viewing (see Supplementary Figure 7). We used the experimental (nonzero) measures of DD (Table 2B). For foveal measures of TD and DD, the multilinear fit improved the correlation coefficient r to 0.522. However, the fit with measures at 12° eccentricity improved to 0.955 (with 5° measures, $r = 0.721$). This correlation with 12° rating measures is significantly better than that with foveal measures ($r = 0.522$ vs. 0.955, $n = 12$, $z = 2.77$, $p = 0.0056$). Figure 7B plots the measured search slope from our experiments against the value predicted by the multilinear regression on the 12° eccentricity rating data. It is interesting that the fits to the search slopes for heterogeneous cat (blue circles) and heterogeneous flower arrays (blue triangles) are not systematically different even though the pattern of search *times* was different (Figure 5).

The multilinear regression on TD and DD at 12° eccentricity was given by

$$\log_{10}(\text{search slope}) = 3.809 - 2.256 \log_{10}(\text{TD}) + 0.882 \log_{10}(\text{DD}) \quad (2A)$$

Removing the logarithms, we might suppose that search *time* would be given by

$$\text{search time} = kk + 6442 \cdot \left[\frac{\text{DD}^{0.882}}{\text{TD}^{2.256}}\right] \cdot \text{size ms} \quad (3)$$

However, this is not the overall best fit because the initial fitting of search slopes (Table 1A, Figure 7B) will have ignored the considerable divergence among the intercepts (Table 1B). We therefore used a nonlinear iterative fit of a four–parameter formulation:

$$\text{search time} = kk + k \cdot \left[\frac{\text{DD}^{\beta_1}}{\text{TD}^{\beta_2}}\right] \cdot \text{size ms} \quad (4)$$

Using the TD and DD values at 12°, the best-fitting values ($r = 0.904$, $n = 36$) were given by

$$\text{search time} = 533 + 2222 \cdot \left[\frac{\text{DD}^{0.267}}{\text{TD}^{1.222}}\right] \cdot \text{size ms} \quad (4A)$$

The best fit of Equation 4A with foveal values of TD and DD only gave $r = 0.639$; with 5° eccentricity values, the fit gave $r = 0.823$. Figure 7C plots the measured search times from our experiments against the value predicted by the fit at 12°. The predictions for the heterogeneous flower arrays seem to be systematically too high (some blue triangles lie well below the line of equality) reflecting the difference in pattern of search times for the two families shown in Figure 5. The fit of Equation 4A should be improvable if we could model the single parameter kk as varying with the type of search array, making it some nonlinear function of foveal TD to account for the finding that "low" arrays have greater intercept (kk) than "high" or "metamer" arrays. However, we do not have enough data to allow extra fit parameters. Overall, the fits in Figure 7 quantitatively confirm the conclusion of Figure 4 that search slopes and search times are largely dependent on peripheral perceptions of the differences among the array elements: The correlation coefficients are substantially better when using the 12° peripheral ratings than when using the foveal ratings.

Search array condition	Cat family		Flower family	
	Foveal	Peripheral	Foveal	Peripheral
(A) Target–distractor distance (TD)				
Homogeneous				
Low	7.7 ± 5.5	**20.1 ± 9.5**	11.2 ± 8.0	*13.9 ± 12.1*
Metamer	**35.9 ± 18.9**	**18.5 ± 15.2**	*31.5 ± 19.7*	*10.8 ± 12.3*
High	**44.2 ± 19.9**	39.9 ± 20.7	*37.3 ± 14.3*	40.7 ± 11.6
Heterogeneous				
Low	10.0 ± 8.6	17.4 ± 11.0	14.2 ± 10.5	14.0 ± 14.0
Metamer	23.1 ± 19.3	10.1 ± 10.9	30.6 ± 20.5	9.0 ± 10.5
High	32.3 ± 14.8	35.6 ± 15.5	41.3 ± 20.6	39.3 ± 16.8
(B) Distractor–distractor distance (DD)				
Homogeneous				
Low	0.0 ± 0.0	4.3 ± 6.0	1.2 ± 3.1	3.5 ± 6.2
Metamer	0.4 ± 1.3	6.1 ± 9.7	*(1.7)*	*(2.1)*
High	0.3 ± 0.9	2.8 ± 5.1	2.1 ± 6.0	0.7 ± 1.7
Heterogeneous				
Low	19.9 ± 8.0	25.8 ± 13.2	17.4 ± 11.2	18.1 ± 14.2
Metamer	27.1 ± 18.9	13.6 ± 13.7	35.9 ± 20.6	12.3 ± 12.0
High	42.0 ± 19.5	44.8 ± 18.6	48.8 ± 23.5	43.7 ± 19.6

