Vision Research 116 (2015) 142–151

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Clutter perception is invariant to image size

Gregory J. Zelinsky [a,b,*], Chen-Ping Yu [b]

[a] Department of Psychology, Stony Brook University, USA
[b] Department of Computer Science, Stony Brook University, USA

ARTICLE INFO

Article history:
Received 26 September 2014
Received in revised form 7 April 2015
Available online 14 May 2015

Keywords:
Visual clutter
Proto-objects
Image segmentation
Color clustering
Superpixels
Image size

ABSTRACT

Two experiments evaluated the effect of retinal image size on the proto-object model of visual clutter perception. Experiment 1 had 20 participants order 90 small images of random-category real-world scenes from least to most cluttered. Aggregating these individual rankings into a single median clutter ranking and comparing it to a previously reported clutter ranking of larger versions of the identical scenes yielded a Spearman's $\rho = .953$ ($p < .001$), suggesting that relative clutter perception is largely invariant to image size. We then applied the proto-object model of clutter perception to these smaller images and obtained a clutter estimate for each. Correlating these estimates with the median behavioral ranking yielded a Spearman's $\rho = .852$ ($p < .001$), which we showed in a comparative analysis to be better than six other methods of estimating clutter. Experiment 2 intermixed large and small versions of the Experiment 1 scenes and had participants ($n = 18$) again rank them for clutter. We found that median clutter rankings of these size-intermixed images were essentially the same as the small and large median rankings from Experiment 1, suggesting size invariance in absolute clutter perception. Moreover, the proto-object model again successfully captured this result. We conclude that both relative and absolute clutter perception is invariant to retinal image size. We further speculate that clutter perception is mediated by proto-objects—a preattentive level of visual representation between features and objects—and that using the proto-object model we may be able to glimpse into this pre-attentive world.

1. Introduction

Everyone knows what clutter is; it is the typically negative percept resulting from the disordered organization of an excessive number of objects. Most previous work on clutter has focused on its consequences for task performance. The clearest example of this is the decrease in visual search efficiency accompanying an increase in clutter (Bravo & Farid, 2008; Henderson, Chanceaux, & Smith, 2009; Mack & Oliva, 2004; Neider & Zelinsky, 2011; Rosenholtz, Li, & Nakano, 2007). This effect of clutter on search efficiency led some researchers to suggest that clutter might be used as a surrogate measure of search set size (Neider & Zelinsky, 2011; Rosenholtz, Li, & Nakano, 2007), the number of objects appearing in a search display. This suggestion in turn led to the development of several computational methods for quantifying clutter (e.g., Bravo & Farid, 2008; Lohrenz et al., 2009; Rosenholtz, Li, & Nakano, 2007; van den Berg, Cornelissen, & Roerdink, 2009) so as to predict search efficiency in real-world scenes, stimuli in which a set of discrete objects cannot be defined objectively. A goal of our study is to further evaluate one of these methods for quantifying clutter—the proto-object model of clutter perception (Yu, Samaras, & Zelinsky, 2014).

There is a fundamental relationship between clutter and visual attention. Much attention research has been devoted to identifying those mental processes that can be performed simultaneously, without creating interference or performance costs, and those that result in performance costs when combined. The former have been termed *pre-attentive* and the latter *post-attentive* (or simply, *attentive*), referring to the fact that individual processes must be selected for serial execution so as to avoid incurring costs. This distinction largely shaped the massive literature on visual search (see Wolfe, 1998; for a review), but dates even farther back to the seminal attention studies using dichotic listening paradigms (see Pashler, 1998; for a review). In addition to a small set of basic visual features that can be extracted and used in parallel (Wolfe & Horowitz, 2004), our perception of clutter is likely pre-attentive; we seem able to effortlessly estimate how much "stuff" there is over a region of space (see Alvarez, 2011; for a review). This follows from the fact that clutter perception is likely derived from summary statistics computed over local pooling regions (Rosenholtz, Huang, & Ehinger, 2012; van den Berg, Cornelissen, & Roerdink, 2009), as has been proposed for crowding

* Corresponding author at: Psychology B-240, Stony Brook University, Stony Brook, NY 11794-2500, USA.
E-mail address: Gregory.Zelinsky@stonybrook.edu (G.J. Zelinsky).

http://dx.doi.org/10.1016/j.visres.2015.04.017
0042-6989/

(Balas, Nakano, & Rosenholtz, 2009; Rosenholtz, Huang, Raj, et al., 2012). Setting aside for now the question of what units are actually summed, the assumption is that this summary is obtained pre-attentively and does not involve an actual count of discrete things; visual information is accumulated in parallel and summed to derive the clutter estimate. Indeed, our percept of clutter may be more than just the product of a pre-attentive process; it may be our perception of pre-attention.

Size also matters. Limiting this statement to vision, examples range from older work showing that larger shapes are given greater perceptual weight by the oculomotor system, causing saccades to land closer to bigger objects (Findlay, 1982), to recent work showing that common objects have a canonical size that affects their recognition speed and accuracy (Konkle & Oliva, 2011). Shifting focus from objects to scenes, another goal of our study asks how the retinal size of visual scenes impacts the perception of scene clutter.

Understanding the relationship between retinal image size and clutter perception is important. We live in a very cluttered world, filled with busy city streets, messy desks, and computer screens packed with icons numbering in the dozens if not hundreds. The sizes of these screens, however, have taken two trends. One has been to make computer monitors bigger so as to fit even more things in our field of view. The other has been to decrease the size of these screens so that we can put them in our purses and pockets and carry them wherever we go. This latter trend creates an obvious problem: given that the screens through which we increasingly interact with the world have become smaller, but the number of apps and other icons that we put on these screens has increased or stayed about the same, our world is becoming perceptually compressed into increasingly smaller spaces.

What are the consequences of this compression for our perception of clutter? One possibility is that perceived clutter might increase with decreasing retinal size. Objects in smaller scenes are closer together. If the absolute distance between objects affects perceived clutter, small scenes should be perceived as being more cluttered than larger ones. Another possibility is that retinal size doesn't matter for clutter perception, and that what is important is the number of perceived objects (or proto-objects). If so, decreasing the size of an image should not affect perceived clutter so long as this manipulation is not so drastic as to change the number of objects that are perceived.

2. Experiment 1

To investigate the relationship between retinal image size and perceived clutter we adopt the joint behavioral and computational approach reported recently by Yu, Samaras, and Zelinsky (2014). These authors asked participants to rank order 90 images of random category scenes from least to most cluttered. Note that a time-unlimited clutter ranking task is perfectly suited to the broader goal of our study, to demarcate the boundary between pre-attention and attention by obtaining an explicit estimate of clutter (pre-attention) that is minimally confounded with more goal-directed (attentive) tasks such as search, scene memory, or even free viewing. Yu et al. then modeled this clutter ranking by computing proto-objects for each scene and ordering them based on the number of proto-objects in each. Model success was assessed by correlating these behavioral and computational rankings. In Experiment 1 we used the scenes from Yu et al. to obtain another behavioral ranking of clutter, only these scenes were one-quarter the size of those used in the earlier study. This consistency in both stimuli and task allows the new behavioral clutter ranking to be compared directly to the one from Yu and colleagues so as to determine the effect of retinal image size on clutter perception. Additionally, we test the proto-object model of clutter perception from Yu et al. to determine how well it predicts the effect of changing image size on these behavioral clutter rankings, and compare these predictions to those from other models of clutter perception.

