Natural image clutter degrades overt search performance independently of set size

Semizer, Yelda
yelda.semizer@rutgers.edu

Michel, Melchi M.
melchi.michel@rutgers.edu

Department of Psychology, Rutgers University, New Brunswick, NJ

Abstract

While studies of visual search have repeatedly demonstrated that visual clutter impairs search performance in natural scenes, these studies have not attempted to disentangle the effects of search set size from those of clutter *per se*. Here, we investigate the effect of natural image clutter on performance in an overt search for categorical targets when the search set size is controlled. Observers completed a search task that required detecting and localizing common objects in a set of natural images. The images were sorted into high and low clutter conditions based on a clutter metric. The search set size was varied independently, by fixing the number and positions of potential targets across set size conditions within a block of trials. Within each fixed set size condition, search times increased as a function of increasing clutter, suggesting that clutter degrades overt search performance independently of set size.

1 Introduction

Interacting with the world involves, as frequent and ubiquitous subtasks, the detection and localization of objects in our visual environment. These subtasks are called visual searches. One fundamental property common to all visual searches is uncertainty regarding the positions of target objects. This study examines the properties of the visual environment and of the visual system that contribute to this position uncertainty. In particular, our goal is to investigate how visual clutter affects performance when observers search natural images for categorical targets.

Position uncertainty can be due to either extrinsic or intrinsic sources. For example, an observer searching an unfamiliar bookshelf for a particular book will probably have some uncertainty about the location of the book. In this case, (i.e., when the observer does not know the book's location *a priori*) the uncertainty is a result of imprecise specification of the likely target location. This type of position

uncertainty increases with the number of potential target locations and is called *extrinsic* position uncertainty. However, even when the observer is familiar with the bookshelf and knows the order of its books, she might still have a hard time localizing the book in the visual periphery. This uncertainty is a result of the limitations intrinsic to the visual system and it is called *intrinsic* position uncertainty.

Regardless of whether it is extrinsic or intrinsic, position uncertainty impairs performance for detecting, discriminating, and localizing stimuli. This is indicated by decreases in detection and localization accuracy (Burgess & Ghandeharian, 1984; Eckstein, Thomas, Palmer, & Shimozaki, 2000), by increases in detection thresholds (Cohn & Wardlaw, 1985; Palmer, Verghese, & Pavel, 2000), and by increases in search times (Egeth, Atkinson, Gilmore, & Marcus, 1973; Treisman & Gelade, 1980). While research on the effects of position uncertainty has typically focused on extrinsic sources of uncertainty (e.g., Bochud, Abbey, & Eckstein, 2004; Burgess & Ghandeharian, 1984; Swensson & Judy, 1981), a few studies have explicitly focused on intrinsic sources (e.g., Michel & Geisler, 2011; Pelli, 1985; Tanner, 1961). Evidence from these studies, and from studies of visual crowding (e.g., Bouma, 1970; Levi, 2008; Pelli, Palomares, & Majaj, 2004; Pelli et al., 2007) suggests that the ability to identify and localize features declines systematically in the periphery. Indeed, position uncertainty has been repeatedly implicated as a primary contributor to crowding (Krumhansl & Thomas, 1977; Pelli, 1985; Popple & Levi, 2005; Wolford, 1975). For example, similar to crowding (Bouma, 1970; Levi, 2008; Levi, Hariharan, & Klein, 2002), intrinsic position uncertainty also increases approximately linearly with eccentricity (Michel & Geisler, 2011). Moreover, the eccentricity-dependent effects of position uncertainty seem to persist in overt search tasks (Semizer & Michel, 2017).

As an inherent property of the observer's visual system, intrinsic position uncertainty cannot be experimentally controlled. However, its effect on performance can be observed by manipulating the visual environment. In a recent study, Semizer and Michel (2017) introduced an experimental technique that modulates the effects of intrinsic uncertainty independently of extrinsic uncertainty by manipulating the distribution of clutter in synthetic noise displays. Using this technique, the authors showed that intrinsic position uncertainty substantially limits overt search performance and that its effects are especially evident when the amount of extrinsic uncertainty is controlled. Does this result generalize to real-world searches?

In many ways, synthetic visual stimuli have been incredibly useful for vision research. Synthetic stimuli provide researchers with a great deal of flexibility and control, enabling them to manipulate individual stimulus features and to determine how these contribute to performance in a variety of tasks. In visual search, for example, measuring performance in synthetic search displays has allowed researchers to discover how observers use information about peripheral target visibility to select fixations (Najemnik & Geisler, 2005; Geisler, Perry, & Najemnik, 2006; Najemnik & Geisler, 2008; Michel & Geisler, 2009; Zhang & Eckstein, 2010; Verghese, 2012), how intrinsic position uncertainty and clutter in the periphery

degrade performance (Michel & Geisler, 2011; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Semizer & Michel, 2017), how the template for known search targets is structured (Eckstein, Beutter, Pham, Shimozaki, & Stone, 2007), and how observers integrate information about the target across fixations (Caspi, Beutter, & Eckstein, 2004; Kleene & Michel, 2018), all while controlling extraneous properties of the search display (e.g., spectral spatial frequency statistics, environmental contingencies, target location probabilities, etc.) in ways that would be difficult or impossible with natural scenes. However, their highly controlled nature means that synthetic displays may provide only limited insight into how observers search in naturalistic settings.

For example, the search targets used in synthetic displays typically exhibit very little variability across trials, and observers are therefore assumed to represent them with little uncertainty. In contrast, the targets of natural searches typically exhibit many sources of variability. Objects in natural scenes appear in various positions and orientations, occlude one another, and change appearance depending on the lighting conditions. Moreover, individual exemplars may vary considerably within a natural object category. These sources of variability introduce additional uncertainty that might overwhelm any effects of intrinsic uncertainty on search performance. Thus, it is important to verify that the factors that explain search performance in synthetic displays generalize to account for searches in more naturalistic displays.