Table 2. Estimated average target–distractor distances (TD, A) and distractor–distractor distances (DD, B) for foveal viewing and at 12° eccentricity. *Notes*: Eleven observers participated. (A) Based on the ratings of the difference between the target and a distractor. Each of the TD estimates for homogenous search arrays is the average of the 11 observers' ratings for two different distractor images ($n = 22$). Each of the TD distances for the heterogeneous search arrays is the average of the 11 observers' ratings for seven stimuli ($n = 77$); these include the two stimuli counted for the homogeneous arrays. Standard deviations are shown. The highlighting is discussed in the text. (B) Here the ratings were measured for the perceived difference between pairs of distractor stimuli. For the heterogeneous DD distance, the values in the table are the averages for 11 observers' ratings to only 11 of the possible pairs. The homogeneous DD values in the table are based on 11 observers' ratings for one or two image pairs except for the homogeneous "metamer" cat family in which no pairs were examined; these values in brackets are the averages of the appropriate "high" and "low" ratings.

Discussion

Search time has been hypothesized to depend upon TD and DD (Duncan & Humphreys, 1989). In this study, we have attempted to quantify these differences using ratings of natural image stimuli (To et al., 2011; To et al., 2010) in both peripheral and foveal vision. We show that it is possible to create "metamer" images that are easily distinguishable in foveal vision but difficult to distinguish in peripheral vision, supporting previous work that has shown that certain types of stimulus changes are difficult to detect peripherally (To et al., 2011). We then showed that peripheral ratings of TD and DD can be used to quantitatively predict search slopes, highlighting the important role of peripheral vision in search tasks.

Using a search experiment, we found relatively slow reaction times and steep search slopes for search displays with low discriminability between target and distractors both peripherally and foveally as predicted by Duncan and Humphreys (1989). Similarly, search displays with high discriminability between target and distractors both peripherally and foveally had faster search times and shallower search slopes. However, the

key condition in our study was the "metamer" condition, in which the distractors were easy to discriminate from the target foveally but more difficult peripherally. We have found that the "metamer" condition had similar search times to the "low" condition and in fact even steeper slopes than the "low" condition, implying that the peripheral TD is more important than the foveal discriminability in determining overall search efficacy. Modeling using previously established formulations of the role of TD and DD in search (Duncan & Humphreys, 1989; Rosenholtz et al., 2012) also showed that peripheral TD and DD are better predictors of search times and slopes than foveal TD and DD. Although there have been previous studies highlighting the importance of peripheral information in search (Geisler et al., 2006; Loschky & McConkie, 2002; Rosenholtz et al., 2012), our results are the first to show this pattern for visual search using naturalistic images.

Our results suggest that the discriminability of targets and distractors in the periphery is critical in understanding search time. The fastest search times ("pop out") are seen when the TDs are high and the DDs are low (i.e., the distractors in search arrays are homogeneous). In this case, the information required to

make the discrimination does not need to be highly detailed; for example, an observer may simply need to pick out the target that matches the color of the remembered template, and this is easy to do using just peripheral information, leading to fast search times. However, for slower search times (such as when TDs are low and DDs are high), more detailed information is needed to make a judgment, which is not available with peripheral vision. This may be because the differences are too small to be distinguished (such as in the "low" condition) or because they are of the wrong type to be easily discriminated in the periphery (such as in the "metamer" condition). Observers therefore must move their eyes to each target in turn to foveate them and allow discrimination, increasing search times.

The increased search slope in the "low" and "metamer" conditions compared to the "high" condition reflects the increased difficulty of the search task for the former conditions as the set size increases. This could indicate that participants need to foveate a greater number of images to be sure of finding the correct target. Interestingly, we found that the search slope of the "metamer" condition was actually even higher than that for the "low" condition. This can perhaps be explained by the results of the ratings experiment, which showed that the rated TD peripherally was in fact slightly smaller for the "metamer" condition compared to the "low" condition (Figure 6B), suggesting that even more targets would need to be foveated and checked on average for the "metamer" stimuli in the larger set sizes, leading to increased search slopes. However, an alternative explanation for the increased search slopes for both "metamer" and "low" conditions is that, with a greater number of options, participants need to spend longer foveating individual images to be sure of whether it is the target or not. This is perhaps particularly plausible in the "low" condition, in which the distractors are all perceived as being similar to the target even when viewed foveally but may offer a less convincing explanation for the "metamer" condition, in which it should be easy to distinguish the target from the distractors once they have been foveated.