2.1. The proto-object model of clutter perception

Central to our approach is the suggestion that clutter perception can be predicted by how much stuff appears in a scene, where "stuff" is quantified in terms of locally similar image features that become merged into perceptual fragments that we refer to as *proto-objects*. This definition of a proto-object loosely follows the original usage of the term as coined by Rensink and Enns (1995). These authors conceptualized proto-objects as being relatively small and highly volatile clusters of visual features, created by the pre-attentive application of local grouping processes, from which more extensive visual object representations are ultimately built (see also Rensink, 2000). We subscribe to all of these defining properties. Indeed, rather than substantively reconceptualizing what a proto-object is, we see our work as contributing to the further quantification of this construct and its application to real-world objects and scenes.

Several models of attention have appealed to proto-objects in the context of visually complex stimuli. These have taken two basic approaches. One has been to redefine proto-objects as regions of high image salience. For example, Walther and Koch (2006) used a saliency map (Itti & Koch, 2001; Itti, Koch, & Niebur, 1998) to identify salient points in an image, then spread the activation from each back through the intensity, orientation, and color feature maps to obtain a saliency-based segmentation that they referred to as proto-objects (see also Borji, Sihite, & Itti, 2013; Russell et al., 2014, for related approaches). Using this method, Nuthmann and Henderson (2010) compared these salient proto-objects to objects hand labeled from a scene to see which could better describe the *preferred viewing location* (PVL) of behavioral participants performing various scene inspection tasks. They found that proto-objects were less successful than actual objects in describing the PVL effect, at least for proto-objects defined by feature salience. Another approach has been to use color blob detectors (Forssén, 2004) applied directly to unprocessed images (Wischnewski et al., 2009) and video (Wischnewski et al., 2010) to define proto-objects. These proto-objects are then combined with the Theory of Visual Attention (TVA, Bundesen, 1990) to produce a priority map that is used to predict allocations of visual attention.

Our method for deriving proto-objects differs from previous methods in at least two key respects. First, saliency is not considered in our method. We quantify how much stuff there is in a scene, regardless of whether this stuff is salient or not. Second, rather than using blob detectors to segment proto-objects from an image, which at best restricts the shape of proto-objects to coarse elliptical regions, we use more sophisticated image segmentation techniques developed in the computer vision literature. Specifically, we combine superpixel image segmentation (Liu et al., 2011) with a clustering method (Comaniciu & Meer, 2002) to merge featurally-similar superpixels into proto-objects. Note that superpixels themselves are atomic regions of an image containing pixels that are similar in some feature space, but superpixel methods tend to over-segment images. For this reason we treat superpixel segmentation as a preprocessing stage, one that we follow with a merging stage in which neighboring superpixels that are similar in color are combined to create more spatially extended image fragments that we define as proto-objects. Our model then simply counts the number of proto-objects in an image to obtain an estimate of its clutter (Yu, Samaras, & Zelinsky, 2014).

By quantifying the amount of stuff in an image, our proto-object model of clutter perception differs fundamentally from another popular method of quantifying image clutter—the feature congestion model (Rosenholtz et al., 2005; Rosenholtz, Li, & Nakano, 2007). Rather than specifying an image's perceptual fragments, this model computes the local variance of color, luminance, and orientation features in an image. It then builds from this a three-dimensional covariance ellipse, the volume of which is used to estimate visual clutter; the larger the volume the greater the feature congestion and, presumably, the perception of clutter (Rosenholtz, Li, & Nakano, 2007). In recent work Yu, Samaras, and Zelinsky (2014) implemented and compared several models of visual clutter, including the proto-object and feature congestion models, and found that the proto-object model outperformed all of the others in predicting the behavioral clutter ranking of 90 real-world scenes. Moreover, they observed that approaches based on an initial superpixel segmentation of an image generally did better in predicting this clutter ranking than approaches based on feature variability. They concluded that proto-objects, a mid-level perceptual representation residing between features and objects, underlies clutter perception, and may be instrumental in mediating other basic visual tasks. The present work builds directly on this earlier study, asking how the previously obtained clutter ranking will compare to another clutter ranking obtained using smaller versions of the same images, and whether the proto-object model will be as successful in capturing this ranking when evaluated against other models of clutter perception.

2.2. Methods

Because our goal was to compare new behavioral results to those from Yu, Samaras, and Zelinsky (2014), and to use the model introduced there to describe these new results, our behavioral and computational methods followed closely those from the earlier study. Yu, Samaras, and Zelinsky (2014) should therefore be consulted for additional methodological details and demonstrations of model reliability and robustness.

2.2.1. Behavioral

2.2.1.1. Participants. Twenty undergraduate students from Stony Brook University participated for course credit, none of whom participated in the clutter ranking task from Yu, Samaras, and Zelinsky (2014). All participants had normal or corrected-to-normal vision, by self-report, and gave informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2.1.2. Stimuli. Stimuli were the same 90 images of random category real-world scenes used by Yu, Samaras, and Zelinsky (2014), only reduced in size from 800×600 pixels to 200×150 pixels. This re-sizing to one-quarter their original dimension was done using bi-cubic interpolation in Matlab (version R2011a). Scenes were originally selected from the SUN09 image database (Xiao et al., 2010), and Yu, Samaras, and Zelinsky (2014) should be consulted for constraints on this scene selection.

2.2.1.3. Procedure and apparatus. Following the identical procedure from Yu, Samaras, and Zelinsky (2014), participants were told that they would see 90 images, one at a time, and would have to rank order these images from least to most visually cluttered using their own idiosyncratic definition of what constitutes clutter.

This clutter ranking task was controlled by in-house software written in Matlab (version R2011a) and running on a Windows 7 PC. The interface and apparatus, again identical to the one used in Yu, Samaras, and Zelinsky (2014), consisted of two LCD monitors, one on top of the other. Scenes to be ranked for clutter were displayed one at a time on the bottom monitor. In the Yu, Samaras, and Zelinsky (2014) study these scenes filled the entire display and subtended a visual angle of $27^\circ \times 20^\circ$; in the present study these scenes subtended a visual angle of approximately $6.75^\circ \times 5^\circ$ and were centrally positioned when displayed on the monitor. Order of scene presentation was randomized across participants. The top monitor displayed scenes that the participant already ranked for clutter, although only two images from this partial ranking were shown at a time (with the less cluttered scene on the left and the more cluttered scene on the right). The participant's task was to scroll through these clutter-ranked scenes with the goal of finding the two neighboring images between which the new image (from the bottom monitor) should be inserted. After making this insertion, that image would be added to the accumulating list of clutter-ranked scenes and a new unranked scene would be displayed on the bottom monitor. This procedure continued until all 90 scenes were ranked. Participants were also given the opportunity to re-order their ranking at any time during the task. This would be accomplished by selecting a previously ranked image, removing it from the ranked set (causing it to be re-displayed on the bottom monitor), then re-inserting it into a new position in the ranking.

There was a 12-image practice set, during which participants could form their idiosyncratic clutter scale and become familiar with the interface used to assign clutter rankings, followed by the test set of 90 images. In total, this ranking task took 60–90 min, and an experimenter remained in the room with participants to observe whether they were making thoughtful clutter ranking decisions.

