Published in final edited form as:
*J Vis*. ; 12(4): . doi:10.1167/12.4.14.

# A Summary Statistic Representation in Peripheral Vision Explains Visual Search

**Ruth Rosenholtz**,
Department of Brain & Cognitive Sciences, Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

**Jie Huang**,
Department of Brain & Cognitive Sciences, MIT, Cambridge, MA, USA

**Alvin Raj**,
Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

**Benjamin J. Balas**, and
Department of Psychology, North Dakota State University, Fargo, ND, USA

**Livia Ilie**
Mathematics Department, MIT, Cambridge, MA, USA

## Abstract

Vision is an active process: we repeatedly move our eyes to seek out objects of interest and explore our environment. Visual search experiments capture aspects of this process, by having subjects look for a target within a background of distractors. Search speed often correlates with target-distractor discriminability; search is faster when the target and distractors look quite different. However, there are notable exceptions. A given discriminability can yield efficient searches (where the target seems to “pop-out”) as well as inefficient ones (where additional distractors make search significantly slower and more difficult). Search is often more difficult when finding the target requires distinguishing a particular configuration or conjunction of features. Search asymmetries abound. These puzzling results have fueled three decades of theoretical and experimental studies. We argue that the key issue in search is the processing of image patches in the periphery, where visual representation is characterized by summary statistics computed over a sizable pooling region. By quantifying these statistics, we predict a set of classic search results, as well as peripheral discriminability of crowded patches such as those found in search displays.

### Keywords

visual search; peripheral vision; crowding; texture; summary statistics



Corresponding author: Ruth Rosenholtz., rruth@mit.edu., Address: Dept. of Brain & Cognitive Sciences, CSAIL, M.I.T., Cambridge, MA 02139, USA..

## Introduction

When we search for an object, our eyes perform a series of rapid coordinated eye movements, known as saccades, terminating when the fovea reaches the target. Sometimes we notice the target even in the periphery, and it seems to “pop out”. In other cases, peripheral vision helps guide our eye movements to the target. The periphery, being much larger than the fovea, is inherently more likely to contain the target. Therefore, to understand search we must understand the strengths and limitations of peripheral vision (Erkelens & Hooge, 1996, Carrasco & Frieder, 1997; Carrasco et al, 1998; Geisler, Perry, & Najemnik, 2006).

Peripheral vision is substantially worse than foveal vision. Only a finite number of nerve fibers can emerge from the eye, and rather than providing uniformly mediocre vision, the eye trades off sparse sampling in the periphery for sharp, high resolution foveal vision. If we need finer detail (for example for reading), we move our eyes to bring the fovea to the desired location. This economical design continues into the cortex: the cortical magnification factor expresses the way in which cortical resources are concentrated in central vision at the expense of the periphery.

The phenomenon of visual crowding illustrates that the loss of information in the periphery is not merely due to reduced acuity. A string of letters like ‘BOARD’ can be quite difficult to read in the periphery. An observer might see the letters in the wrong order, perhaps confusing the word with ‘BORAD’. They might not see an ‘A’ at all, or might see strange letter-like shapes made up of a mixture of parts from several letters (Lettvin, 1976). Yet if the ‘A’ appears alone, the observer might easily identify it. This is puzzling. Why should the visual system retain the spatial details needed to perceive letters or parts thereof, but neglect to encode the information needed to keep track of the locations of those details?

To make sense of this curious behavior, consider the following: given that peripheral vision involves a loss of information, what information should be retained? Imagine representing a patch in the periphery by a finite set of numbers. These numbers could be the firing rates of a finite set of neurons or some other low-dimensional representation. More concretely, suppose that we wanted to represent the patch in Figure 1a with just 1000 numbers. We could coarsely subsample this patch down to a 32×32 array of pixel values, using standard filtering and sampling techniques. This is akin to peripheral subsampling in the retina, and leads to a representation like Figure 1b. Another option would be to convert Figure 1a to a wavelet-like representation like that in early visual cortex (V1) – local orientation at multiple scales– and then select the most useful 1000 coefficients. Essentially, if each coefficient corresponds to a potential “neuron”, then one can think of choosing the 1000 neurons with the highest expected firing rates. This leads to a representation like that in Figure 1c. Both of these strategies discard the high spatial frequencies, which makes it impossible to tell much about the resulting blobs other than their locations.

If one were searching for a target letter among an array of distractors, clearly it would be non-ideal to represent the location of the items but throw away virtually all details about their appearance. Perhaps we can instead capture more of the details, if we are willing to

sacrifice the exact location of those details. Portilla & Simoncelli (2000) demonstrated, with their texture synthesis algorithm, that the visual appearance of many textures could be captured with about 700 numbers. These numbers characterize summary statistics such as the marginal distribution of luminance; luminance autocorrelation; correlations of the magnitude of responses of oriented, multi-scale wavelets (similar to those found in primary visual cortex), across differences in orientation, neighboring positions, and scale; and phase correlation across scale. This perhaps sounds complicated, but really it is not. As Freeman & Simoncelli (2011) have articulated, each correlation can be computed by multiplying the outputs of appropriate pairs of V1-like cells, followed by an averaging or "pooling" operation. Perhaps the visual system could use these summary statistics to represent not only texture, but more generally to encode useful information about an arbitrary local image patch. Figure 1d shows a sample of "texture" synthesized to have approximately the same summary statistics as found in Figure 1a. The results are intriguing. Apparently, in order to match the statistics of Figure 1a, a patch needs to contain an evenly spaced array of letter-like objects. The exact details and locations are somewhat jumbled, but the model captures the appearance of the original in important ways. The properties of this representation are reminiscent of those of peripheral vision, particularly as exemplified by the phenomena of visual crowding.

Indeed, recent research on crowding has suggested that the representation in peripheral vision consists of summary statistics computed over local pooling regions (Balas, Nakano, & Rosenholtz, 2009; Parkes et al, 2001; Pelli & Tillman, 2008; Levi, 2008). In particular, discriminability based on the above set of summary statistics has been shown to predict performance recognizing crowded letters in the periphery (Balas, Nakano, & Rosenholtz, 2009). These summary statistics also perform favorably at capturing texture appearance, when compared with several alternative sets of statistics (Portilla & Simoncelli, 2000; Balas, 2006). Freeman & Simoncelli (2011) have shown that observers have difficulty discriminating between a natural scene and an image synthesized to match its local summary statistics, in brief presentation. (See our FAQ at http://persci.mit.edu/mongrels/about.html for more discussion of why these statistics are a good first guess for our model.)

Does it make sense for peripheral vision to retain statistical information about a pattern's appearance, while losing the arrangement of the pattern elements? Consider the different roles played by foveal and peripheral vision: Foveal vision contains powerful machinery for object recognition, but covers a tiny fraction of the visual field. A major role of peripheral vision, by comparison, is to monitor a much wider area, looking for regions that appear interesting or informative, in order to plan eye movements. In visual search in particular, the task is to look for a target, say, the letter O. At each instant, the subject must quickly survey the entire visual field, seeking out regions worthy of further examination. If the informational bottleneck (Nakayama, 1990) has reduced everything to fuzzy blobs (Figures 1b and 1c), then there is no way to choose among the blobs. However, if one at least knows the general location of O-like stuff, then an eye movement can be launched in a promising direction and the search process can proceed.

This view of peripheral vision should have a significant impact on our understanding of visual search. To put this in context, consider some classic puzzles in the search literature, as

well as previous attempts to explain them. A reasonable intuition, when studying search, might be that it would be easier to find a target if it was visually dissimilar from the distractors. This intuition holds true in a number of cases (Duncan & Humphreys, 1989; Palmer, Verghese, & Pavel, 2000): e.g. it is easier to find an O among X's than to find an O among Q's. However, much of the theoretical and empirical research on visual search has been driven by phenomena in which confusability of individual items fails to predict search difficulty. First, performance varies substantially across conditions for which discriminability of single items is trivial. For example, it is easy to tell a red O from a red X, and easy to tell a red O from a green O. However, it is hard to search for a red O amidst an array of red X's and green O's. Mixing the distractors (producing a "feature conjunction" task) makes search difficult (Treisman & Gelade, 1980; although see Eckstein, 1998, for the suggestion that the increase in difficulty in conjunction search does result from target-distractor discriminability). Similarly, looking for a randomly oriented T among randomly oriented L's is difficult (Wolfe, Cave, & Franzel, 1989), even though it is trivial to discriminate between a T and an L. It seems that search is difficult when the target and distractors differ only in the configuration of their parts – here, arguably, the horizontal and vertical bars that make up both the T and the L's. Second of all, the sheer prevalence of search asymmetries (Treisman & Souther, 1985; Treisman & Gormican, 1988; Wolfe, 2001) argues against search being governed by confusability of target and distractors. By definition, the confusability between O and Q is the same as that between Q and O. Therefore one might expect that finding an O among Q's should be just as hard as finding a Q among O's. This prediction fails completely (Treisman & Souther, 1985). Furthermore, factors unrelated to target-distractor similarity significantly impact search difficulty: The spacing between display items matters (Wertheim, Hooge, & Krikke et al., 2006), and item heterogeneity and/or the extent to which items group matters (Duncan & Humphreys, 1989; Verghese & Nakayama, 1994; Rosenholtz, 2001a).

Some of the puzzles described above led to the development of Feature Integration Theory (FIT) (Treisman & Gelade, 1980), and later Guided Search (Wolfe, 1994). FIT postulates that search should be easy if the target and distractors are sufficiently distinct along a single basic feature dimension, which may be used to "guide" attention. On the other hand, search tasks with targets defined by two or more attributes are more difficult, as they require that one serially deploy visual attention to individual items, in order to "bind" their features for recognition.

Since FIT strongly relies on the choice of features, the field has spent many years trying to enumerate the "basic features", and researchers have tended to adjust the list of basic features with each new search result (Rosenholtz, 2001b). The results have been puzzling, as many of these features have been quite low-level, (e.g. color and orientation), whereas other work has suggested that higher-level features such as 3D shape, lighting direction, (e.g., Enns & Rensink, 1990) or apparent reflectance (Sun & Perona, 1996) might be basic features. More generally, a number of researchers have shown results that are inconsistent with FIT (e.g. Wolfe, 1994; Carrasco et al, 1995; Carrasco et al, 1998; Vlaskamp et al, 2005; Wertheim et al, 2006; Reddy & VanRullen, 2007) or have offered alternative explanations for a number of the basic search phenomena (e.g. Duncan & Humphreys, 1989; Geisler & Chou, 1995; Palmer et al, 1993; Verghese & Nakayama, 1994; Eckstein, 1998; Palmer et al,

2000; Gheri et al, 2007; Rosenholtz et al, 2012). Nonetheless, the view persists that serial search is mediated by rapid attentional shifts, necessary to bind features of individual items.