Recent work using eye-tracking methodology has considered these two possibilities in experiments using naturalistic search stimuli with which the distractors could differ from the target in zero, one, two, or three features (Alexander & Zelinsky, 2012). They found that increasing TD similarity changed both search guidance (leading to an increase in time taken to fixate the target, for example) and also led to longer foveal search decisions (e.g., giving longer target verification times), suggesting that both factors may be in play for the differences between the "low" and "high" conditions in our experiment. Similar results have also been found in studies using simple patterns as search stimuli (Becker,

Ansorge, & Horstmann, 2009). It would be interesting to extend the current study to consider whether "metamer" stimuli show different patterns with the prediction being that search guidance might be strongly slowed (and thus more similar to the "low" condition) and search decisions would be unaffected (being more similar to the "high" condition).

Although search in our "metamer" and "low" conditions share many similarities overall, it does not seem that they are treated identically by the visual system. The rating experiment, of course, shows that "metamer" stimuli are much more recognizable when they are viewed foveally compared with peripherally unlike the "low" stimuli; any differences in detail may reflect the role of foveal vision in search. For instance, observers were asked to indicate where in the display they thought the target was, and by far, the most errors were made with the "low" stimuli. In addition, we found that search intercepts are generally higher for "low" arrays. Intercepts are not generally considered when considering search efficiency (Zelinsky & Sheinberg, 1997), but our results imply that they may offer important insights in addition to search slopes into the different types of search strategy that may be used for different conditions. A higher intercept for the "low" arrays may again reflect the increased difficulty in foveal discrimination for these stimuli as they imply that even when the set size is zero (i.e., the target is on its own) it takes longer to make a judgment.

One interesting finding in our current results is that there were differences between the two types of metameric stimuli used in the experiment with the search times for the cat "metamer" array in the heterogeneous condition being longer than the "low" array, which contrasts with the flower "metamer" array, for which the search times were shorter than the corresponding "low" array. It is probably not surprising that different naturalistic images give different results in detail although the differences here are not easily explicable in terms of TD and DD in the ratings experiment. It has been shown that high-level semantic structure can affect search (Eckstein, Drescher, & Shimozaki, 2006; Henderson, Weeks, & Hollingworth, 1999; Moores, Laiti, & Chelazzi, 2003; Neider & Zelinsky, 2006), and therefore, the differences seen for the different image families in this experiment may reflect differences in how semantic information affects search processes. For example, the cat target stimulus in this experiment had an extremely clear, canonical orientation whereas the flower target stimulus had a much less obviously "correct" orientation due to its many axes of approximate symmetry. This would allow a subject to much more easily distinguish the target from, say, any rotated distractors in the case of the cat stimulus, making the "low" search condition relatively easier than for the flower stimulus. This explanation

would mean that the search results would not necessarily be expected to match those found in the ratings experiment; it is possible for participants to find it easier to see that there are differences between two flower stimuli than cat stimuli (leading to larger difference ratings for the flower stimuli) while simultaneously finding it easier to identify the true cat as the target (leading to slower search times for the flower stimuli in the "low" condition). Of course, it is difficult to generalize extensively from these results as only two initial naturalistic images were tested. Future work could use a wider range of images to test this hypothesis.

Modeling our results using solely the 12° peripheral measures of TD and DD produced good fits with r values of over 0.8. We were able to model search slope better than search time (which includes the extra high intercept in the "low" condition). We can therefore explain a great deal of the variation in the data using a relatively simple formulation of how targets and distractors are related to each other. However, it is probable that this modeling could be improved by adding further terms. We did not consider the interaction between TD and DD suggested by Duncan and Humphreys (1989) in which DD has less effect as TD increases. We also did not add extra terms for the foveal or 5° peripheral ratings. Given the high intercept in the "low" condition, perhaps search time could be better modeled by allowing kk (Equation 4) to depend somehow upon foveal values. We conducted modeling using only one peripheral eccentricity (12°) as we found that there was a high degree of correlation between the data at the two peripheral eccentricities used in the ratings experiment and even with the foveal ones. This was especially true for the DD ratings because in homogeneous arrays DD will be near zero whatever eccentricity is used and in heterogeneous arrays the ratings will all be higher (although to different extents at different eccentricities). Last, we used averages of TD and DD in each condition even though specific distractors differed between arrays and the specific differences between patches varied idiosyncratically. Although it seems that a complete model should include some of these aspects, a larger experimental data set would be required to permit more complex model fitting.