One of the major challenges associated with naturalistic tasks in the context of visual search is to quantify the amount of clutter in natural images. Unlike in artificial displays, clutter cannot be directly manipulated in natural images. However, a variety of models have been proposed to quantify scene clutter. These include edge density (Mack & Oliva, 2004), feature congestion (Rosenholtz, Li, Mansfield, & Jin, 2005; Rosenholtz, Li, & Nakano, 2007), subband entropy (Rosenholtz et al., 2007), the scale invariant clutter measure (Bravo & Farid, 2008), and the proto-object model (Yu, Samaras, & Zelinsky, 2014). Using these measures, several studies have shown that clutter degrades performance for search in various types of naturalistic displays including geographic maps (Rosenholtz et al., 2007), quasi-realistic scenes (Neider & Zelinsky, 2011), natural scenes (Henderson, Chanceaux, & Smith, 2009), images displaying contents of bags (Bravo & Farid, 2008), and photo-collages of objects (Bravo & Farid, 2004, 2008).

However, these findings confound different potential sources of position uncertainty. As a scene gets cluttered, the number of possible target locations (i.e., set size) also increases. This increase in set size augments the position uncertainty due to extrinsic sources. At the same time, due to intrinsic sources of position uncertainty, the ability to exclude irrelevant signals in the periphery decreases in highly cluttered scenes (Michel & Geisler, 2011; Semizer & Michel, 2017). These two concurrent effects of clutter make it challenging to separate the contributions of extrinsic versus intrinsic uncertainty on performance in highly cluttered images.

The goal of the current study was to separate out the contributions of intrinsic versus extrinsic sources of position uncertainty and to characterize them in a naturalistic search task that requires searching natural images for categorical targets. As in Semizer and Michel (2017), we approached this goal by controlling and manipulating set size independently of clutter. Instead of imposing synthetic clutter, we used an existing clutter measure (Bravo & Farid, 2008), chosen for its efficiency and its demonstrated correlation with search performance, to quantify the existing clutter in a set of natural images. The images were sorted into high and low clutter conditions based on this clutter measure. The "relevant set size" (Palmer, 1994, 1995), which governed the extrinsic position uncertainty was varied independently by manipulating the number and positions of cues indicating potential target locations. Within each fixed set size condition, search times increased as a function of increasing clutter, suggesting that clutter degrades overt search performance independently of set size.

2 Methods

2.1 Observers

A total of twenty-five observers participated in the study. One of the observers was an author; the remaining observers were naïve to the purpose of the experiment and received compensation for their participation. All observers had normal or corrected-to-normal vision.

2.2 Apparatus

Stimuli were presented on a 22-in Philips 202P4 CRT monitor at 100 Hz. The resolution was set to 1280×1024 pixels. Observers were seated 70 cm away from the display so that the display subtended $15.8^\circ \times 21.1^\circ$ of visual angle. The stimuli displays were programmed using MATLAB software (Mathworks) and the Psychophysics Toolbox extensions (Brainard, 1997). Observers' eye movement signals were monitored and recorded using an Eyelink 1000 infrared eye tracker (SR Research, Kanata, Ontario, Canada) at 1000 Hz. Head position was stabilized using a forehead and chin rest.

2.3 Stimuli

Images of natural scenes often contain contextual information that effectively reduces the search set size (Castelhano & Heaven, 2011; Neider & Zelinsky, 2006; Oliva & Torralba, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). To minimize this contextual information, we chose a set of images displaying the contents of bags in arbitrary arrangements (see Figure 1). These images were retrieved from the "What's in your bag?" group on Flickr[1]. We selected five of the most common objects in the image set (cellphones, glasses, iPods, keys, and pens/pencils) to serve as the categorical search targets. If a target object was present in the image, it was either present as a single instance or, in the case of

collective objects, as a single group of instances in close proximity (e.g., keys attached to a keychain). There was never more than one instance or group of the target object present in the image.

Figure 1: Example displays for the low clutter (on the left) and the high clutter (on the right) conditions with keys as the search target. Keys are located near the center in both images. Images are retrieved from "What's in your bag?" group on https://www.flickr.com.

2.3.1 Creating the image data set

The image data set was created by processing raw images in four separate stages: initial filtering, transformation, labelling, and selection. Each stage was described in detail next.

Initial filtering stage. Images were downloaded and subsequently checked for duplicates and quality (e.g., blurs, artifacts, etc.). We avoided scaling the size of small images up to preserve image quality. Therefore, images whose maximum dimension smaller than the height of the stimulus window (1024 pixels) were excluded.

Transformation stage. The clutter measure used in our experiment is sensitive to the image size (see the Measuring clutter section). To control for any potential effects of image size on quantifying clutter, we resized the minimum dimension to 1024 pixels.

Next, we considered the variability in color across images. In order to control for the effects of color on performance, colored images were converted to grayscale intensity images by removing the hue and saturation information while keeping the luminance information. RGB values were converted to grayscale values by computing a weighted sum of the channels using the intensity transformation

$$I = 0.299R + 0.587G + 0.114B, \tag{1}$$

where I represents the grayscale intensity, and R, G, and B corresponds to red, blue, and green channels, respectively.[2] The clutter was computed for both colored and grayscale versions of each image (see the Measuring clutter section). The distribution of clutter was similar across search images containing different target object categories (see Figure 2, left panel). Finally, to control the variability in luminance

[1] A subset of images from this group were also used in a search task by Bravo and Farid (2008).

and contrast levels across images, the average luminance of each image was set to 40 cd/m^2 and its contrast level (root-mean-square, RMS) was adjusted to 0.4.