Ultimately, FIT is, as its name implies, a theoretical framework rather than a model. FIT can describe results from previous experiments, but cannot easily generate predictions for a new search task, in which target and distractors differ along a previously untested dimension. The field lacks a clear model that can predict search difficulty given the target and distractors, plus the layout of the display. What we propose here is a testable model that meets this criterion and is applicable to arbitrary search displays.

We argue for an alternative both to search being governed by target-distractor similarity, and to FIT. We base this argument upon the insight that search is constrained by peripheral vision, and on evidence from visual crowding that in the periphery, the visual system represents each local patch by a fixed set of summary statistics (Parkes et al., 2001; Levi, 2008; Pelli & Tillman, 2008; Balas, Nakano, & Rosenholtz, 2009). These local patches can be quite large, as the literature on crowding suggests that they grow approximately linearly with eccentricity – i.e. with distance to the center of fixation – and that flankers interfere with perception of a target if they lie within a radius of about 0.4 to 0.5 × the eccentricity (Bouma, 1970). Such patches will, for typical search displays, often contain more than one item. This suggests that rather than thinking about the similarity between a single target and a single distractor, we should be thinking about the similarity between peripheral patches containing a target (plus distractors) and those containing only multiple distractors. After all, that is the visual system's real task as it confronts a search display, as illustrated in Figure 2a.

In Figure 2a, the target (Q) is not visible near the current fixation (red crosshairs), so the subject continues searching. Where to look next? A reasonable strategy is to seek out regions that have promising statistics. The green and blue discs represent two hypothetical pooling regions in the periphery, one containing the target (plus distractors), the other containing only distractors. If the statistics in a target patch are noticeably different from those of non-target patches, then this can guide the subject's eyes toward the target. However, if the statistics are inadequate to make the distinction, then the subject must proceed without guidance.

Our prediction is that to a first approximation, search will be easy if and only if the visual statistics of target patches are very different from those of non-target patches.[1] Essentially, we would like to know how close the target patches are to the non-target patches in some statistical appearance space. In principle, we could build a fully specified metric model of appearance space. This would involve two parts. First, we would specify the information being lost in peripheral vision, and second we would estimate the capabilities of human pattern recognition at utilizing what information is retained.

[1]Our model hypothesizes that only a particular set of summary statistics get through the bottleneck of vision. We suggest that a major factor in search performance arises from whether this available information allows discrimination of target+distractor from distractor-only patches. However, in certain cases, differences in search performance may also arise from processing differences at later stages of vision. For instance, peripheral recognition of letters may be better than that of non-letter symbols due to much more training on the over-learned letters, rather than due to there being more information available at early stages for the letter task.

For the first step of specifying the information encoded by peripheral vision, we have developed the *Texture Tiling Model* (Balas et al, 2009; Rosenholtz, 2011; Rosenholtz et al, 2012). In this model, the visual system tiles the visual field with overlapping pooling regions, which grow linearly with eccentricity. Within each pooling region, the visual system represents its input via the rich set of summary statistics enumerated above.

How useful are these statistics for distinguishing between target and non-target patches in peripheral vision? Unfortunately, there are presently no reliable algorithms for mimicking human pattern recognition in general. Nor can one justify running a machine classifier to find the discriminability of the corresponding statistical "vectors." Human observers have unknown uncertainty in measuring the summary statistics; adding noise to mimic this uncertainty would at this point be like fitting approximately 1000 parameters using only a handful of data points. Therefore we "simulate" human pattern recognition by using actual human observers. To do this, we first visualize the loss of information in peripheral vision by using texture synthesis (Portilla & Simoncelli, 2000) to generate a set of images with approximately the same summary statistics as the original target and non-target image patches; we call these synthesized images "mongrels" (Balas, Nakano, & Rosenholtz, 2009). By asking human observers to discriminate between target+distractor and distractor-only mongrels, we can obtain a measure of how inherently discriminable target patches are from non-target patches, on the basis of their summary statistics (Figure 2b). To test our hypothesis, we can then examine whether this "mongrel discriminability" predicts search performance. (Details and further intuitions are given in Experiment 1, Methods.)

Previous evidence supports the idea that peripheral vision constrains the difficulty of visual search. For instance, set size effects increase with eccentricity (Carrasco et al, 1995; Carrasco & Yeshurun, 1998). This effect is not merely due to slower deployment of attention to more distant eccentricities, as scaling targets and distractors by a cortical magnification factor eliminates the eccentricity effect (Carrasco & Frieder, 1997; Carrasco, McLean, Katz, & Frieder, 1998). Scaling with eccentricity also reduces set size effects, further suggesting that the limits of peripheral vision are at work (Carrasco & Frieder, 1997, Carrasco, McLean, Katz, & Frieder, 1998). That scaling display elements affects search is consistent with the theory proposed here. The approximately linear relationship between pooling region size and eccentricity (as well as between acuity and eccentricity) leads to the property of scale invariance (van Doorn, Koenderink, & Bouman, 1972). The amount of information available in the visual field remains roughly constant as one moves closer or farther away, except near the fovea, where the linear relationship breaks down.

Further evidence for the governing of search by the limits of peripheral vision comes from Najemnik & Geisler (2005, 2008, 2009). They measure the detectability of a sine wave grating in noise, at a number of peripheral locations. This detectability, along with either a Bayesian ideal searcher or more biologically plausible heuristic (Najemnik & Geisler, 2009) is predictive of both the number and pattern of fixations during search.

Other researchers have specifically identified peripheral crowding as a limiting factor on visual search. Vlaskamp, Over, & Hooge (2005) found that the addition of flankers, particularly similar flankers, slowed search for a target letter placed on a strip. Wertheim et

al. (2006) asked observers to rate the degree of what they called “lateral masking” (terminology that has frequently been used interchangeably with “crowding” in the literature) for conjunction vs. disjunction displays as well as for Q among O search, and showed that these subjective judgments predicted search difficulty. A recent study by Gheri et al (2007) showed that an objective measure of the degree of crowding was predictive of search efficiency for a set of feature search conditions. However, their metric of crowding – elevation of orientation thresholds when flankers are present vs. absent – lacks a clear connection to the visual system’s task in visual search. These studies all suggest an intriguing linkage between crowding and search, but are limited in their ability to predict search performance across a wide range of conditions by the lack of a measure of the degree of crowding that is flexible, reliable, and has a clear connection to the search task. Other evidence for the association between crowding and search comes from Geisler & Chou (1995). Though they were not explicitly interested in the effects of crowding, they showed that peripheral discrimination of cued, individual items within crowded search displays correlated with search performance.

In the current study, we go beyond the suggestion that search is limited by peripheral vision. We suggest a specific peripheral representation, and describe how that representation constrains search. Our methodology allows us to sample image patches from arbitrary search displays, and obtain a measure of the statistical discriminability of target from non-target patches. We hypothesize that the discriminability of target-present vs. target-absent mongrels predicts whether a given search task will be easy or hard. We test this hypothesis on a set of classic search experiments, examining phenomena that have historically been used as critical tests of models of visual search.

As mentioned above, the Texture Tiling Model hypothesizes pooling regions which grow with eccentricity. This component, along with the spatial inhomogeneity of many search displays, means that pooling regions vary significantly in their contents; particularly, in the number and density of elements within a given pooling region. We will begin by ignoring this potential source of complexity, and then reintroduce it later. In Experiment 1, we test the feasibility of our model by measuring the mongrel discriminability of crowded target +distractor vs. distractor-only patches, where the patches all contain the same number and average density of items. We find that this mongrel discriminability is predictive of search difficulty. In Experiment 2, we first examine how peripheral discriminability of target +distractor from distractor-only patches varies with the number of items in the patch. We find that the answer depends strongly on the choice of target and distractor. We then measure the mongrel discriminability of patches as a function of the number of items in the patch, and show that this is predictive of the peripheral discriminability, as well as of search performance.

## Experiment 1: Can mongrel discriminability predict search difficulty?

In Experiment 1, we carried out five classic search tasks, and quantified search difficulty in each task. Though results for these tasks already exist in the literature, we re-ran all five, to ensure consistent display conditions and to use the same subjects for all. In parallel, we had subjects perform a second task, in which they discriminated between images with the same

summary statistics as target+distractor patches, and images with the same summary statistics as distractor -only patches. We show that the latter mongrel discriminability between crowded target+distractor and distractor-only patches is predictive of performance on the five search tasks.

## Methods, Search Task

**Subjects—**Ten subjects (6 male) participated in the search experiment after giving written informed consent. Subjects' ages ranged from 18-40 years. All subjects reported normal or corrected-to-normal vision, and received monetary compensation for their participation.

**Stimuli and procedure—**Our visual search experiments resemble classic search experiments in the literature. We tested five classic search conditions: Conjunction search (targets defined by the conjunction of luminance contrast and orientation), search for T among Ls, search for O among Qs, search for Q among Os, and feature search for a tilted line among vertical lines. Examples of targets and distractors can be seen in Figure 4's leftmost two columns.

Stimuli were presented on a 40cm × 28 cm monitor, with subjects seated 75 cm away in a dark room. We ran our experiments in MATLAB, using the Psychophysics Toolbox (Brainard, 1997). Eye movements were recorded at 240 Hz using an ISCAN RK-464 video-based eyetracker for the purposes of later quantitative modeling of the number of fixations to find the target. The search displays consisted of a number of items (the "set size"), consisting of either all distractors (target absent trial) or one target and the rest distractors (target present trial). Target-present and target-absent displays occurred with equal probability.

Each search task had four set size levels: 1, 6, 12, or 18 total items. Stimuli were randomly placed on 4 concentric circles, with added positional jitter (up to 1/8th degree). The radii of the circles were 4, 5.5, 7, and 8.5 degrees of visual angle (v.a.) at a viewing distance of 75 cm. Each stimulus (target or distractor) subtended approximately 1 degree v.a. Example target-present stimuli for two of the conditions, Q among Os and T among Ls, are shown in Figure 3a, for set size=18.

On each trial, the search display was presented on the computer screen until subjects responded. Subjects indicated with a key press whether each stimulus contained or did not contain a target, and were given auditory feedback. If the subject made an error, another trial was added to the block to replace that trial. Thus, each subject finished 144 correct trials for each search condition (72 target-present and 72 target-absent), evenly distributed across four set sizes. The order of the search conditions was counterbalanced across subjects, and blocked by set size. (While blocking by set size is less standard than randomizing, one could argue that blocking is a better strategy as it allows one more control over what the observer knows about the stimulus and task. In comparing our results to those in the literature, there seems to be little effect of blocking on the conclusions drawn in this paper.)