2.2.2. The proto-object model

This study used the proto-object model of clutter perception from Yu, Samaras, and Zelinsky (2014). This model consists of two basic stages: A superpixel segmentation stage to obtain image fragments, followed by a clustering and merging stage to assemble these fragments into proto-objects (see Fig. 1). Specifically, entropy rate superpixels (Liu et al., 2011) were obtained for each image using $k = 500$ initial seeds, and the median HSV color was obtained for each superpixel. Mean-shift clustering (Cheng, 1995; Comaniciu & Meer, 2002), using a feature space bandwidth of 2.5, was applied to these superpixel color medians to obtain an optimal number of color clusters in the feature space. Each superpixel was assigned to a color cluster based on its median color similarity, and then merged with adjacent superpixels falling into the same cluster to obtain proto-objects. The number of proto-objects for a given scene was normalized between 0 and 1 by dividing the segmented number by 500, the initial k number of superpixel seeds. This value is then used as the clutter estimate, with a higher normalized value indicating a more cluttered scene. The following sections provide additional details and background regarding the key superpixel segmentation and superpixel clustering stages.

2.2.2.1. Superpixel segmentation. Although early superpixel methods analyzed the eigenspace of intensity values in an image to produce a segmentation (Shi & Malik, 2000; see also Arbelaez et al., 2011), more generally, superpixel segmentation refers to the process of "seeding" an image then iteratively growing the seed coverage based on an objective function that incorporates local feature similarity. This iterative process ends when this objective function is optimized, resulting in an over-segmented partitioning of an image into *superpixels*.

Following the comparative evaluation of different superpixel segmentation methods conducted by Yu, Samaras, and Zelinsky (2014), the entropy rate method (Liu et al., 2011) was adopted and used in the present study. This method initially distributes seeds uniformly over an input image. The number of seeds is

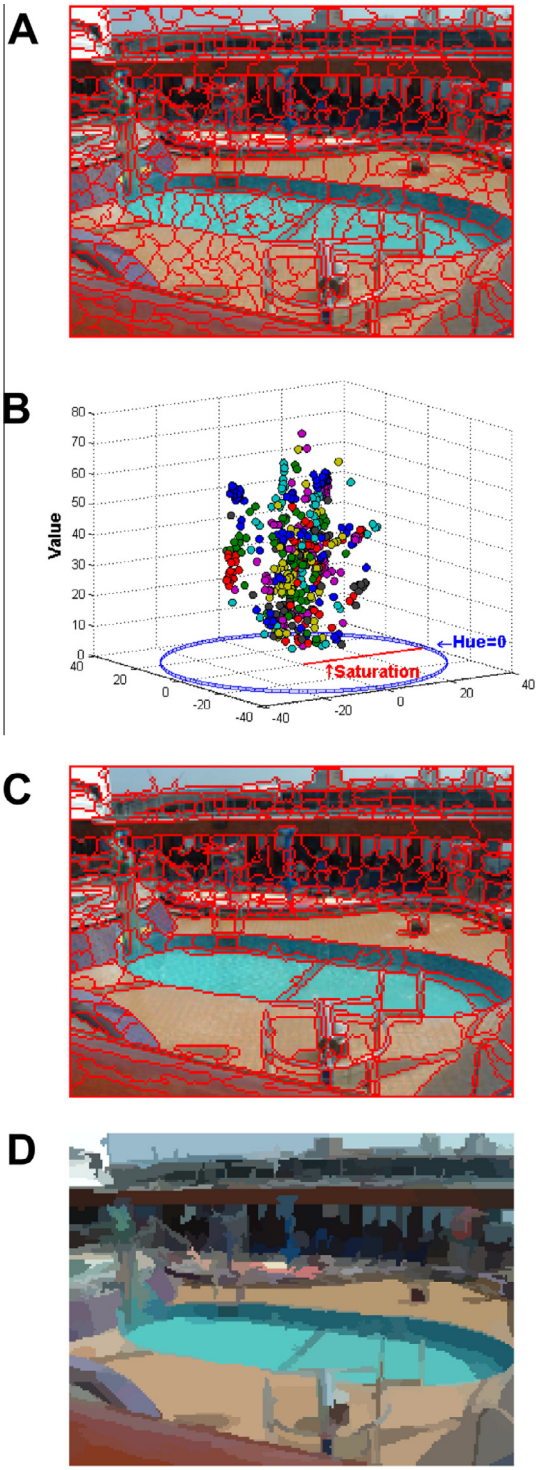

Fig. 1. Illustration of the proto-object model procedure for a representative scene. (A) An entropy rate superpixel segmentation using $k = 500$ seeds. (B) 177 clusters of median superpixel color using a mean-shift feature bandwidth of 2.5 in HSV color space. (C) 359 proto-objects obtained after merging, normalized visual clutter score = 0.719. (D) A visualization of the proto-object segmentation showing each proto-object filled with the median color from the original image pixels.

specified by a user-supplied input parameter, k, and because each seed grows into a superpixel, this parameter determines the number of superpixels that will be extracted from an image. Each seed is then grown by maximizing an objective function that considers edge strengths and local affinity until a stationary segment coverage is achieved. The proto-object model uses superpixel image segmentation as a pre-process because there are two properties that make it undesirable as a direct estimate of clutter: it produces highly over-segmented images (note in Fig. 1A that the relatively uniform water in the pool is segmented into multiple superpixels due to multiple seeds falling within this region) and it requires that the desired number of superpixels be input as a parameter (k). This latter property is particularly undesirable because it is our goal to use the number of fragments segmented from an image as the clutter estimate. For these reasons we therefore include a second clustering stage that uses feature similarity to merge these superpixel image fragments into more spatially extended regions that we call proto-objects.

2.2.2.2. Superpixel clustering. Neighboring superpixels having similar features are merged into proto-objects based on a cluster analysis performed on the color feature space. We use the mean-shift algorithm (Cheng, 1995; Comaniciu & Meer, 2002) to obtain these clusters. Mean-shift clusters data into groups by iteratively shifting every data point to a common density mode, with the data converging to the same density mode forming a cluster. Bandwidth parameters determine the search area for the shift directions. Mean-shift is often used as an image segmentation method, and indeed is one of the methods that we include in our analysis comparing the proto-object model to other approaches. However, the application of mean-shift inside the proto-object model differs from this standard application in that we use the algorithm, not for segmentation, but only for clustering. We apply mean-shift solely to the space of color medians in an image, where each median corresponds to a superpixel. This application means that only the feature-space bandwidth parameter is used; the spatial bandwidth parameter is needed only for segmentation. The algorithm returns the optimal number of color clusters in this space, which, again following Yu, Samaras, and Zelinsky (2014), was HSV in the present study. Superpixels are then assigned to clusters based on the median color obtained for a given image fragment in HSV feature space. From these assignments, superpixels falling into the same color cluster and sharing a boundary (i.e., neighbors) are finally merged into a larger proto-object.