Labelling stage. Images were annotated by labeling the type of potential target objects present in them. Then, target locations were marked by drawing circumscribing polygons around the target objects. The vertices of these polygons were recorded. At the end of this stage, each image was associated with an annotation consisting of: a list of target objects within the image, a list of vertices describing the circumscribing polygon for each target object, and the clutter value for the image.

Selection stage. For each of five target categories, 800 test images were selected. The target object was present only in half of these images. Test images were chosen based on the following criteria.

First, we wanted our clutter conditions to represent instances of distinctly high and low clutter. Therefore, for each target category, we selected only images at the extreme ends of the clutter distribution (i.e., $< 30^{th}$ percentile and $> 70^{th}$ percentile) as potential test images.

Next, we expected that target size might impact search performance in target-present images. To control for any size effects, we first measured the size of each target object by computing the area of its circumscribing polygon. Target size varied depending on the target category (see Figure 2, right panel). For example, on average, cellphones were larger than keys. To limit the effects of unusually-sized objects, we restricted the variability in target size by including images only if $t \in [\frac{1}{4}m, 4m]$, where t is the target size and m is the median target size.

A final inclusion criterion considered the variants of targets. If we suspected that observers might not be familiar with a particular variant of target object, images displaying that variant were not selected. For example, in the case of cellphones, images did not include any flip-phones. Similarly, in the case of iPods, only images with iPods with a particular shape, a rectangular screen at the top and a circular area at the bottom, were included. Further, images with objects that looked highly similar to targets were also excluded. For example, images containing an iPod touch (which might look like an iPhone to the observer) were excluded. Similarly, in the case of pens, we excluded images that included makeup pencils.

At the end of this process, 800 images were selected for each target category. 400 test images were selected for the target-present trials by prioritizing the amount of clutter and checking for the criteria listed above, and another 400 target-absent images were selected to match the clutter values of the target-present images.

2.3.2 Preparing the search stimuli

Images were formatted to be presented in the search task. Individual images in each of the two clutter conditions were randomly assigned to either the low (5 locations) or high (13 locations) set size

[2]The weights used in conversion of RGB values to grayscale values were based on ITU-R Recommendation BT.601-7 standard for color video encoding.

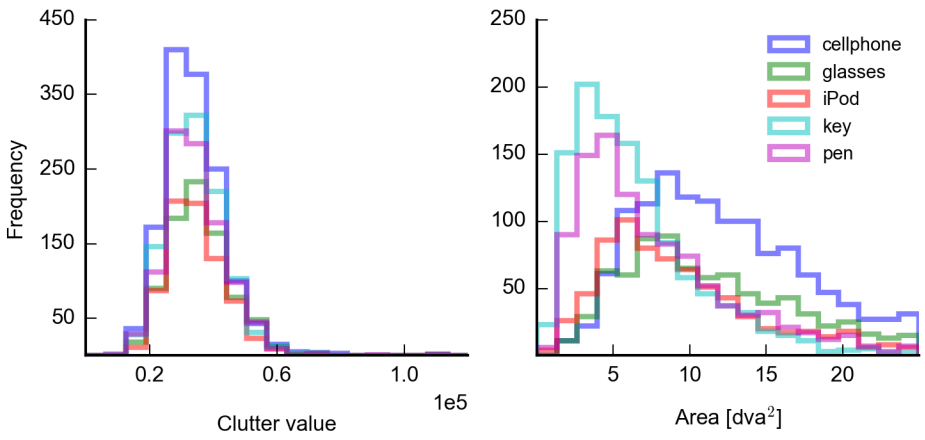

Figure 2: The distribution of clutter values in colored images (left panel) and target size (right panel) for each target category.

conditions. Potential target locations were marked by small circular cues overlaid on the image. Cue size was 0.25° in diameter. Images were shifted and rotated so that only one of these cues appeared within the circumscribing polygon associated with the correct target location. Images were presented in a circular region, 24° in diameter. This region was chosen with the constraint that it contained the target object. Finally, the area around the circular region was set to uniform gray. The final form of images used as stimuli in the search task is shown in Figure 3.

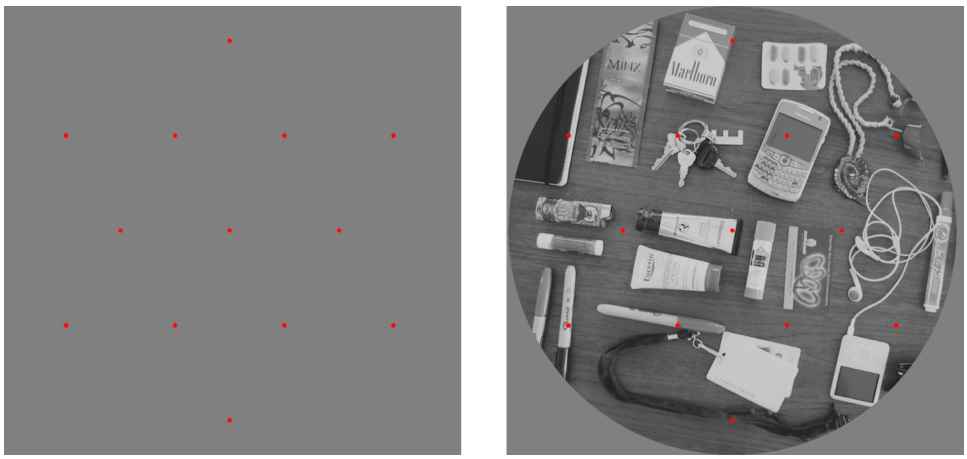

Figure 3: Search task sequence for a trial with keys as the search target. Small red cue markers represent the potential target locations ($N = 13$). The keys are located within the top left quadrant of the image.

2.3.3 Measuring clutter

We quantified image clutter using a modified version of the clutter measure described in Bravo and Farid (2008). We chose this clutter measure because it has been shown to successfully predict search times

in a similar set of images. Additionally, this measure is computationally efficient and scale invariant. Briefly, this measure estimates the amount of clutter in an image as a function of the relationship between the number of "segments" in an image and the scale of segmentation. The details of the segmentation procedure and our implementation of the clutter measure are described below.