## Methods, Mongrel task: Discriminating target+distractor from distractor-only patches using summary statistics

**Subjects—**The discrimination task was carried out by five subjects (4 male). Subjects' ages ranged from 18-45 years. All reported normal or corrected-to normal vision and were paid for their participation.

**Stimuli and procedure—**To measure the discriminability between target+distractor and distractor-only patches using only summary statistics, we used the following methodology. First, we generated 10 unique distractor-only and 10 unique target+distractor patches for the 5 visual search conditions described above (Figure 4, columns 1 & 2).

For each patch, we then synthesized 10 new image patches that closely match the same summary statistics as the original patch (Figure 4, last 4 columns), using Portilla & Simoncelli's (2000) texture synthesis algorithm with 4 scales, 4 orientations, and a neighborhood size of 9. (This choice of parameter settings leads to measurement of approximately 1000 summary statistics per patch.) This algorithm first measures a set of wavelet-based features at multiple spatial scales, then computes a number of summary statistics, including joint statistics that describe local relative orientation, relative phase, and wavelet correlations across position and scale. To synthesize a new texture, the algorithm then iteratively adjusts an initial "seed" image (often, as in this experiment, white noise, but any starting image may be used) until it has approximately the same statistics as the original image patch. The resulting texture synthesized patch, which we call a "mongrel," is nearly equivalent to the original input in terms of the summary statistics measured by the model.

These mongrels essentially give us a visualization of the information encoded in approximately 1000 local summary statistics. Without mongrels, getting intuitions about the available information is quite difficult. In addition, mongrels enable a methodology for testing our model. The general logic is this: we can generate a number of mongrels which share the same local summary statistics as each original stimulus. The model cannot tell these mongrels apart from the original, nor from each other. If these images are confusable, as the model suggests, how hard would a given task be? For example, consider the mongrels in Figure 4c. The two on the left contain the same summary statistics as an image with one T and 5 L's. The two on the right share summary statistics with an image with 6 L's. Yet it is difficult to distinguish between the two pairs. This suggests that if our model is correct, it would be quite difficult to find a T among L's. In fact, quite powerfully, we can gain intuitions about the impact of our model on higher-level visual tasks without needing a model of higher-level vision. We don't need to build an "all-Q's" vs. "Q's plus O" discriminator; a simple visual inspection of mongrels like Figure 4d suggests that our model predicts this task will be difficult.

The value of such intuitions should not be underestimated. However, it is desirable to formalize them into more quantitative predictions. We do this by replacing the subjective assessment of mongrels with a discrimination task. First, we synthesize a number of mongrels. Then, for the case of search, we ask observers to discriminate between mongrels from target-present patches, and mongrels from target-absent patches. We use the mongrel discrimination performance as a measure of the informativeness of the summary statistics

for the given task. This procedure allows us to generate testable model predictions for a wide range of tasks. We have previously used this methodology to study a number of visual crowding tasks, and shown that the model can predict performance on those tasks (Balas, Nakano & Rosenholtz, 2009). (See our FAQ, http://persci.mit.edu/mongrels/about.html, for more discussion of why we favor this methodology.)

During each trial, a mongrel was presented at the center of the computer screen until subjects made a response. Each mongrel subtended 3.8 × 3.8 degrees v.a. at a viewing distance of 75 cm. Subjects were asked to categorize each mongrel according to whether or not they believed a target was present in the original patch. We wish to determine the inherent difficulty discriminating target+distractor from distractor-only patches using summary statistics, and therefore chose to optimize observer performance at this task: Subjects had unlimited time to freely view the mongrels. Observers viewed the mongrels at increased contrast, as in Figure 4. Contrast was increased using the Adobe® Photoshop® contrast adjustment, with a setting of 100. This keeps the Michelson contrast approximately the same, but increases RMS contrast by 8-22%, depending upon the content of the original image. Without increased contrast, pilot subjects reported confusion as to whether they should make use of lower contrast -but quite visible – image features (e.g. the “bar” on a “Q” often appears lower contrast in mongrels than the “O” part). In the pilot experiment, this adjustment reduced training time for naïve subjects, without affecting asymptotic performance. See Figures 4 and 8 for contrast enhanced vs. non-enhanced mongrels.

Each of the five conditions (corresponding to one of our search tasks) had a total of 100 target+distractor and 100 distractor-only patches to be discriminated in this mongrel task, with the first 30 trials (15 target+distractor and 15 distractor-only) serving as training, to familiarize observers with the nature of the stimuli. Observers received auditory feedback about the correctness of their responses throughout the experiment.

Here, then, is our operational hypothesis: (1) that given a fixed set of summary statistics, one can generate mongrels with approximately the same statistics as each given image patch; (2) that we let observers view those mongrels however they like (freely moving their eyes, for as long as they wish), so as to derive as much information from them as they can; and (3) that the discriminability of target-present from distractor-only mongrels will predict search performance.

## Results

**Search difficulty—**As is standard in the search literature, we quantify search difficulty as the slope of the best-fit line relating mean reaction time (RT) to the number of items in the display. Only correct trials were included in this analysis. Figure 3b plots the mean reaction time against set size of search display, along with the best linear fit. The results are shown in the legend of Figure 3b. These results are consistent with previously reported search studies.

**Mongrel discriminability—**Before presenting the quantitative results of our experiments, it is worth examining the mongrels in more detail. Since the set of mongrels corresponding to a given stimulus share the same statistical representation, viewing them allows us to see the confusions and ambiguities inherent in the representation.

Consider "feature search" for a tilted line among vertical (Figure 4a), a task known to be easy (Treisman & Gelade, 1980; Treisman & Schmidt, 1982; Kanwisher, 1991). The target +distractor mongrels for this condition clearly show a target-like item, whereas the distractor-only mongrels do not. Patch discrimination based upon statistics alone should be easy, predicting easy search. We predict that this task should be possible in the periphery, without moving one's eyes.

Conjunction search for a white vertical among black verticals and white horizontals shows some intriguing "illusory conjunctions" (Treisman & Gelade, 1980; Treisman & Schmidt, 1982) – white verticals and black horizontals (Figure 4b). This makes the patch discrimination task more difficult than for our feature search example, and correctly predicts more difficult search. Pelli et al (2004) have previously suggested that many of the illusory conjunction results in the literature may actually be due to crowding rather than to a lack of "focal attention". That our model of crowding (Balas et al, 2009) predicts such illusory conjunctions provides additional evidence for this claim.

Search for a T among L's (Figure 4c) is known as a difficult "configuration search" (Wolfe, Cave, & Franzel, 1989). In fact, the mongrels for this condition show T-like items in some of the distractor-only patches, and no T-like items in some of the target+distractor mongrels. Patch discrimination based upon statistics looks difficult, predicting difficult search. Similarly, search for O among Q appears quite difficult according to the statistical representation (Figure 4d), but mongrels for a Q among O search (Figure 4e) contain evidence of Q's in the target+distractor mongrels, predicting an easier search (Treisman & Souther, 1985).

Performance of the mongrel task in each condition was described by discriminability, *d'*, computed in the standard way: Using the correct identification of a target+distractors mongrel as a Hit and the incorrect labeling of a distractoronly mongrel as a False Alarm,

$$d' = z(\text{Hit rate}) - z(\text{False Alarm rate})$$

where *z(p)* indicates the z-score corresponding to proportion *p*. This measure of the discriminability of the mongrel images gives us a measure of the discriminability of target +distractor from distractor-only patches based on their summary statistics, here referred to as the mongrel discriminability.

Our model proposes that to a first approximation, discriminability based on summary statistics should predict whether a given search task is difficult or not. Specifically, when distractor-only patches have summary statistics which are not easily discriminable from target+distractor ones, the corresponding search task should be difficult, and vice versa. To examine our model's prediction, we carried out correlation analysis between each task's search reaction time slope and corresponding mongrel discriminability. Figure 5a plots $\log_{10}$(search slope) on these 5 tasks versus $\log_{10}(d')$ from our mongrel experiment. The data shows a clear relationship between search performance and the mongrel discriminability of target+distractor from distractor-only patches ($R^2 = .99$, $p < 0.001$). The significant relationship echoes the insights gleaned from viewing the mongrels, and agrees with our

predictions. When it is difficult to discriminate between target+distractor patch statistics and distractor-only statistics, search is inefficient; when the statistics are easy to discriminate, search is efficient.

The results of Experiment 1 demonstrate the feasibility of thinking of visual search in terms of summary statistic representation in peripheral vision. However, Experiment 1 was a test of a simplified version of the model, in which all patches contained the same number of items, with roughly the same spacing, and so on. In Experiment 2, we examine the discriminability of more realistically variable patches.

## Experiment 2: Peripheral and mongrel discriminability of patches of varying numerosity

In the full version of the Texture Tiling model, the visual input is tiled with overlapping pooling regions whose size grows with distance from the point of fixation (patch eccentricity). As a result, the size and contents of each pooling region in a search display depend upon the exact display and the current fixation, and may vary considerably from patch to patch. It is likely that peripheral discriminability varies depending on the contents of a patch as well.

Previous studies have demonstrated that visual search performance can be substantially influenced by factors like heterogeneity among distractors (Nagy & Thomas, 2003), and the number of distractors and their spacing around the target (Wertheim et al., 2006; Reddy & VanRullen, 2007). Peripheral discriminability decreases as the number of items in the peripheral patch goes up (Levi, 2009; Põder & Wagemans, 2007), though the magnitude of this effect likely depends upon the task as well as the contents of the patch.

In Experiment 2, we mimic some of the variability in patches of a search display by measuring the peripheral discriminability of target+distractor from distractor-only patches as a function of the number of items in each patch (the *numerosity*). We then measure the mongrel discriminability of patches of varying numerosity, and show that it is predictive of peripheral discriminability, as well as of search.

### Methods

**Subjects—**A new group of five subjects (four male) participated in this task. Subjects' ages ranged from 20-26 years (Mean = 22.8). All subjects provided written informed consent, and were paid for their participation. Subjects took part in both the peripheral discriminability and mongrel tasks.

**Stimuli and procedures, peripheral discrimination task—**We ran a peripheral discrimination task corresponding to each of the five search conditions in Experiment 1. On each trial, a peripheral patch appeared on the computer display at an eccentricity of 12 degrees v.a. Each target or distractor subtended approximately 1 degree v.a. If present, the target was always located at the center of the peripheral patch, and we instructed subjects to that effect. This task requires peripheral discrimination, and not visual search. The goal was

to measure the effect of crowding upon difficulty distinguishing between target+distractor and distractor-only peripheral patches.