2.3. Results and discussion

We evaluate the relationship between image size and clutter perception in three ways. First, we compare the behavioral clutter ranking obtained for the large 800×600 pixel images by Yu, Samaras, and Zelinsky (2014) to the clutter ranking of the smaller 200×150 pixel images from the present study. This comparison will give us our first answer to how visual clutter perception changes with retinal image size. Second, we compare the clutter ranking of the small images to a rating of the same images obtained from the proto-object model of clutter perception. This comparison will tell us whether predictions from this model, and, more generally, the proto-object approach to quantifying clutter, can generalize to smaller images. Third, we evaluate how well the proto-object model compares to other clutter estimation methods in predicting the clutter ranking of small images, and how this evaluation differs from the model comparison made by Yu, Samaras, and Zelinsky (2014). This comparison will tell us the model that should be preferred when estimating clutter perception across images of different sizes.

2.3.1. Evaluating the effect of scene size on behavioral clutter rankings

Does the retinal size of a scene affect its position in a ranking of similarly-sized scenes for visual clutter? To answer this question we first obtained a single representative scene ranking from the 20 individual rankings from our participants. Yu, Samaras, and Zelinsky (2014) did this for the 800×600 pixel versions of these

scenes by finding the median of each image's ranked position in the ordered set, and we followed an identical procedure for the smaller 200 × 150 pixel images used in the present study. Ordering these medians from least to most cluttered produced a single ranking, which we then correlated with the comparable median ranking from Yu, Samaras, and Zelinsky (2014). As shown in Fig. 2A, these clutter rankings correlated highly, yielding a Spearman's ρ of 0.953 ($p < .001$). This near perfect correlation suggests that the perception of clutter between scenes is highly invariant to retinal image size. Regardless of whether these scenes were viewed as large or small, participants ranked them similarly for visual clutter. Moreover, because reducing the size of an image is equivalent to a form of blurring, this finding also suggests that visual clutter perception is invariant to blur, at least at the levels explored in this study. The fact that retinal image size decreases with increasing viewing distance means that same-sized scenes viewed from different distances might also be ranked similarly for clutter, although this specific manipulation was not tested.

2.3.2. Evaluating the proto-object model as a predictor of clutter in small scenes

The proto-object model from Yu, Samaras, and Zelinsky (2014) proved to be a good predictor of clutter rankings in the large 800 × 600 pixel scenes used in that study; will this model be as successful in predicting clutter rankings in smaller versions of these images? To answer this question, we computed a proto-object segmentation for the 90200 × 150 pixel images used in the present study, counted the number of proto-objects in each, then normalized this count by dividing it by the number of initial superpixels ($k = 500$) to obtain a clutter rating for each scene. Correlating these model ratings with the behaviorally-derived median ranking yielded a Spearman's $\rho = .852$ ($p < .001$), using $k = 500$ superpixel seeds and a mean-shift feature bandwidth parameter of 2.5. Those scenes that were judged by participants to be least (most) cluttered tended also to be estimated as least (most) cluttered by the model (Fig. 2B). Moreover, this result is comparable to the $\rho = .814$ correlation reported by Yu, Samaras, and Zelinsky (2014) for the larger scenes, demonstrating that the proto-object model of clutter perception is robust to changes in retinal image size. Fig. 3 shows representative proto-object segmentations for 2 (of the 90) scenes at both large and small image sizes.

It should be noted that the good agreement between the clutter ranking from the proto-object model and the ranking from our participants was based on parameters optimized over the entire image set. This has implications for the conclusions that can be drawn, as the high $\rho = .852$ correlation might reflect a form of training accuracy rather than true prediction. To determine whether the model's performance would generalize to new scenes we performed 10-fold cross validation on our clutter-ranked images. Specifically, we used 90% of the images for training and 10% for testing, and repeated this procedure 10 times using different random splits. Averaging over these 10 tests produced a correlation of .827, numerically lower than the correlation obtained without cross validation but still highly significant ($p < .001$). This finding suggests, not only that the proto-object model is able to predict the clutter ranking of small images, but also that its predictive success is likely to generalize to new image sets that might be used in future experimentation, at least with respect to random categories of real-world scenes. Note also that a similar cross validation performed by Yu, Samaras, and Zelinsky (2014) on the larger 800 × 600 pixel scenes yielded an average correlation of .74 ($p < .001$), suggesting that the proto-object model might be an even better predictor of clutter as image size decreases.

Aside from our choices of superpixel segmentation method (entropy rate superpixels) and color feature space (HSV), both of which were informed by Yu, Samaras, and Zelinsky (2014), the proto-object model of clutter perception has only two parameters: the number of initial superpixel seeds (k) and the feature bandwidth for mean-shift clustering. The Spearman's correlation of .852 (.827 after cross validation) was based on the selection of optimal seed and bandwidth parameters, which begs the question of how robust the model is to different parameter settings. Table 1 answers this question by showing 45 correlations spanning nine levels of superpixel seeds (ranging from 300 to 700) and five feature bandwidths (ranging from 1 to 7).[1] Even with the least desirable setting of parameters, 300 superpixel seeds and a bandwidth of 7, the proto-object model still produced a .713 correlation with the behavioral clutter rankings. Taken together with a similar observation by Yu, Samaras, and Zelinsky (2014), the clear conclusion from this analysis is that the predictive success of the proto-object model, regardless of whether the model is applied to large or small images, is highly robust to changes in its parameters.

2.3.3. Evaluating the effect of scene size on models of clutter perception

At its best, the proto-object model produced a Spearman's correlation of .852 (.827), but how does this compare to other methods of estimating clutter after similar parameter optimization (when applicable)? To answer this question we obtained clutter estimates of our 90 200 × 150 pixel scenes from six other clutter models, then correlated these estimates with the median behavioral clutter ranking, just as we did when evaluating the proto-object model. The models included in this comparative analysis were: the feature congestion model (Rosenholtz, Li, & Nakano, 2007), the power law model (Bravo & Farid, 2008), the C3 model (Beck, Lohrenz, & Trafton, 2010; Lohrenz et al., 2009), an edge density model (Mack & Oliva, 2004), mean-shift image segmentation (Comaniciu & Meer, 2002), and a graph-based segmentation method (Felzenszwalb & Huttenlocher, 2004). These latter two models, although not really models of clutter perception, use image segmentation methods to merge pixels into larger image fragments. Clutter predictions can therefore be easily generated simply by counting the number of segmented fragments, similar to what was done for the proto-object model. Yu, Samaras, and Zelinsky (2014) should be consulted for descriptions of each of these models and methods, and for implementation details.

Fig. 4 shows the results of this model comparison. Plotted are Spearman's ρ correlations for each of the seven evaluated methods applied to the 200 × 150 pixel scenes, together with the corresponding correlations re-plotted from Yu, Samaras, and Zelinsky (2014) obtained from the 800 × 600 pixel scenes. For each method, care was taken to optimize any and all parameters so as to achieve the best possible correlation. Focusing first on the data from the smaller images, although the proto-object model yielded the highest correlation out of all the methods tested, several other models performed nearly as well. Specifically, mean-shift segmentation, the power law model, and the graph-based model all produced correlations of .80 or higher. Notably, all of these best performing methods included an initial superpixel segmentation stage, whereas all of the poorer performing methods, edge density, feature congestion, and the C3 model, did not. This suggests that clutter perception may engage a process that spatially groups feature information into perceptual fragments.

[1] Note that the values in this table are correlations between the noise-free proto-object model and the median clutter ranking from participants. Variability therefore does not exist in these values, making it impossible to conduct satisfying statistical tests to determine if one is significantly different from another. However, our conclusions do not require us knowing whether differences exist among these values. Indeed, the fact that these correlations are all high bolsters our point that any differences, should they exist, are not likely to be meaningful.