Although a strength of the results presented in this paper is the use of naturalistic images, the current visual search task differs from the natural situation in a number of important ways. One limitation is that search in real life may be less well constrained: For example, people may not know exactly what they are looking for or may be searching for a target that is not in fact contained within a scene. The current approach could be easily extended to include target-absent conditions or tasks in which participants are required to spot the "odd one out" target rather than a specified target. However, a more fundamental limitation may be that, when searching in a natural scene, it may be difficult to accurately estimate pairwise differences between target and distractor items due to the difficulty of partitioning the background into discrete distractors. However, recent research has shown that some clutter metrics may provide an alternative to set size as a determinant for visual search tasks with natural scenes (Asher et al., 2013). It may also be possible to generate contexts with different levels of peripheral similarity to a target, perhaps by using a model such as the texture tiling model (Rosenholtz et al., 2012). Future work could therefore use different levels of clutter as a way to tap into the importance of peripheral information in visual search in natural visually continuous scenes.

Conclusion

In conclusion, our results provide evidence that for visual search tasks using discrete patches of naturalistic stimuli, search times and search slopes are primarily determined by peripheral visual measures of perceived differences between targets and distractors and not by foveal difference measures. However, we also found that some aspects of search, such as intercepts and errors, may be influenced by foveal information, highlighting the interplay between peripheral and foveal vision in search tasks. Our results support previous work using simpler, geometric targets (Rosenholtz et al., 2012) and extend it to stimuli that may be more similar to those used in real-world search tasks.

Keywords: visual search, peripheral vision, natural images, search slope

Acknowledgments

*AEH and RVS contributed equally to this article.
Commercial relationships: none.
Corresponding author: Anna E. Hughes.
Email: anna.hughes@ucl.ac.uk.
Address: Division of Psychology and Language Sciences, University College London, London, UK.

References

Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision, 11*(8):9, 1–15, doi:10.1167/11.8.9. [PubMed] [Article]

Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search: The Frankenbear experiment. *Vision Research, 54*, 20–30, doi:10.1016/j.visres.2011.12.004.

Asher, M. F., Tolhurst, D. J., Troscianko, T., & Gilchrist, I. D. (2013). Regional effects of clutter on human target detection performance. *Journal of Vision, 13*(5): 25, 1–15, doi:10.1167/13.5.25. [PubMed] [Article]

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

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. Retrieved from http://arxiv.org/abs/1406.5823

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996). Visual search for colour targets that are or are not linearly separable from distractors. *Vision Research, 36*(10), 1439–1466, doi:10.1016/0042-6989(95)00207-3.

Becker, S. I., Ansorge, U., & Horstmann, G. (2009). Can intertrial priming account for the similarity effect in visual search? *Vision Research, 49*(14), 1738–1756, doi:10.1016/j.visres.2009.04.001.

Bouma, H. (1970, April 11). Interaction effects in parafoveal letter recognition. *Nature, 226*(5241), 177–178, doi:10.1038/226177a0.

Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics, 57*(8), 1241–1261.

Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 673–692.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458, doi:10.1037/0033-295X.96.3.433.

D'Zmura, M. (1991). Color in visual search. *Vision Research, 31*(6), 951–966.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17*(11), 973–980, doi:10.1111/j.1467-9280.2006.01815.x.

Foster, D. H., & Ward, P. A. (1991). Asymmetries in oriented-line detection indicate two orthogonal filters in early vision. *Proceedings of the Royal Society of London B: Biological Sciences, 243*(1306), 75–81, doi:10.1098/rspb.1991.0013.

Fox, J., & Weisberg, H. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage Publications. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14*(9), 1195–1201, doi:10.1038/nn.2889.

Geisler, W. S., Perry, J. S., & Najemnik, J. (2006). Visual search: The role of peripheral information measured using gaze-contingent displays. *Journal of Vision, 6*(9): 1, 858–873, doi:10.1167/6.9.1. [PubMed] [Article]

Gilchrist, I. D., Heywood, C. A., & Findlay, J. M. (1999). Saccade selection in visual search: Evidence for spatial frequency specific between-item interactions. *Vision Research, 39*(7), 1373–1383.

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25*(1), 210–228, doi:10.1037/0096-1523.25.1.210.

Hess, R. F., & Field, D. (1993). Is the increased spatial uncertainty in the normal periphery due to spatial undersampling or uncalibrated disarray? *Vision Research, 33*(18), 2663–2670, doi:10.1016/0042-6989(93)90226-M.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrische Zeitschrift [Biometrical Journal], 50*(3), 346–363, doi:10.1002/bimj.200810425.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics, 5*(3), 299, doi:10.2307/1390807.