The surrounding distractors appeared evenly spaced on a notional circle of radius 2.8 degrees v.a from the target. Thus the distractors lay within the critical spacing of visual crowding (Bouma, 1970), and within the range of distances between neighboring items in the search displays of Experiment 1. In addition, each peripheral patch was further flanked by a hexagonal array of asterisks (*), evenly spaced on a notional circle of radius 7 degrees v.a. from the target. The circle of asterisks helps localize the center of the patch, i.e. the location of the possible target.

For each condition, we ran four different numerosity levels, with the peripheral patch containing 2, 3, 4 or 5 items. That is, on target present trials, a target was flanked by 1, 2, 3 or 4 distractors; on target absent trials, the center of the peripheral patch was occupied by a distractor, with the rest kept the same as target present trials. Examples of peripheral patches in the crowding task with varying densities are shown in Figure 6.

Subjects were asked to maintain their fixations on the central fixation point of the screen throughout each trial, and we informally monitored their eye movements, using a video camera, to confirm that they complied. The observers had no prior knowledge about on which side (left or right) the patch would appear, as this varied randomly from trial to trial. This, plus the 180 msec presentation time, discouraged eye movements toward the target. Brief presentation time has been commonly used in tasks involving peripheral vision (e.g. Carrasco, Evert, Chang & Katz, 1995; Carrasco & Frieder, 1997; Carrasco, McLean, Katz & Frieder, 1998; Balas, Nakano & Rosenholtz, 2009). The task was to respond if a target was present in the center of the peripheral patch or not.

Each subject completed 100 trials for each numerosity level in each condition, half of which were target present. The first 20 trials among the 100 trials were treated as practice and were excluded from data analysis. Trials were blocked by condition, with the order of the five conditions randomized for each subject.

**Stimuli and procedures, mongrel task—**How does the mongrel discriminability of patches vary with the number of items in the patch, and is this mongrel discriminability predictive of actual peripheral discriminability of the crowded patches? To answer these questions, we also conducted a set of mongrel tasks corresponding to the peripheral tasks described above. For these tasks, we generate mongrels of patches sampled from the actual search displays in Experiment 1. For each of the five search conditions tested in Experiment 1, we randomly sampled patches with a size of 6.4 × 6.4 degrees v.a., centered at an eccentricity of 7 degrees v.a. (the third concentric circle in Experiment 1), as in Figure 7. Although studies have found spatial pooling regions to be elongated radially and more elliptical in shape (Toet, & Levi, 1992), these square patches still closely capture the nature of the pooling region contents and are more convenient computationally.

The patches for the mongrel task, unlike the displays for the peripheral task, contained no asterisks. The asterisks in the peripheral task served to help localize the possible target

location, but were placed quite far from the target. A Bouma's Law-sized pooling region centered on the target would not include any asterisks, just as the search displays would not.

We constrained the numerosity of patches to 2, 3, 4, or 5 items, (Figure 8, columns 1 & 2) and no patches contained partial items, as in the peripheral task. For distractor-only patches, all items in the patch were distractors; for target+distractor patches, one item in the patch was a target and the rest were distractors.

We extracted 10 distractor-only and 10 target+distractor patches for each numerosity level in each search condition. We blurred these patches in order to mimic the acuity loss for a Landolt C at 7 degrees eccentricity (Rodieck, 1998). (Gaussian blur with $\Sigma \approx 0.04$ degrees v.a. To mimic presentation at, say, 12 degrees eccentricity, one would instead blur with $\Sigma \approx 0.06$ degrees v.a. Both of these blurs are inconsequential for the stimuli in our experiments; acuity is not the limiting factor.) For each patch, we synthesized 10 mongrels that share approximately the same summary statistics as the original patch, as described for the mongrel task in Experiment 1. The two rightmost columns of Figure 8 show example Q among O mongrels. The original patches in this experiment precisely capture the item arrangement, crowdedness, and the ratio of item size to patch size in our search displays. The resulting mongrels are hence less dense than the mongrels in Experiment 1. Aside from this difference in density, the mongrels in Experiments 1 and 2 look qualitatively similar. The mongrels viewed by the observers were increased in contrast as in the previous experiment; Figure 8 shows mongrels prior to contrast enhancement.

During the mongrel task, stimuli were presented on a 51 cm × 32 cm monitor, with subjects seated 61 cm away. Each mongrel subtended 6.4 × 6.4 degrees (v.a.). Subjects free-viewed the mongrels for unlimited time, and were asked to respond if a target was present in the original patch. Note that even though the peripheral discriminability task uses limited viewing time – largely to limit eye movements – we continue to use unlimited viewing time for the mongrel task. This is because the purpose of the mongrel task is to determine how informative our model's summary statistics are for a given task – what is the best one can expect to do, given the hypothesized available information. Therefore, we allow free-viewing to optimize observer performance. If, in the future, one wanted to explicitly model the effect of peripheral viewing time on performance, this would best be done by making the information available in the model a function of peripheral viewing time, rather than through limiting the viewing time of mongrels.

For each of the five search conditions, each numerosity level (2, 3, 4, or 5 items in patch) had 100 distractor-only and 100 target+distractor mongrels to be discriminated, with the first 30 trials (15 target+distractor and 15 distractor-only) serving as practice and excluded from analysis. The trials were blocked by condition, with the order of five conditions randomized for each subject. All other experimental procedures and parameters were the same as those used in the mongrel task of Experiment 1.

## Results

**Peripheral discrimination—**Performance in the peripheral discrimination task was measured as peripheral discriminability, *d'*. We first examined how patch numerosity affects

peripheral *d'*. We computed *d'* separately for each numerosity level in each of the five conditions, Figure 9a presents the peripheral *d'* in each of the five conditions. Consistent with previous report (Levi, 2009; Põder & Wagemans, 2007), we found that patch numerosity does have an impact on peripheral discriminability. Except for the easiest condition, Tilted among Vertical, all the other four conditions demonstrated a significant linear decrease in *d'*: Conjunction search, $F(1, 4) = 7.01$, $p < .05$; T among Ls, $F(1, 4) = 100.18$, $p < .001$; O among Qs, $F(1, 4) = 21.77$, $p = .01$; and Q among Os, $F(1, 4) = 26.46$, $p = .007$.

It is worth noting that in the two most difficult conditions (T among L, and O among Q), peripheral discriminability decreased dramatically as numerosity increased from 2 to 3, followed by a fairly flat slope for numerosities 3 through 5. A numerosity of 2 may be a special case in our peripheral discriminability experiment: one may place a Bouma's Law-sized pooling such that it includes the central array item (potentially the target), and no other items other than perhaps the flanking asterisks (see Figure 6, panel 1). For numerosity=2, the peripheral *d'* measured in Experiment 2 might overestimate the peripheral discriminability during visual search, when other search items may be present outside the patch.

**Mongrel discriminability—**Performance in each mongrel task allows us to measure a mongrel *d'*. Figure 9b plots mongrel *d'* results in the five conditions against patch numerosity. Overall, mongrel discriminability resembles the pattern of peripheral discriminability (Figure 9a). Figure 10 plots peripheral *d'* vs. mongrel *d'*. Performance was quite similar in the two tasks, and highly correlated ($R^2 = 0.84$, $p < 0.001$). The information available in our proposed summary statistic representation is quite predictive of peripheral patch discriminability.

Mongrel discriminability overall shows a decreasing trend as the number of items in a patch increases. Repeated-measures ANOVA showed a significant effect of numerosity, $F(3, 12) = 26.81$, $p < 0.001$. In addition, we find a significant two-way interaction of numerosity × condition, $F(12, 48) = 9.40$, $p < 0.001$. Specifically, the three most difficult search conditions, T among L, O among Q and conjunction search, all show a significant downward trend as a function of numerosity (conjunction search, $F(1, 4) = 62.26$, $p < 0.001$; T among L, $F(1, 4) = 16.03$, $p < 0.001$; O among Q, $F(1, 4) = 13.38$, $p < 0.001$). Conjunction search showed the most dramatic decrease in *d'* with patch numerosity. Examination of the mongrels for this condition suggests for higher numerosity patches, mongrels more likely contain illusory conjunctions, and thus subjects more often committed false alarms on distractor-only mongrels.

We observed no effect of numerosity on our two easier search conditions, tilted among vertical ($p = 0.44$) and Q among O ($p = 0.10$). In feature search for a tilted line among vertical, the easiest condition, even when mongrels were fairly crowded (number of items = 5), subjects still maintained high discriminability (*d'* = 3.98) of target+distractor mongrels relative to distractor-only ones.

Do these new measurements of mongrel discriminability predict search difficulty, as we showed in Experiment 1? To test our model's prediction, we computed an overall mongrel

discriminability for each condition by aggregating each subject's performance across all trials (lumping together all numerosities) in Experiment 2. Replicating the finding in Experiment 1, $\log_{10}$(search slope) is significantly correlated with $\log_{10}$(aggregate mongrel *d'*) ($R^2 = .87$, $p = 0.02$, see Figure 5b).

In summary, Experiment 2 shows that mongrel discriminability is indeed predictive of peripheral discriminability for the stimuli of our 5 search conditions. Patch numerosity has a detrimental effect on both peripheral discriminability and mongrel discriminability, particularly in more difficult conditions. When mongrels are created from patches extracted directly from the search task, there still exists a significant relationship between search difficulty and mongrel discriminability, in agreement with our model prediction.

We have argued that the visual system's task in visual search is not to distinguish between individual targets and distractors, but rather to discriminate between sizable, crowded patches which may contain multiple items. Furthermore, we have argued, both in this paper and in Balas et al. (2009), that those patches are represented by a rich set of summary statistics. Experiments 1 and 2 lend credence to this view of search, by demonstrating that mongrel discriminability of crowded target+distractor and distractor-only patches can predict the qualitative difficulty of a set of classic search tasks. Furthermore, this discriminability was predictive of peripheral discriminability of such patches.

## Discussion

Many of the fundamental puzzles of visual search are puzzles because of an implicit assumption that the visual system's task is to discriminate between an individual target and an individual distractor. The difference between feature and conjunction search, search asymmetries, and other issues all indicate that search is not about discrimination of individual items. We suggest, instead, that search hinges on the performance of peripheral vision in directing our eye movements to the target, and that peripheral vision operates not on individual items but on patches which may contain multiple items. Intuitively, the capabilities and limitations of peripheral vision should have a significant impact on search performance, and one must assess this impact before one can make sense of what other factors – such as selective attention – may influence search.