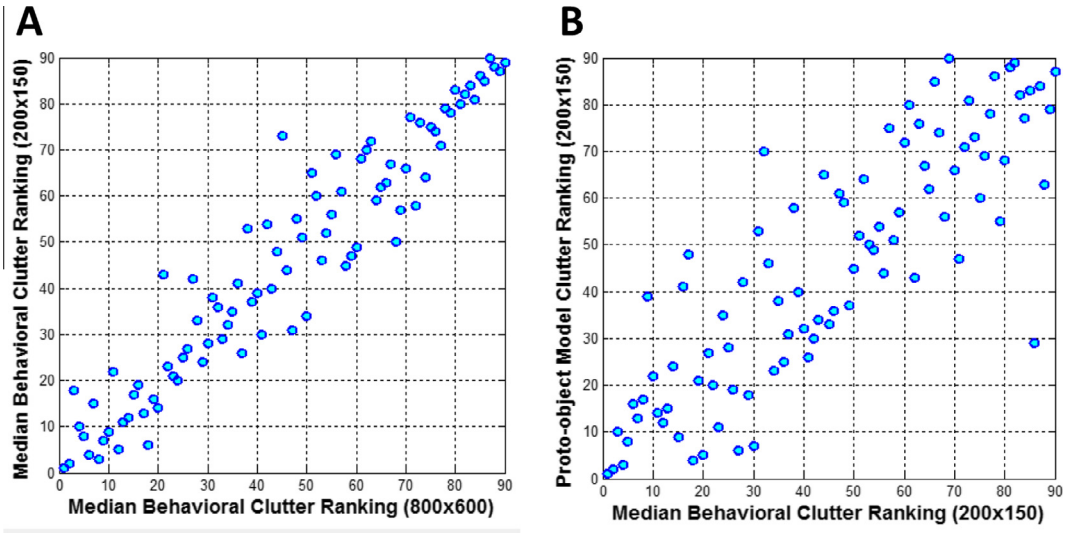

Fig. 2. (A) The median behavioral clutter ranking of the 90 scenes viewed at a resolution of 200 × 150 pixels (present study) plotted as a function of the median behavioral clutter ranking of the same 90 scenes viewed at a resolution of 800 × 600 pixels (Yu, Samaras, & Zelinsky, 2014). Spearman's ρ = .953. (B) The median behavioral clutter ranking of the 90 200 × 150 pixel scenes plotted as a function of the clutter ranking from the proto-object model for the same 200 × 150 pixel scenes. Spearman's ρ = .852.

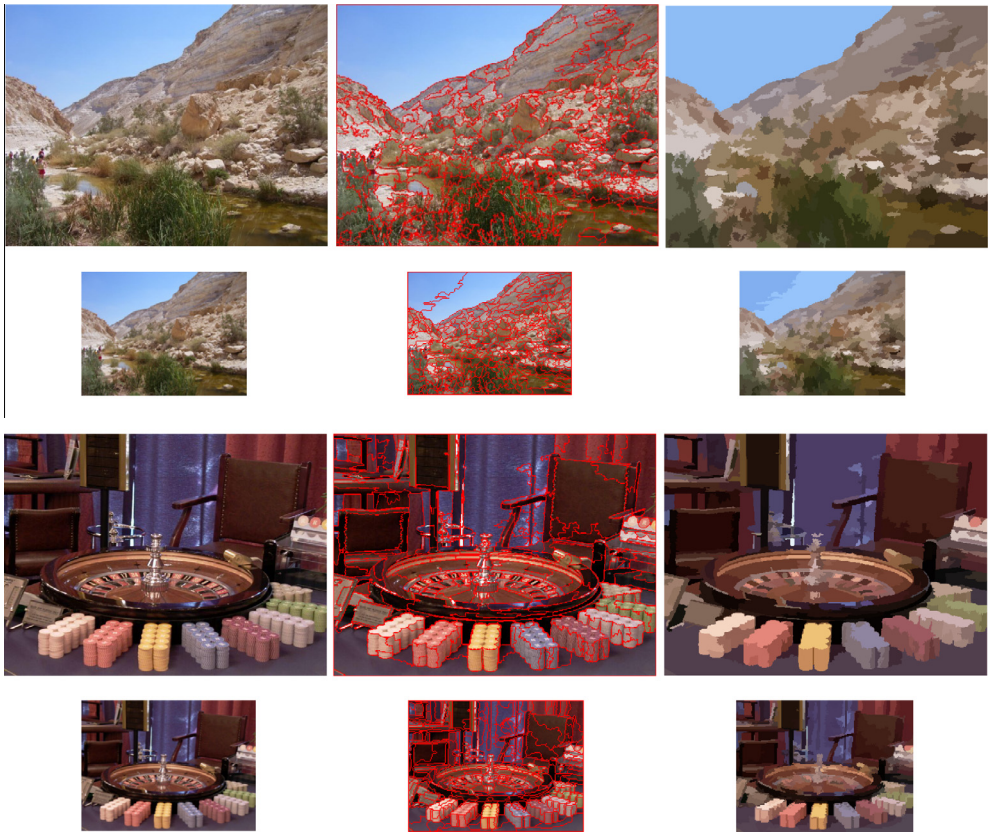

Fig. 3. Examples of 2 of the 90 images used in this study (*left column*), shown with their proto-object segmentations (*middle column*) and reconstructions obtained after filling each proto-object with its median color (*right column*). *First and third rows*: segmentations and reconstructions from Yu, Samaras, and Zelinsky (2014) using 800 × 600 pixel images. *Second and fourth rows*: corresponding segmentations and reconstructions from the present study using images scaled to one-quarter their size, shown slightly larger here for better visibility. Results are based on entropy rate superpixel segmentation with 600 (large images) and 500 (small images) initial seeds, and mean-shift clustering bandwidths of 4 (large images) and 2.5 (small images) within an HSV color feature space.

How does image size affect the performance of these models? One clear trend from Fig. 4 is that all of the evaluated methods performed better for the smaller images compared to the larger, and for some methods this difference was dramatic. Fig. 5 quantifies these differences as the percentage drop in Spearman's ρ resulting from application of a method to large images relative to its small image correlation ($[1 - \text{large}_\rho/\text{small}_\rho] \times 100$). This analysis shows that the proto-object model's performance on the 200 × 150 pixel image set declined by less than 5% after application to the larger 800 × 600 pixel images. The same cannot be said for the other

Table 1
Spearman's correlations between the median behavioral clutter ranking and rankings from the proto-object model as a function of the number of initial entropy rate superpixel seeds and mean-shift feature bandwidth.

Mean-shift	# Of superpixel seeds								
Bandwidth	**300**	**350**	**400**	**450**	**500**	**550**	**600**	**650**	**700**
1	.717	.729	.740	.723	.727	.718	.734	.724	.720
2.5	.832	.836	.848	.833	**.852**	.834	.846	.846	.831
4	.846	.825	.808	.842	.813	.833	.797	.797	.826
5.5	.764	.799	.785	.796	.784	.803	.774	.761	.762
7	.713	.750	.763	.757	.751	.734	.735	.749	.770

Note: The highest correlation is indicated in bold.

evaluated methods, which showed performance drops in the 13–22% range. We know from Fig. 2A that clutter ranking is highly invariant to changes in image size. A model of clutter perception should be able to capture this size invariance to clutter. Only the proto-object model is able to do this. Out of all the models included in this comparative evaluation, not only is the proto-object model the best predictor of clutter perception in both large and small images, it is also the model that was least affected by changing image size.