Segmentation algorithm. To count the number segments in each image, we used the graph-based segmentation algorithm introduced by Felzenszwalb and Huttenlocher (2004). This algorithm segments the image by considering the variability of nearby regions. In particular, it draws boundaries between regions based on pairwise comparisons of the intensities within and across regions. The threshold for drawing these boundaries is controlled by a scale parameter k. Larger k leads the algorithm to favor larger regions and results in smaller number of segments. The algorithm produces perceptually reasonable segments (e.g., see Figure 4) and it runs at a high speed in practice.

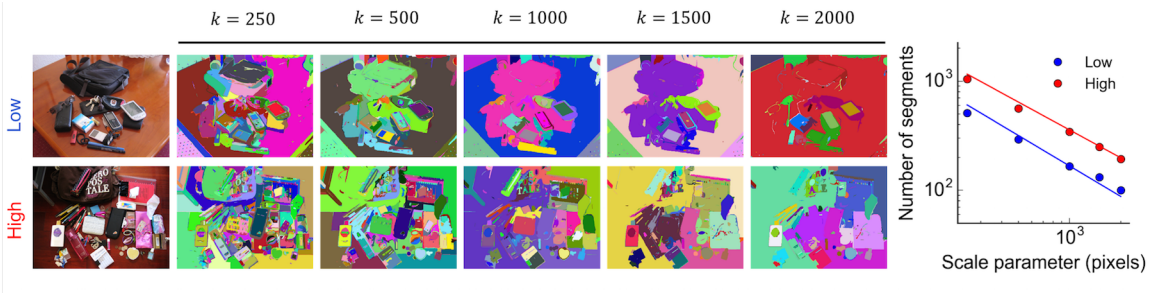

Figure 4: Example segmented images of a low clutter (top) and high clutter (bottom) image at six values of the scale parameter. Color is used only to show segmented regions in the image. Plot on the right shows the number of segments as a function of the scale parameter for each image. Points represent the raw number of segments while the lines represent the log-linear fits.

To get more stable estimates, instead of using the whole image at once, we created random samples from each image. We counted segments obtained for each sample at multiple scales. Then, we computed the geometric mean of segment counts across samples at each scale. At the end of this process, each image was associated with a segment count for each scale.

Clutter measure. We measured the clutter in each image by characterizing the relationship between the scale of segmentation k and the number of segments for that scale $y(k)$. We determined this relationship empirically by varying the scale parameter across a range of values, applying the segmentation algorithm, and counting the resulting number of segments. Figure 4 shows examples of segmented images and the number of segments at several scales of segmentation. For any given image, the number of segments is log-linearly related to the scale of segmentation, such that

$$\ln y(k) = \alpha + \beta \ln k, \tag{2}$$

where ln represents the natural logarithm.

The slope of this relationship is approximately constant ($\beta \approx -0.71$), but the intercept α varies

across images. In particular, for any setting of the scale parameter, highly cluttered images tend to have more segments than the minimally cluttered images. Therefore, we used the intercept of each image to quantify its clutter.

To get robust estimates of these log-linear relationships for each image, we (1) randomly sampled 10 1024×1024 sections of the image, (2) computed the segment counts for each of these samples across a range of scales ($k \in [180, 4095]$), and (3) computed the intercept of the log-linear fit using a least-squares procedure. The slope was computed as the average least-squares slope for all of the images in the data set ($N = 4,953$), and the intercepts for individual images were fitted with this average slope held constant.

In order to evaluate the generalizability/robustness of our clutter measurements, we also quantified clutter in our image data set using alternative clutter measures including edge density (Mack & Oliva, 2004), feature congestion (Rosenholtz et al., 2005, 2007), and subband entropy (Rosenholtz et al., 2007). The clutter measures were all highly correlated (see Table 1), suggesting that that the particular choice of clutter metric is not important.

The code for implementation of the segmentation algorithm is made publicly available by its authors. A MATLAB implementation of the clutter measure using this algorithm as described above can be found at [LINK].

Table 1: Correlation coefficients among clutter measures.

	Segmentation	Edge density	Feature congestion	Subband entropy
Segmentation	-			
Edge density	0.732	-		
Feature congestion	0.748	0.712	-	
Subband entropy	0.564	0.579	0.672	-

Note: All p <0.001.

2.4 Procedure

The design of the experiment was $5 \times 2 \times 2 \times 2$, with one between-subjects variable (target object category) and three within-subjects variables (search set size, clutter level, and target present/absent). At the start of the search experiment, observers were randomly assigned to one of five search target categories (cellphone, glasses, iPod, keys, or pens/pencils). Observers were instructed to detect and locate the target object within an image as quickly and accurately as possible. Additionally, they were told that if the search target was present in an image, there was only one single item or a group of items in close proximity from the search category, and the item was visible.

Before the start of each trial, observers fixated a point at the center of the display while a set of circular cues indicated the potential target locations (see Figure 3). Observers began the trial by pressing a start key. After the trial was initiated, the search display appeared and observers freely searched for the target.

Observers were allowed three seconds to search. After either three seconds had elapsed or the observer pressed a key to end the trial early, the search image disappeared while the set of circular cues indicating the potential target locations remained on the screen. An additional cue appeared at a random location 1° outside the search region. Observers were instructed to make a localization decision. They fixated at either the cue corresponding to the perceived location of the target (if target was present) or the additional cue (if target was absent). After the fixation, observers were required to log their responses with a keypress.

The amount of time spent inspecting each image was recorded as the search time, and was the primary measure of performance. In target present trials, a response was registered as "correct" if the indicated target location was closest to the actual target location among all possible locations. In the target absent trials, a response was registered as "correct" if the indicated location was closest to the absent cue location than to any other location. Observers received auditory feedback indicating the accuracy of their responses.