If search involves discriminability of peripheral patches, this correctly predicts that search depends upon spacing of the display items (Wertheim et al., 2006), as visual crowding in the periphery depends strongly on such spacing (Bouma, 1970). One would also expect distractor heterogeneity to negatively impact search performance in many cases (Duncan & Humphreys, 1989), as such heterogeneity would decrease the fidelity with which a patch could be represented by a limited number of statistical parameters. Furthermore, thinking of search in terms of patch discriminability instead of single-item discriminability sheds light onto the fundamental puzzle of why search asymmetries are so common (Treisman & Souther, 1985; Treisman & Gormican, 1988; Wolfe, 2001). In searching for an O among Q's, we must essentially distinguish a patch containing an O and multiple Q's from a patch containing all Q's. In searching for a Q among O's, we must discriminate a patch containing a Q and multiple O's from a patch containing all O's. For an arbitrary measurement made by

the visual system on such patches, there is no reason why these two tasks should be equivalent. Asymmetries should be everywhere, which is basically what researchers have found. For sparse items, many asymmetries should go away (and do – see Wertheim et al., 2006), as many patches will contain only one item. Exceptions might include search involving tilted and vertical lines (Carrasco & Frieder, 1997), perhaps attributable to greater uncertainty in representing different orientations (Vincent, 2011), and might also include search tasks involving complex stimuli.

We have gone beyond merely suggesting that the quirks of peripheral vision provide a key determinant of search performance, by developing a testable model of representation in peripheral vision. In particular, we have proposed that the visual system, whether for reasons of efficiency, limited cortical real estate, or an information bottleneck (Nakayama, 1990), represents the periphery using a rich, fixed set of local summary statistics: the marginal distribution of luminance; luminance autocorrelation; correlations of the magnitude of responses of oriented, multi-scale wavelets, across differences in orientation, neighboring positions, and scale; and phase correlation across scale. These hypothesized statistics are a good initial guess for our model, though ultimately, determining the right set of summary statistics will require the study of many different stimuli and tasks, not just visual search.

The proposed statistical representation, while it does not allow perfect reconstruction of the periphery, nonetheless captures much useful information, which can help guide our eye movements as we look for objects of interest and explore the world around us. We have shown that discriminability of target-present from target-absent patches, based upon this statistical representation, is predictive not only of peripheral discriminability of crowded patches, but also of search performance.

## Acknowledgments

This work was funded by NIH-NEI EY019366 to R.R., and by NIH (NCRR, NIGMS) P20GM103505 to B.B.
Commercial relationships: none.

## References

Balas B. Texture synthesis and perception: Using computational models to study texture representations in the human visual system. Vision Research. 2006; 46:299–309. [PubMed: 15964047]

Balas BJ, Nakano L, Rosenholtz R. A summary-statistic representation in peripheral vision explains visual crowding. Journal of Vision. 2009; 9(12):1–18.

Bouma H. Interaction effects in parafoveal letter recognition. Nature. 1970; 226:177–178. [PubMed: 5437004]

Brainard DH. The Psychophysics Toolbox. Spatial Vision. 1997; 10:433–436. [PubMed: 9176952]

Carrasco M, Evert DL, Chang I, Katz SM. The eccentricity effect: Target eccentricity affects performance on conjunction searches. Perception & Psychophysics. 1995; 57(8):1241–1261. [PubMed: 8539099]

Carrasco M, Frieder KS. Cortical magnification neutralizes the eccentricity effect in visual search. Vision Research. 1997; 37(1):63–82. [PubMed: 9068831]

Carrasco M, McLean TL, Katz SM, Frieder KS. Feature asymmetries in visual search: Effects of display duration, target eccentricity, orientation & spatial frequency. Vision Research. 1998; 38:347–374. [PubMed: 9536360]

Carrasco M, Yeshurun Y. The contribution of covert attention to the set-size and eccentricity effects in visual search. Journal of Experimental Psychology: Human Perception & Performance. 1998; 24(2): 673–692. [PubMed: 9554103]

Duncan J, Humphreys GW. Visual search and stimulus similarity. Psychological Review. 1989; 96:433–458. [PubMed: 2756067]

Eckstein MP. The lower visual search efficiency for conjunctions is due to noise and not serial attentional processing. Psych. Science. 1998; 9(2):111–118.

Enns JT, Rensink RA. Sensitivity to three-dimensional orientation in visual search. Psychological Science. 1990; 1:323–326.

Erkelens CJ, Hooge I. Th. C. The role of peripheral vision in visual search. Journal of Videology. 1996; 1:1–8.

Freeman J, Simoncelli EP. Metamers of the ventral stream. Nature Neuroscience. 2011; 9:1195–1201.

Geisler WS, Chou KL. Separation of low-level and high-level factors in complex tasks: Visual search. Psychological Review. 1995; 102:356–378. [PubMed: 7740093]

Geisler WS, Perry JS, Najemnik J. Visual search: The role of peripheral information measured using gazecontingent displays. Journal of Vision. 2006; 6:858–873. [PubMed: 17083280]

Gheri C, Morgan MJ, Solomon JA. The relationship between search efficiency and crowding. Perception. 2007; 36:1779–1787. [PubMed: 18283928]

Kanwisher N. Repetition blindness and illusory conjunctions: errors in binding visual types with visual tokens. Journal of Experimental Psychology: Human Perception and Performance. 1991; 17:404–21. [PubMed: 1830084]

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

Levi DM. Crowding – an essential bottleneck for object recognition: A mini-review. Vision Research. 2008; 48:635–654. [PubMed: 18226828]

Levi DM. Crowding in peripheral vision: Why bigger is better. Current Biology. 2009; 19:1988–1993. [PubMed: 19853450]

Nagy AL, Thomas G. Distractor heterogeneity, attention, and color in visual search. Vision Research. 2003; 43(14):1541–1552. [PubMed: 12782068]

Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. Nature. 2005; 434:387–391. [PubMed: 15772663]

Najemnik J, Geisler WS. Eye movement statistics in humans are consistent with an optimal search strategy. Journal of Vision. 2008; 8:1–14. [PubMed: 18484810]

Najemnik J, Geisler WS. Simple summation rule for optimal fixation selection in visual search. Vision Research. 2009; 49:1286–1294. [PubMed: 19138697]

Nakayama, K.; Blakemore, C. Vision: Coding & Efficiency. Cambridge University Press; Cambridge, England: 1990. The iconic bottleneck and tenuous link between early vision processing and perception; p. 411-422.

Palmer J, Ames CT, Lindsey DT. Measuring the effect of attention on simple visual search. Journal of Experimental Psychology: Human Perception and Performance. 1993; 19:108–130. [PubMed: 8440980]

Palmer J, Verghese P, Pavel M. The psychophysics of visual search. Vision Research. 2000; 40:1227–1268. [PubMed: 10788638]

Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M. Compulsory averaging of crowded orientation signals in human vision. Nature Neuroscience. 2001; 4:739–744.

Pelli DG, Palomares M, Majaj NJ. Crowding is unlike ordinary masking: Distinguishing feature integration from detection. Journal of Vision. 2004; 4:1136–1169. [PubMed: 15669917]

Pelli DG, Tillman KA. The uncrowded window of object recognition. Nature Neuroscience. 2008; 11:1129–1135.

Põder E, Wagemans J. Crowding with conjunctions of simple features. Journal of Vision. 2007; 7(2): 23, 1–12. [PubMed: 18217838]

Portilla J, Simoncelli E. A parametric texture model based on joint statistics of complex wavelet coefficients. International Journal of Computer Vision. 2000; 40:49–71.

Reddy L, VanRullen R. Spacing affects some but not all visual searches: Implications for theories of attention and crowding. Journal of Vision. 2007; 7(2):3.1–3.17. [PubMed: 18217818]

Rodieck, RW. The first step in seeing. Sinauer Associate; 1998.

Rosenholtz R. Visual search for orientation among heterogeneous distractors: Experimental results and implications for signal-detection theory models of visual search. Journal of Experimental Psychology: Human Perception & Performance. 2001a; 27(4):985–999. [PubMed: 11518158]

Rosenholtz R. Search asymmetries? What search asymmetries? Perception & Psychophysics. 2001b; 63(3):476–489. [PubMed: 11414135]

Rosenholtz R. What your visual system sees where you are not looking. Proc. SPIE (Human Vision and Electronic Imaging XVI). 2011; 7865:786510.

Rosenholtz R, Huang J, Ehinger KA. Rethinking the role of top-down attention in vision: effects attributable to a lossy representation in peripheral vision. Frontiers in Psychology. 2012; 3:13. doi: 10.3389/fpsyg.2012.00013. [PubMed: 22347200]

Sun J, Perona P. Early computation of shape and reflectance in the visual system. Nature. 1996; 379:165–168. [PubMed: 8538766]

Toet A, Levi DM. The two-dimensional shape of spatial interaction zones in the parafovea. Vision Research. 1992; 32(7):1349–1357. [PubMed: 1455707]

Treisman A, Gelade G. A feature-integration theory of attention. Cognitive Psychology. 1980; 12:97–136. [PubMed: 7351125]

Treisman A, Gormican S. Feature analysis in early vision: evidence from search asymmetries. Psychological Review. 1988; 95(1):15–48. [PubMed: 3353475]

Treisman A, Schmidt H. Illusory conjunctions in the perception of objects. Cognitive Psychology. 1982; 14:107–141. [PubMed: 7053925]

Treisman A, Souther J. Search asymmetry: A diagnostic for preattentive processing of separable features. Journal of Experimental Psychology: General. 1985; 114(3):285–309. [PubMed: 3161978]

van Doorn AJ, Koenderink JJ, Bouman MA. The influence of the retinal inhomogeneity on the perception of spatial patterns. Kybernetic. 1972; 10:223–230.