3. Experiment 2

Experiment 1 demonstrated that *relative* clutter perception is invariant to image size—small images were ranked for clutter the same as larger images. This need not have been the case. It could have been that different-sized images might have engaged different-scale filters used to extract spatial information, resulting in different clutter rankings. Our data argue against this possibility. Image size did not affect how cluttered one image appeared relative to another.

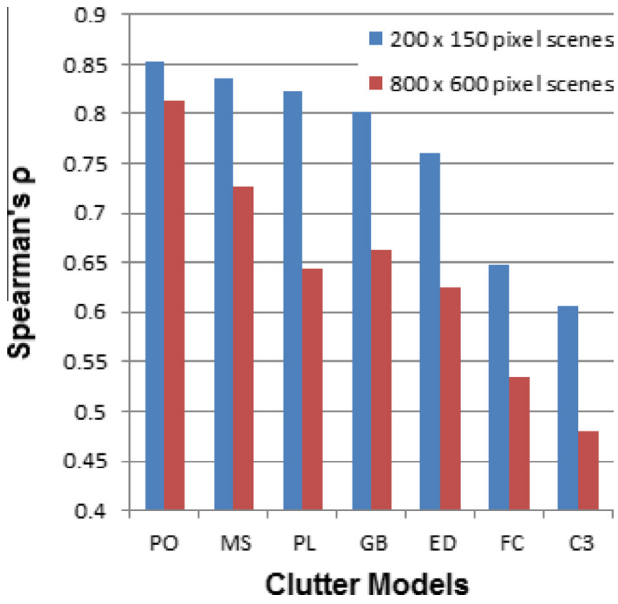

Fig. 4. Spearman's correlations between the median behavioral clutter ranking and ratings from seven clutter prediction methods for the 200 × 150 pixel scenes (blue bars, present study) and the 800 × 600 pixel scenes (red bars, from Yu, Samaras, & Zelinsky, 2014). Results are ordered from most correlated (left) to least correlated (right) based on the 200 × 150 pixel image set. PO: proto-object clutter model (Yu, Samaras, & Zelinsky, 2014). MS: mean-shift image segmentation (Comaniciu & Meer, 2002). PL: power law clutter model (Bravo & Farid, 2008). GB: graph-based image segmentation (Felzenszwalb & Huttenlocher, 2004). ED: edge density (Mack & Oliva, 2004). FC: feature congestion clutter model (Rosenholtz, Li, & Nakano, 2007). C3: color clustering clutter model (Lohrenz et al., 2009). All $p < .001$.

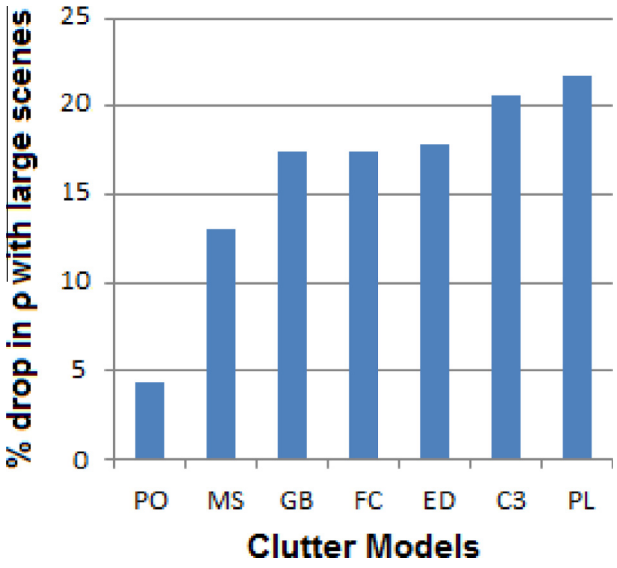

Fig. 5. Percent drop in the Spearman's correlations between predicted and behavioral rankings for the 800 × 600 pixel scenes relative to the correlations for the 200 × 150 pixel scenes for each of the clutter prediction methods tested.

But what of *absolute* clutter perception—are smaller images perceived as more or less cluttered than larger ones? In some sense this is the more interesting question, one that might even impact the type of content that we view on our devices. For example, content providers may make viewing recommendations depend on the device used to view the content. If an action movie will appear cluttered when viewed on a phone, the viewing experience may be more enjoyable if you save the action movie for a large screen and view a sitcom on the phone instead.

Interestingly, the relationship between retinal image size and absolute clutter perception is unknown, and a case can be made for an effect appearing in either direction. To the extent that clutter perception depends on the number of discrete "things" in an image, then as an image gets smaller so too might the number of things. Small things might dissolve into just "stuff". This would be expected to produce a positive relationship between image size and clutter perception, with smaller images tending to be perceived as less cluttered. The opposite relationship might also exist. To the extent that clutter perception is affected by the absolute distance between neighboring objects, this distance will necessarily be greater for larger images. This would result in an inverse relationship between image size and clutter—as image size increases objects spread farther apart resulting in the scene appearing less cluttered. Of course finding no difference between large and small images would suggest that absolute clutter perception, like relative clutter perception, is invariant to image size.

The goal of Experiment 2 was to discover the nature of the relationship between retinal image size and absolute clutter, and to determine if the proto-object model could account for this relationship. We did this by again asking participants to produce a clutter ranking of the same 90 scenes used in Experiment 1, only now these scenes were a mixture of large and small images (one-quarter the size). If smaller images are perceived as more cluttered, they will tend to be clustered toward the "more cluttered" end of the ranking; if smaller scenes are perceived as less cluttered they will be clustered toward the "less cluttered" end. If image size is unrelated to clutter perception, the prediction is that a ranking of intermixed large and small images should not differ from a ranking of only large or only small images.

3.1. Methods

Eighteen undergraduate students from Stony Brook University participated for course credit, none of whom participated in Experiment 1 or the clutter ranking task from Yu, Samaras, and Zelinsky (2014). All participants had normal or corrected-to-normal vision, by self-report, and gave informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Stimuli were the same 90 images of random category real-world scenes used by Yu, Samaras, and Zelinsky (2014), with the only difference being that half of the scenes, selected at random, were reduced in size from 800×600 pixels (large) to 200×150 pixels (small), the same reduction to one-quarter size used in Experiment 1.

The procedure was identical to what was described for Experiment 1, participants were told that they would see 90 images, one at a time, and would have to rank order these images from least to most visually cluttered using their own idiosyncratic definition of what constitutes clutter. Order of scene presentation was randomized both within and across participants, meaning that small and large images appeared unpredictably in the presentation sequence and varied from participant to participant. Image size was counterbalanced across participants; half of the participants viewed half of the scenes as small, while the other participants viewed the same scenes as large.