Trials were blocked by the relevant set size. Each block consisted of 50 experimental trials. At the start of each block, observers completed a 13-point calibration routine covering the central 22° of gaze angle. The calibration was repeated until the average test-retest calibration error across gaze points fell below 0.25°. The calibration routine could be repeated if necessary during a block. If a blink was detected during a trial, the trial was aborted, and the observer was notified. Data from aborted trials were discarded, but the image from the discarded trial was repeated later in the experiment.

Observers completed the study in two one-hour sessions on separate days. Each session contained 8 blocks, resulting in a total of 800 trials. The block order was randomized across sessions and observers.

Observers were trained and refamiliarized with the task by completing 8 practice trials at the start of the experiment and a single practice trial at the start of each block. Data from the practice trials were excluded from the analysis.

3 Results

3.1 Search times

Figure 5 shows average search times in the target present trials and target absent trials. Each faint line represents data from five observers searching for one type of target (shapes) in either high clutter (red lines) or low clutter (blue lines) condition as a function of relevant set size. Two main trends are evident: (a) search times tend to increase as the relevant set size increases, and (b) search times tend to increase as the amount of clutter increases.

Search times were analyzed by conducting a $5 \times 2 \times 2 \times 2$ mixed design ANOVA, with one between-subjects variable (target object category) and three within-subjects variables (relevant set size, clutter

level, and target present/absent).

The ANOVA revealed main effects of clutter level, $F(1, 20) = 333.62, p < 0.001$, of relevant set size, $F(1, 20) = 49.43, p < 0.001$, and of target present/absent $F(1, 20) = 170.94, p < 0.001$. The main effect of target category did not reach significance, $F < 1, n.s.$ There was a significant two-way interaction between clutter level and target category, $F(4, 20) = 4.03, p = 0.015$, clutter level and relevant set size $F(1, 20) = 8.96, p = 0.007$, and clutter level and present/absent $F(1, 20) = 26.35, p < 0.001$, as well as a significant two-way interaction between set size and target present/absent $F(1, 20) = 12.22, p = 0.002$. The four-way interaction was also significant, $F(4, 20) = 4.39, p = 0.01$.

The average search time was larger in the high clutter condition ($M = 1.38, SE = 0.01$) than in the low clutter condition ($M = 1.18, SE = 0.01$), suggesting that clutter degrades search performance. This pattern of results appeared across all target categories, but the effect was larger for some target categories than others. The search time was also larger in the large set size ($M = 1.37, SE = 0.01$) than in the small set size ($M = 1.19, SE = 0.01$), confirming our manipulation of set size. Finally, on average, target absent trials resulted in larger search times ($M = 1.55, SE = 0.01$) than the target present trials ($M = 1.00, SE = 0.01$).

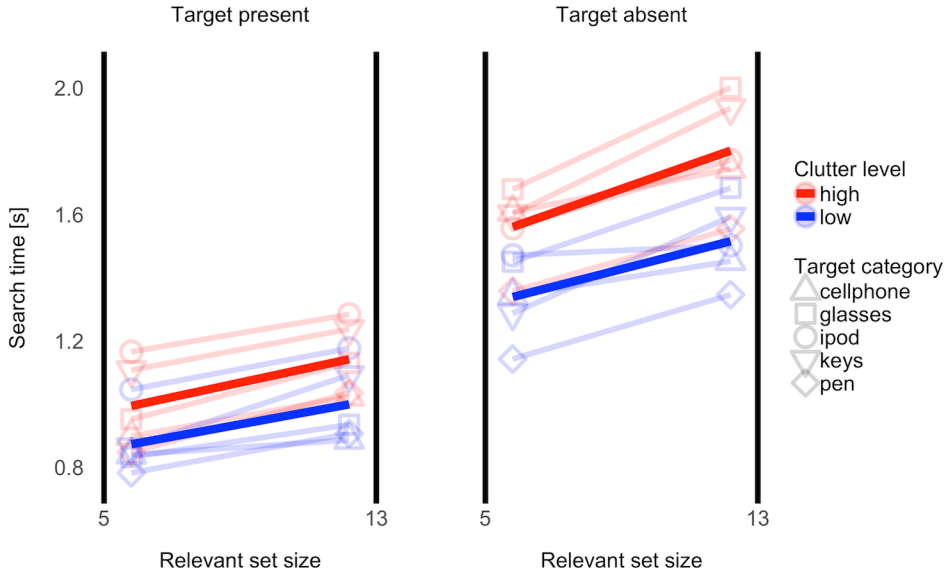

Figure 5: Average search time as a function of relevant set size in the target present trials (left panel) and in the target absent trials (right panel). Each combination of line and symbols represents data from five observers searching for one type of target (shapes) in either high clutter (red lines) or low clutter (blue lines) condition. Average search times across target categories are represented by the heavy lines.

3.2 Fixation distributions

As a further check on our manipulation of set size, we examined observers' fixation distributions during search. If observers make use of the target location information provided by the cues when

planning their fixations, they should be more likely to fixate the cued locations than other locations in the display. Figure 6 shows the fixation distributions for each of the set size conditions, aggregated across all observers and trials (with the first and last fixations excluded). Indeed, observers appear to use cue location information when selecting their fixation locations, confirming the effectiveness of our set size manipulation.