Verghese P, Nakayama K. Stimulus discriminability in visual search. Vision Research. 1994; 34:2453–2467. [PubMed: 7975284]

Vlaskamp BNS, Over EAC, Hooge ITC. Saccadic search performance: The effect of element spacing. Experimental Brain Research. 2005; 167:246–259. [PubMed: 16078032]

Wertheim AH, Hooge ITC, Krikke K, Johnson A. How important is lateral masking in visual search. Experimental Brain Research. 2006; 170:387–401. [PubMed: 16328267]

Wolfe JM. Guided Search 2.0: A revised model of visual search. Psychonomic Bulletin & Review. 1994; 1(2):202–238. [PubMed: 24203471]

Wolfe JM. Asymmetries in visual search: An introduction. Perception & Psychophysics. 2001; 63(3): 381–389. [PubMed: 11414127]

Wolfe JM, Cave KR, Franzel SL. Guided search: An alternative to the Feature Integration model for visual search. Journal of Experimental Psychology: Human Perception & Performance. 1989; 15(3):419–433. [PubMed: 2527952]

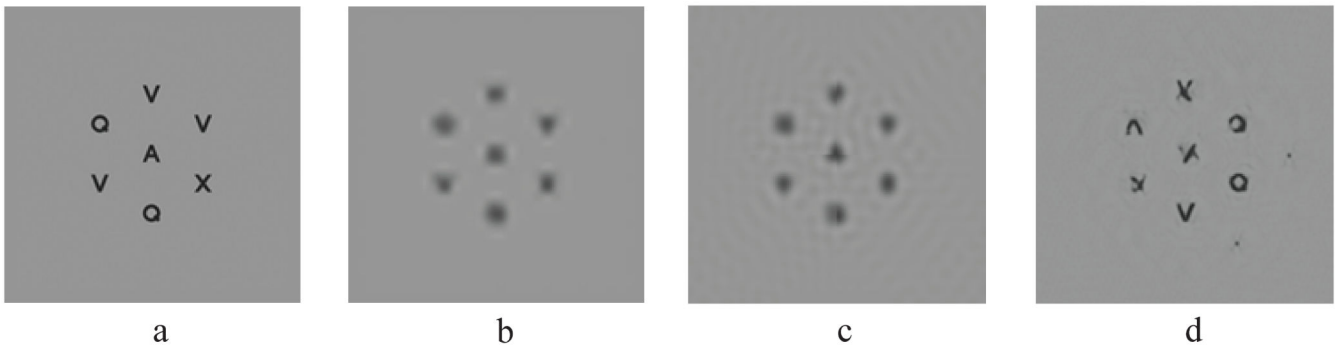


**Figure 1.**

Possible coarse encoding strategies for peripheral vision (a demo). (a) An original image patch, to be viewed peripherally. Suppose, hypothetically, that we want to represent this patch with only 1000 numbers. (b) Subsampling to reduce to a 32x32 image. Clearly this would be a poor representation. One can tell that the original stimulus consisted of 7 items in an array, but we have no idea that those items were made up of lines, nor that they formed letters. (c) Representation by local orientation at multiple scales, as in early visual cortex (V1), followed by reduction to 1000 numbers leads to a similarly poor result. This encoding used the discrete cosine transform; using a more biologically plausible wavelet transform leads to similar (but worse) results. (d) For the same 1000 numbers, one can encode a whole bunch of summary statistics, e.g.: the correlation of responses of V1-like cells across location, orientation, and scale; phase correlation, marginal statistics of the luminance, and autocorrelation of the luminance. Here we visualize the information available in this representation by synthesizing a new “sample” with the same statistics as those measured from (a), using a technique (and statistics) from Portilla & Simoncelli (2000). This encoding captures much more useful information about the original stimulus. Figure originally published in Rosenholtz (2011).

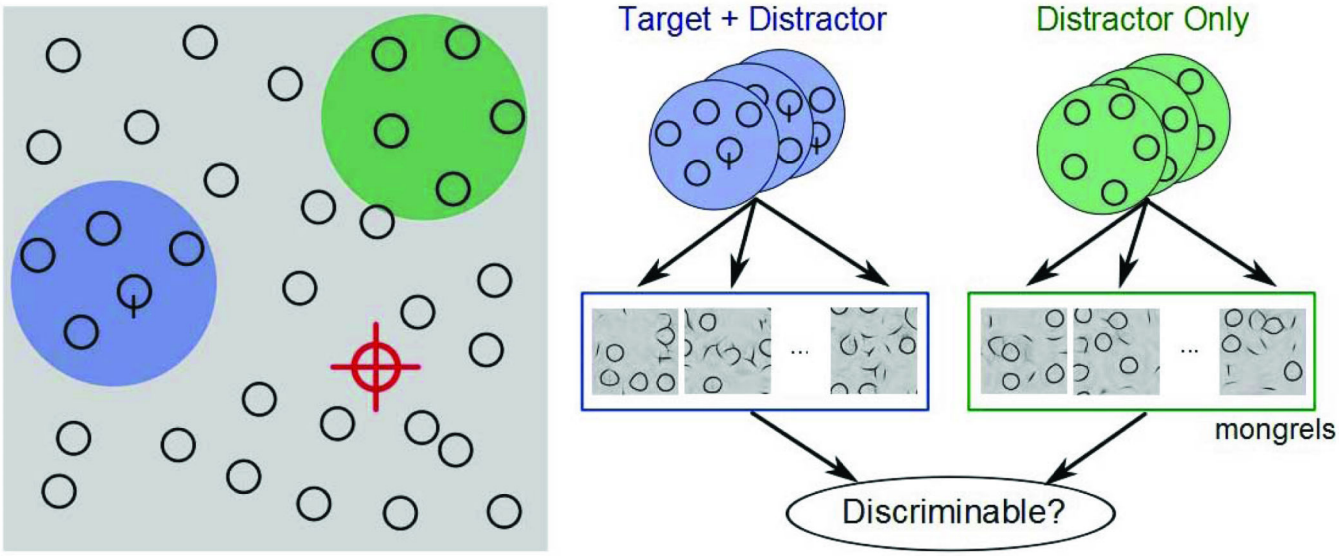


**Figure 2.**

(a) In visual search, we propose that on each fixation (red cross), the visual system computes summary statistics over a number of local patches. Some patches contain a target and distractors (blue), whereas most contain only distractors (green). The job of the visual system is to distinguish between promising and unpromising peripheral patches and to move the eyes accordingly. (b) We hypothesize, therefore, that peripheral patch discriminability, based on a rich set of summary statistics, critically limits search performance. To test this, we select a number of target + distractor and distractor-only patches, and use texture synthesis to generate a number of patches with the same statistics ("mongrels"). We then ask human observers to discriminate between target + distractor and distractor-only synthesized patches, and examine whether this discriminability predicts search difficulty. Figure originally published in Rosenholtz (2011).

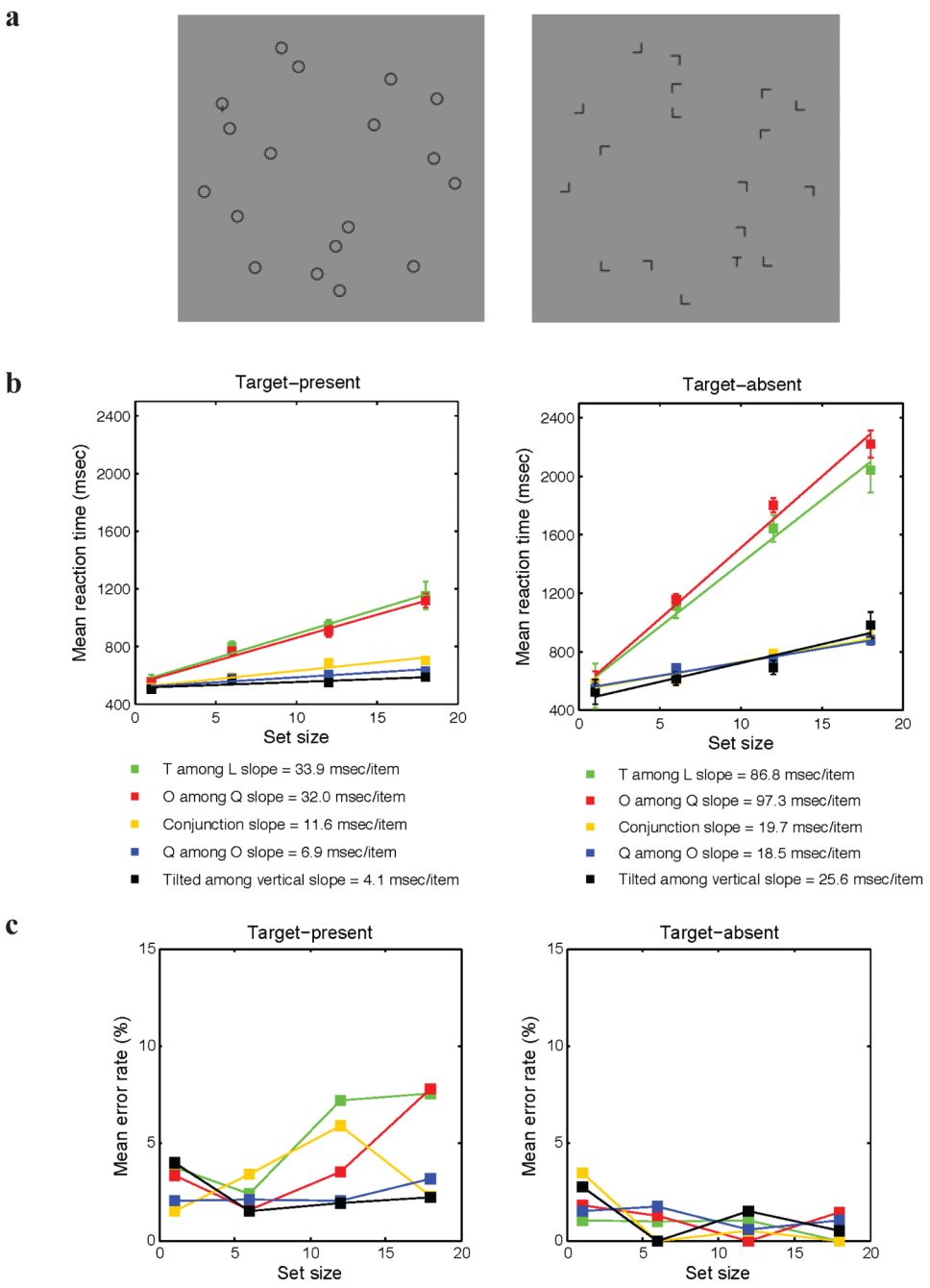


**Figure 3.**

We replicated 5 classic search experiments. (a) Example stimuli for Q among O (left) and T among L (right) search, set size = 18. (b) Mean reaction times (RTs) for correct trials, averaged across subjects. Error bars show standard errors. The legend gives the slopes of the RT vs. set size functions. (c) Error rates on target-present and target-absent trials, for each combination of set size and search condition.