The apparatus used to collect the clutter rankings was also identical to the one used by Yu, Samaras, and Zelinsky (2014) and in Experiment 1, only now the viewing conditions were a mixture of both visual angles. Scenes to be ranked for clutter were displayed one at a time on the bottom monitor of the two-monitor setup, with the large images filling the screen and subtending a visual angle of $27^\circ \times 20^\circ$ (as in Yu, Samaras, & Zelinsky, 2014) and the small images centered on the screen and subtending a visual angle of $6.75^\circ \times 5^\circ$ (as in Experiment 1). All other aspects of the apparatus and its functionality were identical to what was described for Experiment 1.

There was a 12-scene practice set consisting of a randomly interleaved mixture of six large and six small images, followed by the 90-scene test set. This experiment also took 60–90 min, and an experimenter again remained in the room with the participants.

3.2. Results and discussion

We predicted that if small and large images are perceived differently with respect to absolute clutter, an intermixed set of these images should be ranked differently for clutter when compared to a ranking of either all large or all small images. Specifically, the small images from this intermixed set should be ranked differently from the images in the all-small set (Experiment 1) due to the opportunity now for absolute clutter comparisons to the larger images. The same logic holds true for the large images when compared to the ranking from Yu, Samaras, and Zelinsky (2014). A final prediction is that the ranking of the small images from Experiment 2 should correlate weakly with the ranking of the large images from Experiment 2, as the two would have been perceived differently with respect to absolute clutter. Finding evidence for any or all of these predictions would be support for absolute clutter perception being dependent on the size of an image. However, finding uniformly strong correlations across these three comparisons would be support for the opposite conclusion, that absolute clutter perception is also invariant to image size.

To test these predictions we first had to obtain a median clutter ranking of the Experiment 2 images for comparison to the others, and this required segregating the intermixed images into large and small rank-ordered sets. Recall that, across participants, each of the scenes used in Experiment 2 was ranked for clutter when it was viewed as large (by nine participants) and when it was viewed as small (by nine participants). This made it possible to create two median clutter rankings of the 90 scenes, one for large images and the other for small images. Correlating the small-image median ranking with the ranking from Experiment 1 yielded a Spearman's $\rho = .920$ ($p < .001$). Correlating the large-image ranking with the ranking of large images from Yu, Samaras, and Zelinsky (2014) yielded a similar result, Spearman's $\rho = .932$ ($p < .001$). The fact that these two correlations are essentially the same, and both nearly perfect, suggest that it does not matter whether images were viewed as large, small, or a mixture of the two, all were ranked similarly for clutter. We also compared the large-image ranking to the small-image ranking and, as expected from the previous result, found that the two were highly correlated, Spearman's $\rho = .887$ ($p < .001$). Collectively, these analyses argue strongly for absolute clutter perception, like relative clutter perception, being largely invariant to retinal image size.

Parallel analyses were performed to evaluate whether the proto-object model would show a similar invariance to absolute differences in image size. We again used the small-image and large-image rankings obtained from Experiment 2, but instead of correlating these to other behavioral rankings we now correlated them to the ratings of small and large images obtained from the proto-object model. This was done using the optimal superpixel seed and mean-shift bandwidth settings identified by Yu, Samaras, and Zelinsky (2014) for the large images ($k = 600$, bandwidth = 4) and in Experiment 1 for the small images ($k = 500$, bandwidth = 2.5). Doing this, we found that the clutter ratings from the proto-object model correlated highly with both the small-image ranking (Spearman's $\rho = .761$, $p < .001$) and the large-image ranking (Spearman's $\rho = .769$, $p < .001$). Moreover, the fact that these correlations are essentially identical argues that the proto-object model, like the behavioral perception of absolute clutter, is largely invariant to changes in image size.

We also explored the possibility that participants ranking the intermixed images may have adopted a single setting of their clutter "parameters", ones tuned for either large images or small images, and did not switch back and forth between different settings depending on the size of the image being ranked. To evaluate the effect of using fixed settings we correlated the small and large-image rankings from Experiment 2 with ratings from proto-object models trained using the non-corresponding parameter settings. We found that a proto-object model optimized for small images ($k = 500$, bandwidth = 2.5) still correlated highly with the large-image median ranking (Spearman's $\rho = .790$, $p < .001$), as did a proto-object model optimized for large images ($k = 600$, bandwidth = 4) when correlated with the small-image ranking (Spearman's $\rho = .787$, $p < .001$). These correlations are comparable to, and certainly not lower than, the correlations found when parameters were optimized based on image size. This suggests that the above-demonstrated invariance of the proto-object model to changes in absolute image size does not depend on parameter optimization. To the extent that our behavioral participants had comparable parameters, it also suggests that it may not have mattered if they used a fixed setting of these parameters, either for large or small images, or toggled back and forth between the two, their perception of clutter would have been essentially the same.

4. General discussion

In this study we evaluated whether the proto-object model could account for effects of image size on clutter perception. We

all know clutter when we see it, but does this perception of clutter change when images are viewed as large or small? This question was addressed in two experiments, both of which used a clutter ranking task. In Experiment 1 we had participants rank order small images of scenes for visual clutter, then correlated their median clutter ranking with a similarly obtained ranking of larger versions of the same scenes from Yu, Samaras, and Zelinsky (2014). The result was surprising in its clarity; a near perfect correlation. It did not matter whether the scenes were viewed as large or reduced to one quarter in size, they were ranked for clutter nearly the same. This means that *relative* clutter perception is invariant to retinal image size, as are the operations underlying relative clutter judgments. In Experiment 2 we intermixed large and small images and again asked participants to rank order the images for perceived clutter. The task therefore required absolute judgments of clutter across images of different sizes. Here, too, the results were clear. Regardless of whether the images were all small, all large, or size-intermixed, participants ranked them for clutter much the same, suggesting that *absolute* clutter perception, like relative clutter perception, is also invariant to retinal image size. In addition to adding to our basic understanding of clutter, these findings also have applied significance. They suggest that it may not matter whether image content is viewed on a small cell phone or tablet or a large monitor or television, clutter perception depends on the image content and not the medium used to view this content.

These findings also have implications for attention, and in particular the things that attention might select. If clutter perception reflects the amount of stuff over some region of space, what is actually being summed to compute this percept? Most models of clutter assume that this stuff can be quantified in terms of a set of basic visual features, with those used to construct edges certainly included in this set (Henderson, Chanceaux, & Smith, 2009; Mack & Oliva, 2004). The summary would therefore be an estimate of the number of features in this space, with features being the presumed units that can increase in number and result in the perceptual congestion that we experience as clutter (Rosenholtz, Li, & Nakano, 2007). Objects are another potential unit of attention (Scholl, 2001), so presumably the summary could also be of these. While not excluding either of these possibilities, our data suggest a third alternative, that there is another pre-attentive unit existing between features and objects, proto-objects, and that these are the units that are summed to derive the clutter percept.

This conclusion follows from the fact that the proto-object model of clutter perception introduced in Yu, Samaras, and Zelinsky (2014) was successful in predicting the clutter ranking of scenes from our behavioral participants. The model builds on a superpixel segmentation of an image by adding a process that merges neighboring superpixels that share a common color cluster. The products of this merging are spatially extended regions of coherent features—proto-objects. In Experiment 1 we applied the proto-object model to a set of small images ranked by participants and counted the number of proto-objects in each to obtain clutter estimates. We found a .852 correlation between these estimates and the median behavioral clutter ranking, with this correlation remaining high even after cross validation (.827). In Experiment 2 we showed similarly high correlations between the estimates of the proto-object model and small-image and large-image median rankings of size-intermixed images. Taken together with the findings reported by Yu, Samaras, and Zelinsky (2014), we can conclude that the proto-object model nicely captures the size invariance to clutter found for both relative and absolute clutter perception judgments, and is the only existing model of clutter estimation that is able to do so.