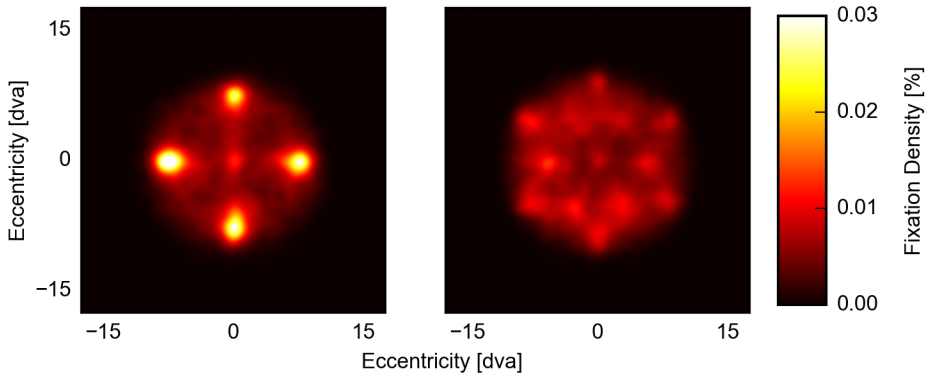

Figure 6: Aggregated fixation distributions across all of the observers, for set size 5 (left panel) and for set size 13 (right panel). The first and final fixations were excluded from the analysis.

3.3 Search target sizes

The difference in search performance across target categories could be caused in part by differences in average target sizes. To investigate this possibility, we examined how search time changed a function of target size. Although we restricted the size of the targets to a limited range, there was still some degree of variability. Target size was defined as either the area of its circumscribing polygon or the length of the longest axis of this polygon. To remedy the curvilinear relationship observed between the target area and the search times, the areas were transformed by taking their square root, which resulted in a more linear relationship. Figure 7 shows that (a) search times tend to decrease as the search target gets larger in size, and (b) some targets are larger, on average, than others. The analysis showed that search times decrease significantly as target size increases, both when the size was measured as the area ($r = -0.29, p < 0.001$) and when it was measured as the length of longest axis ($r = -0.35, p < 0.001$). These results suggest that target size may be one of the factors driving differences in search performance among target categories.

3.4 Search target categories

Our stimulus set contained some common images across different target categories. That is, in some cases, different observers searched for different targets in the same image. These cases gave us the ability to dissociate effects of the search image from those of the search target and to directly examine the effect

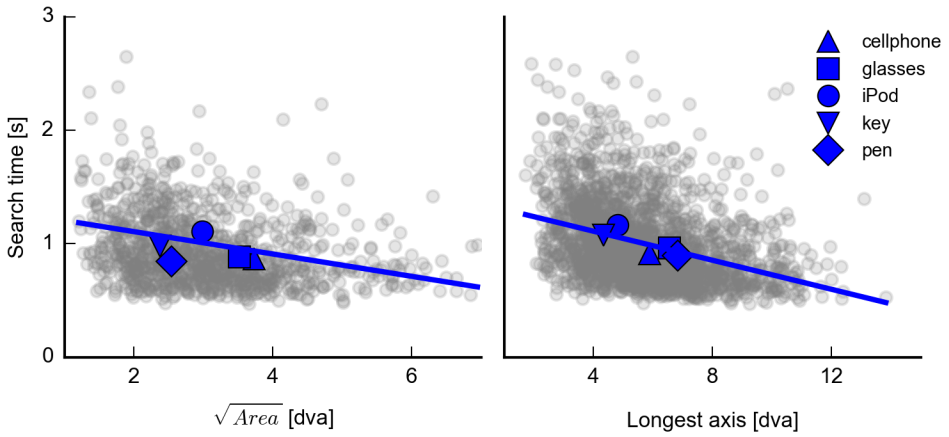

Figure 7: Search times as a function of target size represented by the square root of the area (left panel) or the longest axis (right panel) of the bounding polygon. Each gray dot represents the average search time across five observers for a particular target in an image. Shapes represent the average size for each target category. The best linear fit is given by the blue line.

of target category on search performance. Figure 8 shows the average search times while searching for different targets in the same image. For example, the first plot shows the search time while looking for a cellphone compared to the search time while looking for the other targets in the same image. If the search performance was only determined by the amount of clutter or the relevant set size, then all points would line up on the diagonal. However, these results show that some targets were harder to find than others. For example, on average, observers seem to be faster at locating cellphones than other targets, except pens. These findings suggest that certain features make some targets less susceptible to clutter than others. We discuss potential implications of this result in the Discussion section.

3.5 Error rates

Table 2 shows error rates across conditions. In general, observers were extremely accurate in their judgments.

Table 2: Error rates across conditions.

Clutter level	Relevant set size	Target category					
		cellphone	glasses	iPod	key	pen	average
Low	5	0.038	0.029	0.087	0.033	0.066	0.051
	13	0.044	0.061	0.110	0.053	0.086	0.071
High	5	0.030	0.049	0.096	0.054	0.082	0.062
	13	0.061	0.070	0.124	0.074	0.114	0.089