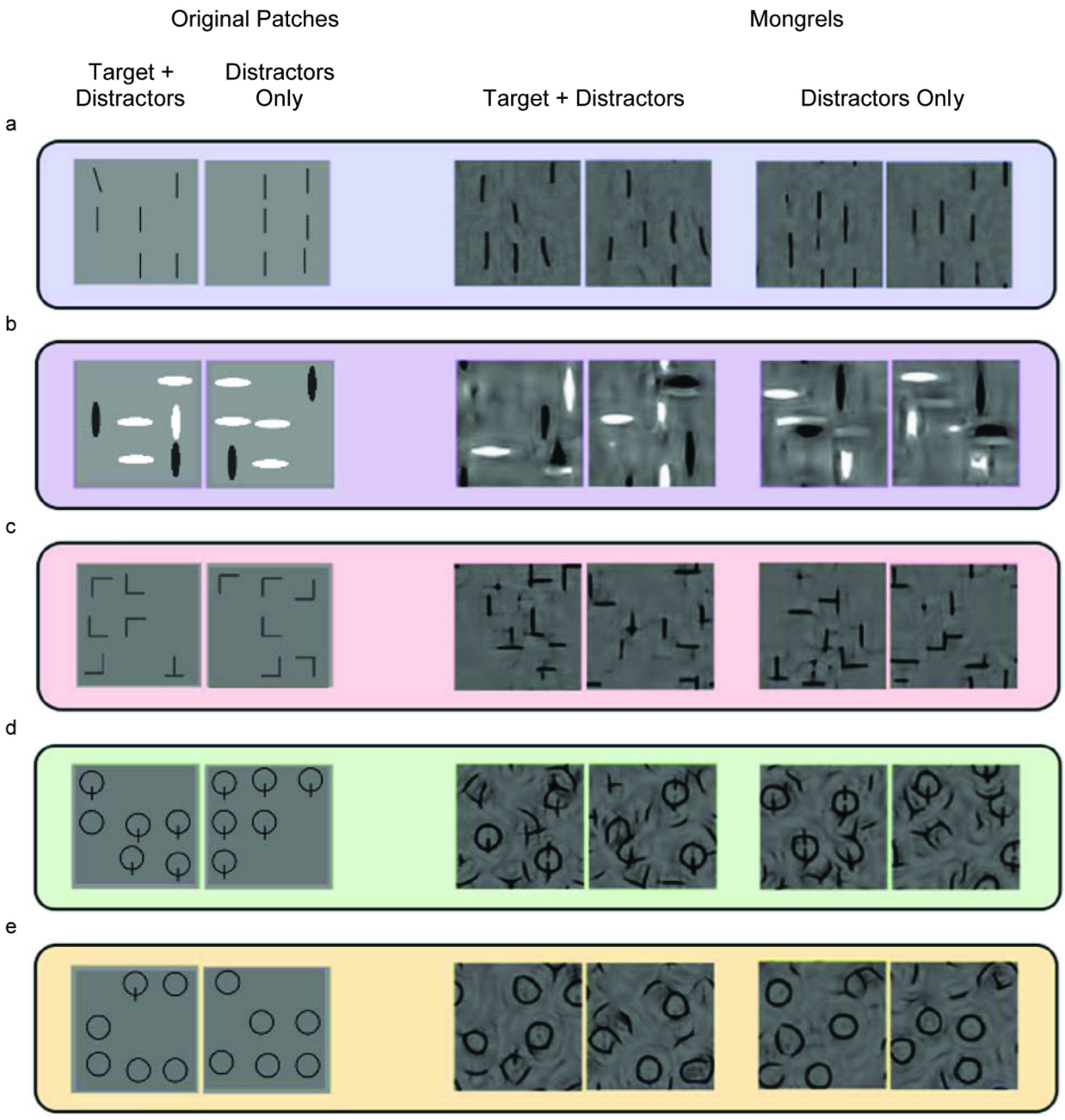


**Figure 4.**

Example target+distractor and distractor-only patches (columns 1 and 2) for 5 classic visual search conditions: (a) tilted among vertical; (b) orientation-contrast conjunction search; (c) T among L; (d) O among Q; and (e) Q among O. For each patch, we synthesized 10 mongrels – images having approximately the same summary statistics as the original patch. Examples are shown in the rightmost 4 columns, at increased contrast). Observers viewed each mongrel for unlimited time, and were asked to categorize them according to whether they thought there was a target present in the original patch.

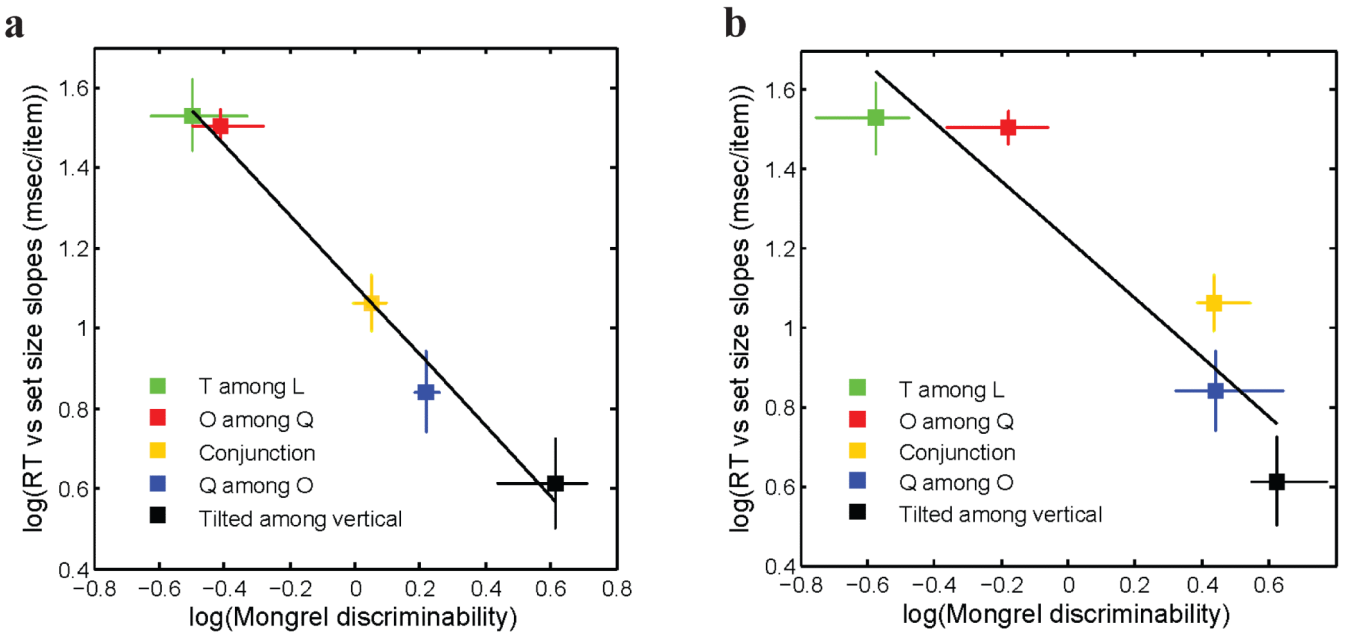


**Figure 5.**

Testing whether peripheral representation of patches by summary statistics predicts visual search performance. The y-axis shows search performance for correct target-present trials, as measured by $\log_{10}$(search efficiency), i.e. the mean number of milliseconds (msec) of search time divided by the number of display items. The x-axis shows $\log_{10}(d')$, the discriminability of the mongrels corresponding to target+distractor and distractor-only patches. Clearly there is a strong relationship between visual search difficulty and the mongrel discriminability, in agreement with our predictions. (y-axis error bars = standard error of the mean; x-axis error bars = 95% confidence intervals for $\log_{10}(d')$.). (a) Experiment 1. Adapted from Rosenholtz et al (2012). (b) Experiment 2.

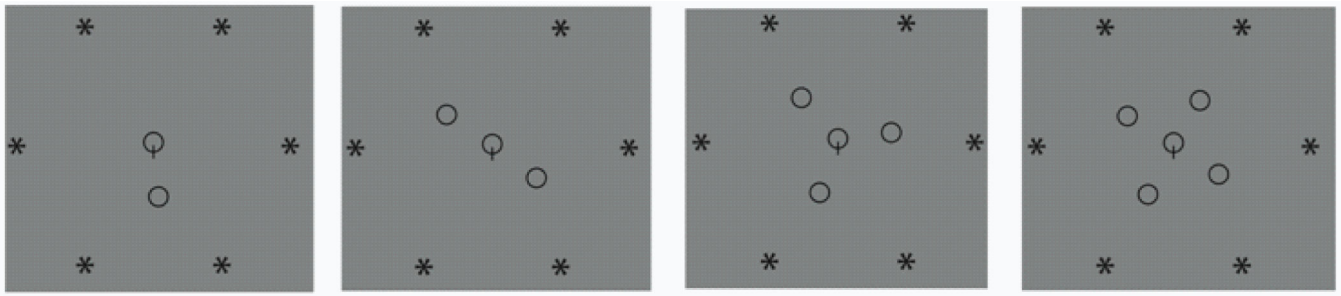

**Figure 6.**
Example peripheral patches used in the crowding task. From left to right are patches with numerosity 2, 3, 4 and 5.

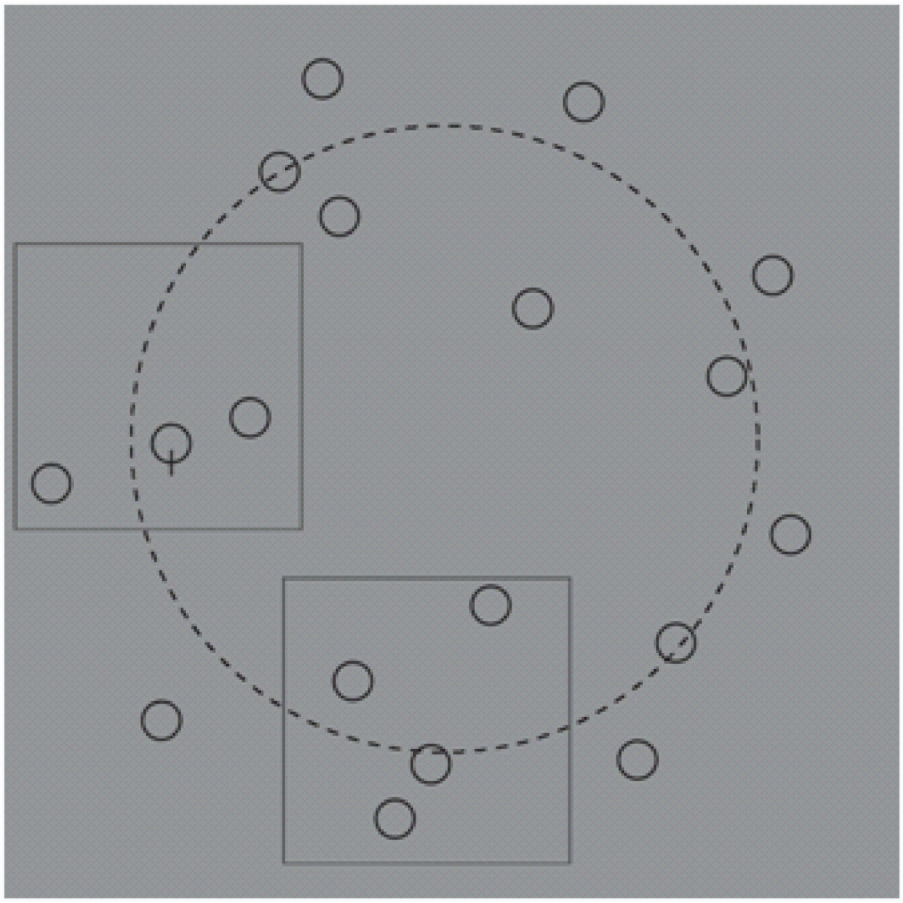

**Figure 7.**
Schematic illustration of sampling of patches from actual search displays, for mongrel synthesis. Each sampled patch has a size of 6.4 × 6.4 degrees v.a., and was located at an eccentricity of 7 degrees v.a. on the original search display.