Although the proto-object model is not a model of attention per se, it informs and constrains these models in a basic respect. Given that attention can only select what is pre-attentively available, we speculate that proto-objects may be the smallest unit of visual space that is available for selection and use by the higher-level routines mediating goal-directed behavior. As such, understanding proto-objects may be fundamental to understanding attention. For one, constraints on the size of proto-objects, or the merging of image fragments into proto-objects, may determine the resolution of visual spatial attention. With the proto-object model we will be able to form testable predictions of what this resolution should be for stimuli ranging from simple patterns to fully realistic scenes. Secondly, if proto-objects constitute the set of things that are available for selection, then the spatial distribution of proto-objects in an image may constrain how attention can be allocated during scene viewing. Stated differently, if proto-objects are the stuff that is preattentively available for selection, the output of the proto-object model might be the input to models of visual selective attention. Future work will test this conjecture by using the proto-object model to predict the allocation of gaze during scene viewing and visual search. It may be the case that the overt allocation of attention is best explained, not by feature-based models (Borji, Sihite, & Itti, 2013; Elazary & Itti, 2008) or object-based models (Einhäuser, Spain, & Perona, 2008; Nuthmann & Henderson, 2010) but rather by models that assume competition on a priority map between proto-objects.

Finally, we interpret our results as further evidence for clutter perception being mediated by a mid-level visual representation, one that we are capturing with our proto-object model. Mid-level visual representations live in the fertile ground between low-level visual features and high-level objects, and evidence from multiple perspectives is mounting for the importance of these representations.[2] We believe that proto-objects represent a key step in the neurocomputational transformation of features into objects, loosely corresponding to the formation of simple closed-form shapes that is thought to occur in cortical area V4 (Cadieu et al., 2007; Pasupathy & Connor, 2002). These visual fragments are the "stuff" of our perception, and we speculate that some percepts and perceptual decisions can be mediated directly from these proto-object representations. We contend that clutter is one of these percepts, and that clutter estimation is one of these perceptual decisions. Clutter estimation is potentially crucial to many everyday tasks. Consistent with the suggested relationship between clutter and a search set size (Neider & Zelinsky, 2011; Rosenholtz, Li, & Nakano, 2007), it provides a rough idea of the size of a search space; an estimate of the difficulty in selecting some particular piece of information from the palate of pre-attentive pieces residing at the mid-level of vision. Future work will be aimed at discovering other behaviors that can be predicted directly from proto-objects, and how these predictions might depend on summary features and spatial statistics computed from these mid-level visual representations. If the pre-attentive world consists of proto-objects, the proto-object model might enable us to glimpse into this world.

Acknowledgments

This work was supported by NIH Grant R01-MH063748 and NSF Grants IIS-1111047 and IIS-1161876 to G.J.Z. We thank Dimitris Samaras and the members of the Eye Cog lab for invaluable discussions.

[2] See http://www.visionsciences.org/symposium_detail.php?year=2014&id=3 for a member-initiated symposium on this topic held at the 2014 meeting of the Vision Sciences Society.

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Science, 15*, 122–131.

Arbelaez, P., Maire, M., Fowlkes, C., & Malik, J. (2011). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 33*(5), 898–916.

Balas, B. J., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12), 1–18. 13.

Beck, M. R., Lohrenz, M. C., & Trafton, J. G. (2010). Measuring search efficiency in complex visual search tasks: Global and local clutter. *Journal of Experimental Psychology: Applied, 16*(3), 238.

Borji, A., Sihite, D. N., & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing, 22*, 55–69.

Bravo, M. J., & Farid, H. (2008). A scale invariant measure of clutter. *Journal of Vision, 8*(1), 1–9.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*(4), 523–547.

Cadieu, C., Kouh, M., Pasupathy, A., Connor, C. E., Reisenhuber, M., & Poggio, T. (2007). A model of V4 shape selectivity and invariance. *Journal of Neurophysiology, 98*(3), 1733–1750.

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17*(8), 790–799.

Comaniciu, D., & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*(5), 603–619.

Einhäuser, W., Spain, M., & Perona, P. (2008). Objects predict fixations better than early saliency. *Journal of Vision, 8*(14), 1–26. 18.

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision, 8*(3), 1–15. 3.

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

Findlay, J. M. (1982). Global visual processing for saccadic eye movements. *Vision Research, 22*, 1033–1045.

Forssén, P. E. (2004). *Low and medium level vision using channel representations* (Doctoral dissertation, Linköping).

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.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194–203.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 20*(11), 1254–1259.

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 23–37.

Liu, M. Y., Tuzel, O., Ramalingam, S., & Chellappa, R. (2011). Entropy rate superpixel segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2097–2104). IEEE.

Lohrenz, M. C., Trafton, J. G., Beck, M. R., & Gendron, M. L. (2009). A model of clutter for complex, multivariate geospatial displays. *Human Factors, 51*(1), 90–101.

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

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.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *Journal of Vision, 10*(8), 1–19. 20.

Pashler, H. (1998). *The psychology of attention*. MIT Press.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience, 5*, 1332–1338. News and Views, 1252–1254.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition, 7*, 17–42.

Rensink, R. A., & Enns, J. T. (1995). Preemption effects in visual search: Evidence for low-level grouping. *Psychological Review, 102*, 101–130.

Rosenholtz, R., Li, Y., Mansfield, J., & Jin, Z. (2005). Feature congestion: A measure of display clutter. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 761–770). ACM.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision, 12*(4), 1–17. 14.

Rosenholtz, R., Huang, J., & Ehinger, K. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology: Consciousness Research, 3*(12), 1–15.

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

Russell, A. F., Mihalaş, S., von der Heydt, R., Niebur, E., & Etienne-Cummings, R. (2014). A model of proto-object based saliency. *Vision Research, 94*, 1–15.

Scholl, B. (2001). Objects and attention: The state of the art. *Cognition, 80*, 1–46.

Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(8), 888–905.

van den Berg, R., Cornelissen, F. W., & Roerdink, J. B. (2009). A crowding model of visual clutter. *Journal of Vision, 9*(4), 1–11.

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks, 19*(9), 1395–1407.

Wischnewski, M., Belardinelli, A., Schneider, W. X., & Steil, J. J. (2010). Where to look next? Combining static and dynamic proto-objects in a TVA-based model of visual attention. *Cognitive Computation, 2*(4), 326–343.

Wischnewski, M., Steil, J. J., Kehrer, L., & Schneider, W. X. (2009). Integrating inhomogeneous processing and proto-object formation in a computational model of visual attention. In *Human centered robot systems* (pp. 93–102).

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*, 1–7.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–71). London: University College London Press.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). Sun database: Large-scale scene recognition from abbey to zoo. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3485–3492). IEEE.

Yu, C., Samaras, D., & Zelinsky, G. J. (2014). Modeling visual clutter perception using proto-object segmentation. *Journal of Vision, 14*(7), 1–16. 4.