Joseph, J. S., Chun, M. M., & Nakayama, K. (1997, June 19). Attentional requirements in a "preattentive" feature search task. *Nature, 387*(6635), 805–807, doi:10.1038/42940.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects oflme4 package) (Version 2.0-11). Retrieved from http://CRAN.R-project.org/package=lmerTest

Levi, D. M. (2008). Crowding – An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635–654, doi:10.1016/j.visres.2007.12.009.

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research, 25*(7), 963–977, doi:10.1016/0042-6989(85)90207-X.

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*(2), 99–117, doi:10.1037/1076-898X.8.2.99.

Lovell, P. G., Gilchrist, I. D., Tolhurst, D. J., & Troscianko, T. (2009). Search for gross illumination discrepancies in images of natural objects. *Journal of Vision, 9*(1):37, 1–14, doi:10.1167/9.1.37. [PubMed] [Article]

Macquistan, A. D. (1994). Heterogeneity effects in visual search predicted from the group scanning model. *Revue Canadienne De Psychologie Expérimentale [Canadian Journal of Experimental Psychology], 48*(4), 495–515.

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6*(2), 182–189, doi:10.1038/nn996.

Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the Optical Society of America A, Optics and Image Science, 7*(7), 1209–1217.

Najemnik, J., & Geisler, W. S. (2005, March 17). Optimal eye movement strategies in visual search. *Nature, 434*(7031), 387–391, doi:10.1038/nature03390.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8*(3):4, 1–14, doi:10.1167/8.3.4. [PubMed] [Article]

Najemnik, J., & Geisler, W. S. (2009). Simple summation rule for optimal fixation selection in visual search. *Vision Research, 49*(10), 1286–1294, doi:10.1016/j.visres.2008.12.005.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614–621, doi:10.1016/j.visres.2005.08.025.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739–744, doi:10.1038/89532.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*(10), 1129–1135, doi:10.1038/nn.2187.

Phillips, S., Takeda, Y., & Kumada, T. (2006). An inter-item similarity model unifying feature and conjunction search. *Vision Research, 46*(22), 3867–3880, doi:10.1016/j.visres.2006.06.016.

Reddi, B. A. J., Asrress, K. N., & Carpenter, R. H. S. (2003). Accuracy, information, and response time in a saccadic decision task. *Journal of Neurophysiology, 90*(5), 3538–3546, doi:10.1152/jn.00689.2002.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision, 12*(4):14, 1–17, doi:10.1167/12.4.14. [PubMed] [Article]

Southwell, R., Hughes, A., Tolhurst, D., & Gilchrist, I. (2014). Visual search speed is driven primarily by visual similarity of objects in peripheral, no foveal vision. *Perception, 43*, 1128.

To, M. P. S., Gilchrist, I. D., Troscianko, T., Kho, J. S. B., & Tolhurst, D. J. (2009). Perception of differences in natural-image stimuli: Why is peripheral viewing poorer than foveal? *ACM Transactions on Applied Perception, 6*(4), 26:1–26:9, doi:10.1145/1609967.1609973.

To, M. P. S., Gilchrist, I. D., Troscianko, T., & Tolhurst, D. J. (2011). Discrimination of natural scenes in central and peripheral vision. *Vision Research, 51*(14), 1686–1698, doi:10.1016/j.visres.2011.05.010.

To, M. P. S., Lovell, P. G., Troscianko, T., & Tolhurst, D. J. (2010). Perception of suprathreshold naturalistic changes in colored natural images. *Journal of Vision, 10*(4):12, 1–22, doi:10.1167/10.4.12. [PubMed] [Article]

Tolhurst, D., Hughes, A., Meacock, O., Gilchrist, I., & Southwell, R. (2014). Uncertainty in ratings for perceived differences in naturalistic image stimuli in peripheral vision. *Perception, 43*, 1129.

Tolhurst, D. J., & Ling, L. (1988). Magnification factors and the organization of the human striate cortex. *Human Neurobiology, 6*(4), 247–254.

Tolhurst, D. J., To, M. P. S., Chirimuuta, M., Troscianko, T., Chua, P. Y., & Lovell, P. G. (2010). Magnitude of perceived change in natural images may be linearly proportional to differences in neuronal firing rates. *Seeing and Perceiving, 23*(4), 349–372.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97–136, doi:10.1016/0010-0285(80)90005-5.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*(1), 15–48, http://doi.org/10.1037/0033-295X.95.1.15.

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 34–49.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495–501, http://doi.org/10.1038/nrn1411.

Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review, 115*(4), 787–835, doi:10.1037/a0013118.

Zelinsky, G. J., & Sheinberg, D. L. (1997). Eye movements during parallel-serial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 244–262.