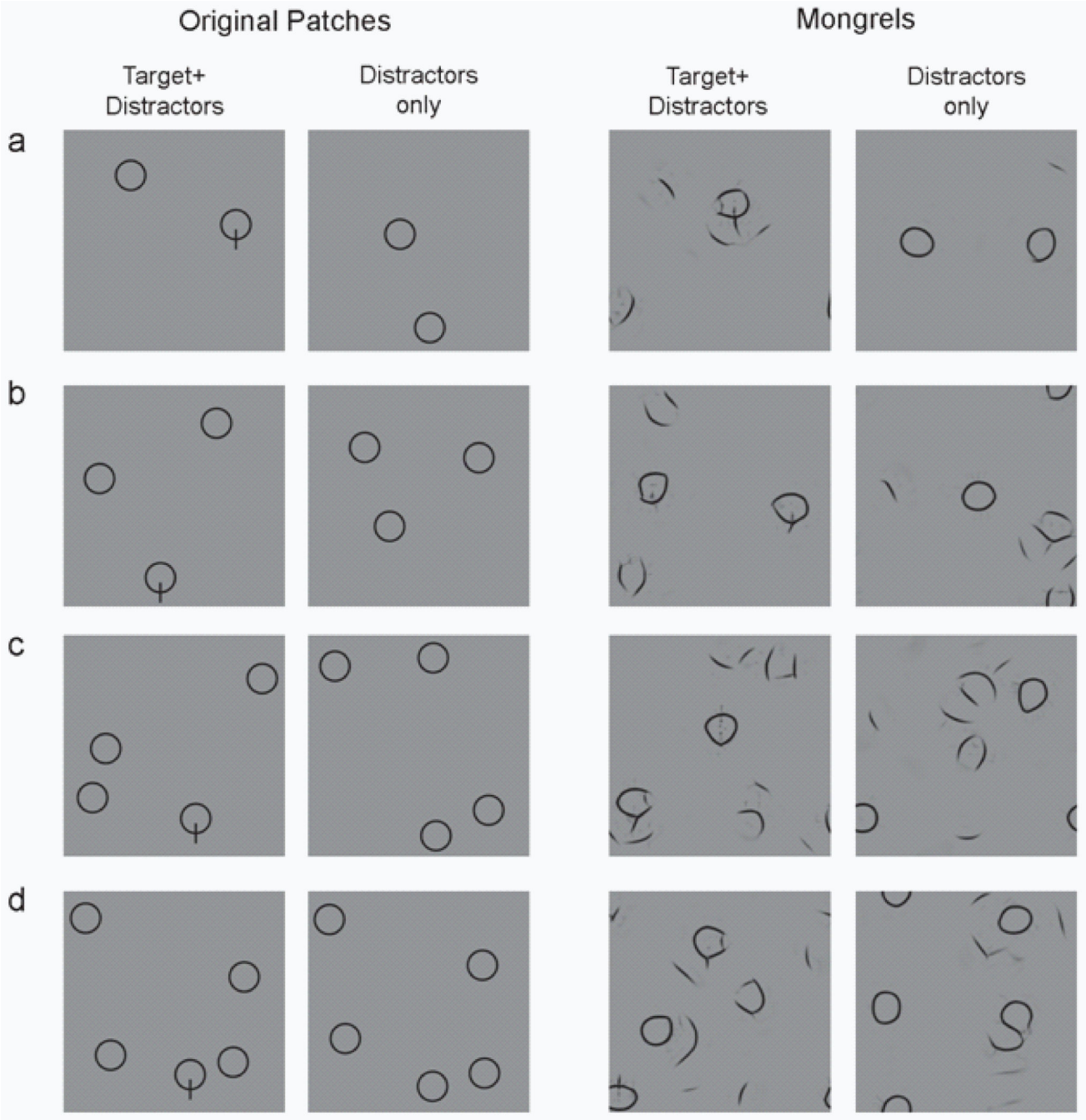


**Figure 8.**

Example target+distractor and distractor-only patches (columns 1 and 2) used in Experiment 2. Patches were sampled from search displays from Experiment 1. For each patch, we synthesized 10 mongrels (column 3 and 4), having approximately the same summary statistics. We sampled patches with four numerosity levels (2, 3, 4, and 5 items), shown in rows (a) – (d).

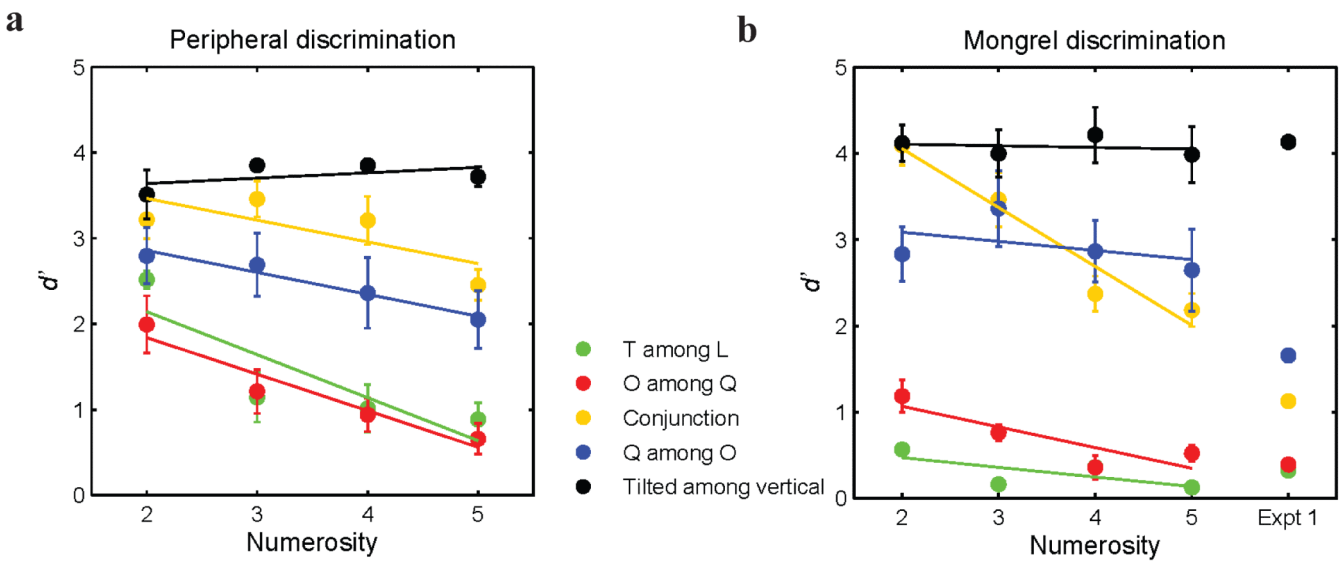


**Figure 9.**

(a) Peripheral *d'*, plotted vs. patch numerosity, for each of the five conditions. (b) Mongrel *d'*, plotted vs. patch numerosity, for each of the five conditions. Solid lines present linear fit to each condition's *d'*. The rightmost markers show mongrel *d'* obtained in Experiment 1. Error bars are within-subject standard errors of mean *d'*.

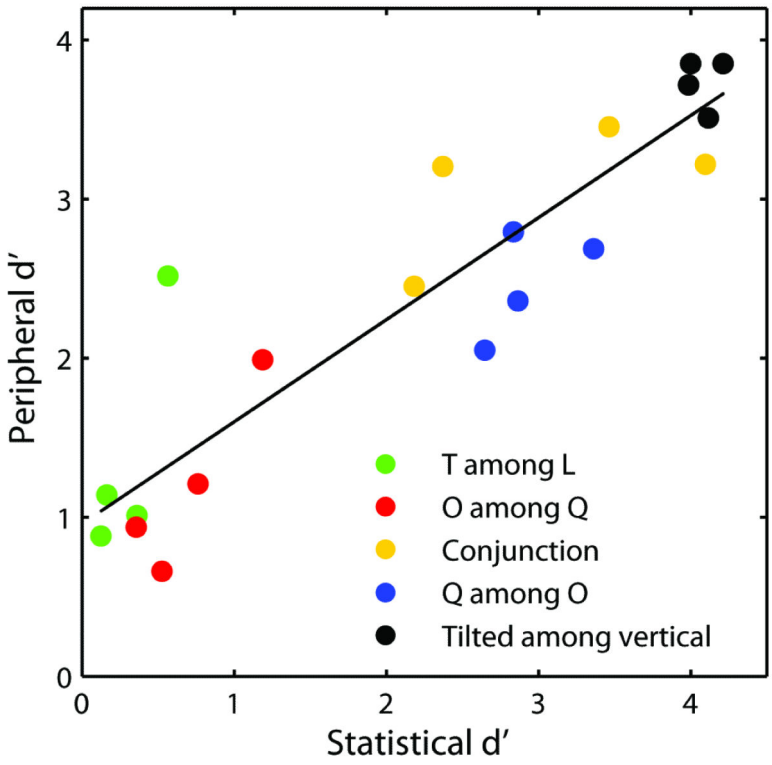


**Figure 10.**
Peripheral *d'* vs. statistical *d'*. Different colors represent different conditions, with multiple points corresponding to different patch numerosities.