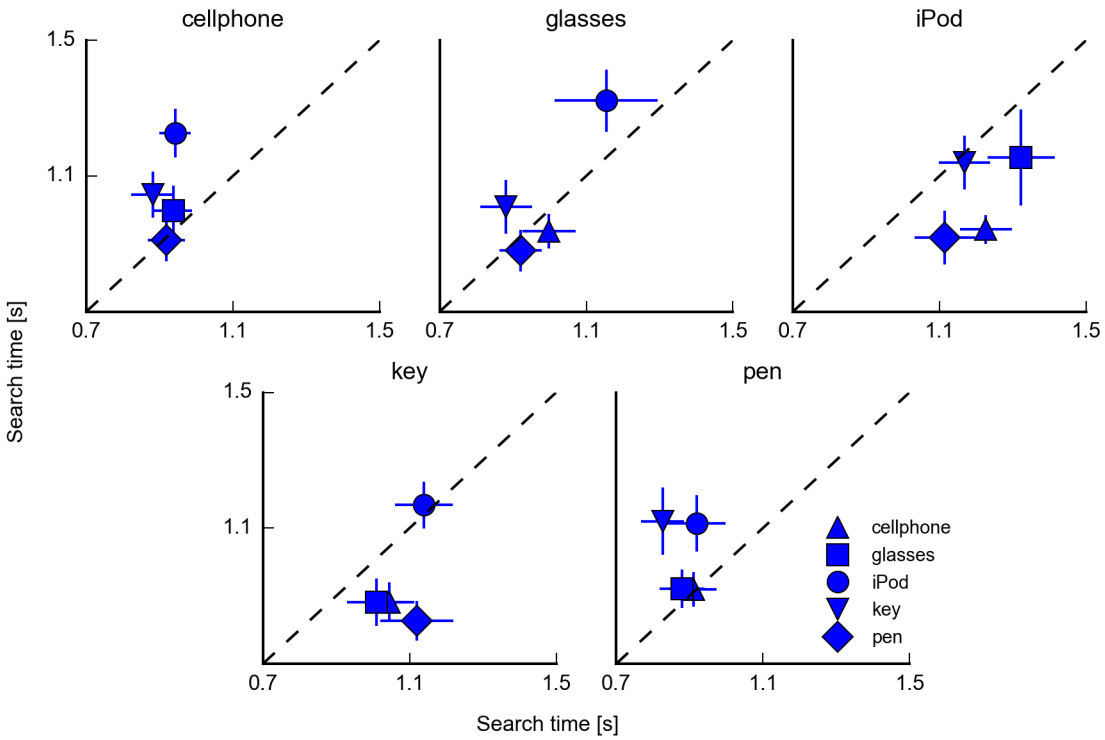

Figure 8: Average search times while searching for different targets in the same image. Each panel compares search time for a particular target category (on the x-axis) to search time for other targets (on the y-axis). Each point represents average search times across a number of images which contained two of the targets. Error bars indicate standard error.

4 Discussion

The purpose of the current study was to determine how clutter affects search for categorical targets in real-world scenes. In particular, we sought to disentangle the effects of extrinsic position uncertainty (i.e., search set size) from those due, through the modulating effect of clutter, to intrinsic position uncertainty (Semizer & Michel, 2017). Our results exhibited several trends:

First, search times increased significantly as the amount of clutter increased. This pattern was evident across different target categories, but the effect was larger for some targets than others. Second, search times increased significantly as the number of possible target locations increased. Additionally, the analysis of observers' fixation distributions showed that observers seemed to fixate frequently at the cued locations. These two sets of findings provide evidence that our manipulation of set size, or extrinsic position uncertainty, was successful.

Third, the stimulus set contained images in which different observers searched for different targets. This gave us an opportunity to directly examine the effect of target category on search performance when the amount of clutter and the relevant set size were held constant. Our findings showed that search times differed depending on the target category. Although the results of the main analysis did not reveal a

significant main effect of target category, this trend should be examined with future studies with more power. The difference in search performance across target categories could be due to differences in the size of search targets. Further analysis revealed that search times tended to decrease significantly as target size increased, both when the size was measured as the area of the circumscribing polygon and when it was measured as the length of the longest axis of this polygon. These findings suggest that certain features make some targets less susceptible to clutter than others. Thus, the strength of the relationship between the clutter and the search performance might depend on the particular search target. Revealing the nature of these specific target features requires future research.

Several studies have previously shown that clutter degrades search performance in naturalistic stimuli (e.g., Bravo & Farid, 2004, 2008; Henderson et al., 2009; Neider & Zelinsky, 2011; Rosenholtz et al., 2007). However, there are various ways in which clutter can lead to the observed performance impairments. For example, clutter has been used as a means of manipulating set size in natural scenes because as a scene gets cluttered, the number of potential target locations also increases (Rosenholtz et al., 2005, 2007). Additionally, increased clutter can force observers to consider irrelevant locations during search, increasing the effects of intrinsic position uncertainty (Semizer & Michel, 2017). Finally, clutter can make search harder by obscuring search targets. Adding clutter to a real-world scenes increases the probability that objects will partially or completely occlude one another. In the current study, we controlled for set size and for occlusions of the search target to isolate those effects of clutter that are due to intrinsic position uncertainty.

The results of the current study, obtained using real-world images, are in broad agreement with those of a previous, related study that showed how clutter degrades search performance in synthetic noise displays (Semizer & Michel, 2017). However, the results of the current study differ in one notable respect. Semizer and Michel (2017) reported that the effect of extrinsic position uncertainty diminished at larger set sizes when the searcher was limited by intrinsic position uncertainty. As a result, search performance was similar across cluttered and uncluttered conditions when the relevant set size was large. However, our results showed that search performance was worse in the high clutter condition than in the low clutter condition regardless of the relevant set size. This difference might be due to either of two reasons (or both): First, the images in our experiment were far less cluttered than the synthetic displays created in the lab. When measured using the same clutter metric, the synthetic stimuli from Semizer and Michel (2017) the yielded clutter values of around $\alpha = 4 \times 10^5$, which was several orders of magnitude larger than the clutter values measured for our images (see Figure 2). Second, the relevant set sizes used in our study were much smaller than those in Semizer and Michel (2017). In the current study the set sizes consisted of either 5 or 13, while the set sizes of Semizer and Michel (2017) ranged from a minimum of 37 to a maximum of 817 potential target locations. Indeed, our results are completely consistent with those of Semizer and Michel (2017) when we consider only the smaller set sizes used in that study.

Overall, our results demonstrate that increased clutter reduces performance in searches of real-world scenes, and does so independently of set size. This suggests that the intrinsic position uncertainty of peripheral vision significantly limits searches of real world scenes in the same way it limits searches of synthetic scenes. Therefore, it is important to account for these effects of intrinsic position uncertainty when evaluating and modeling performance in search tasks.

References

Bochud, F. O., Abbey, C. K., & Eckstein, M. P. (2004). Search for lesions in mammograms: statistical characterization of observer responses. *Medical Physics*, *31*(1), 24–36. doi: 10.1118/1.1630493

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bravo, M. J., & Farid, H. (2004). Search for a category target in clutter. *Perception*, *33*(6), 643–652. doi: 10.1068/p5244

Bravo, M. J., & Farid, H. (2008). A scale invariant measure of clutter. *Journal of Vision*, *8*(2008), 23.1–9. doi: 10.1167/8.1.23

Burgess, A. E., & Ghandeharian, H. (1984). Visual signal detection: II. Signal-location identification. *Journal of the Optical Society of America A: Optics and Image Science*, *1*, 906–910. doi: 10.1364/JOSAA.1.000906

Caspi, A., Beutter, B. R., & Eckstein, M. P. (2004). The time course of visual information accrual guiding eye movement decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 13086–90. doi: 10.1073/pnas.0305329101

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*(5), 890–896. doi: 10.3758/s13423-011-0107-8

Cohn, T. E., & Wardlaw, J. C. (1985). Effect of large spatial uncertainty on foveal luminance increment detectability. *Journal of the Optical Society of America A: Optics and Image Science*, *2*, 820–825. doi: 10.1364/JOSAA.2.000820

Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., & Stone, L. S. (2007). Similar Neural Representations of the Target for Saccades and Perception during Search. *Journal of Neuroscience*, *27*(6), 1266–1270. doi: 10.1523/JNEUROSCI.3975-06.2007

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, *62*(3), 425–451. doi: 10.3758/BF03212096

Egeth, H., Atkinson, J., Gilmore, G., & Marcus, N. (1973). Factors affecting processing mode in visual search. *Perception & Psychophysics*, *13*(3), 394–402. doi: 10.3758/BF03205792

Felzenszwalb, P. F., & Huttenlocher, D. P. (2004). Efficient graph-based image segmentation. *International Journal of Computer Vision*, *59*(2), 167–181. doi: 10.1023/B:VISI.0000022288.19776.77

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), 858–73. doi: 10.1167/6.9.1

Henderson, J. M., Chanceaux, M., & Smith, T. J. (2009). The influence of clutter on real-world scene search : Evidence from search efficiency and eye movements. *Journal of Vision*, *9*(1), 1–8. doi: 10.1167/9.1.32

Kleene, N., & Michel, M. (2018). The capacity of trans-saccadic memory in visual search. *Psychological Review*, *125*(3), 391–408. doi: 10.1037/rev0000099

Krumhansl, C. L., & Thomas, E. A. C. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception & Psychophysics*, *21*(3), 269–279. doi: 10.3758/BF03214239

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., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*, *2*(2), 3. doi: 10.1167/2.2.3

Mack, M. L., & Oliva, A. (2004). Computational estimation of visual complexity. Poster presented at the Twelfth Annual Object, Perception, Attention, and Memory Conference, Minneapolis, MN.

Michel, M., & Geisler, W. (2009). Gaze contingent displays: Analysis of saccadic plasticity in visual search. *Society for Information Display Technical Digest*, *40*(1), 911–914. doi: 10.1889/1.3256945

Michel, M., & Geisler, W. (2011). Intrinsic position uncertainty explains detection and localization performance in peripheral vision. *Journal of Vision*, *11*(1), 1–18. doi: 10.1167/11.1.18

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387–91. 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

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

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: searching for targets in evolving complex scenes. *Journal of Vision*, *11*(14), 1–16. doi: 10.1167/11.14.7

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, *155*, 23–36. doi: 10.1016/S0079-6123(06)55002-2

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*(13), 1703–1721. doi: 10.1016/0042-6989(94)90128-7

Palmer, J. (1995). Attention in visual Search: Distinguishing four causes of a set-size effect. *Current Directions in Psychological Science*, *4*(4), 118–123. doi: 10.1111/1467-8721.ep10772534

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, *40*(10-12), 1227–1268. doi: 10.1016/S0042-6989(99)00244-8

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, *2*, 1508–1532. doi: 10.1364/JOSAA.2.001508

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 1136–1169. doi: 10.1167/4.12.12

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*(2), 20. doi: 10.1167/7.2.20

Popple, A. V., & Levi, D. M. (2005). The perception of spatial order at a glance. *Vision Research*, *45*(9), 1085–1090. doi: 10.1016/j.visres.2004.11.008

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

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion, a measure of display clutter. In *Sigchi* (pp. 761–770). Portland, Oregon.

Rosenholtz, R., Li, Y., & Nakano, L. (2007). Measuring visual clutter. *Journal of Vision*, *7*(2), 17.1–22. doi: 10.1167/7.2.17

Semizer, Y., & Michel, M. (2017). Intrinsic position uncertainty impairs overt search performance. *Journal of Vision*, *17*(9), 1–17. doi: 10.1167/17.9.13

Swensson, R. G., & Judy, P. F. (1981). Detection of noisy visual targets: models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics*, *29*(6), 521–534.

doi: 10.3758/BF03207369

Tanner, W. P. (1961). Physiological implications of psychophysical data. *Annals of the New York Academy of Sciences*, *89*, 752–765. doi: 10.1111/j.1749-6632.1961.tb20176.x

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, *113*(4), 766–786. doi: 10.1037/0033-295X.113.4.766

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

Verghese, P. (2012). Active search for multiple targets is inefficient. *Vision Research*, *74*, 61–71. doi: 10.1016/j.visres.2012.08.008

Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, *82*(3), 184–199. doi: 10.1037/0033-295X.82.3.184

Yu, C.-P., Samaras, D., & Zelinsky, G. J. (2014). Modeling visual clutter perception using proto-object segmentation. *Journal of Vision*, *14*(7), 1–16. doi: 10.1167/14.7.4

Zhang, S., & Eckstein, M. P. (2010). Evolution and optimality of similar neural mechanisms for perception and action during search. *PLoS Computational Biology*, *6*(9). doi: 10.1371/journal.pcbi.1000930