



Published in final edited form as:

*J Vis.* 2011 ; 11(6): . doi:10.1167/11.6.19.

## Object crowding

**Julian M. Wallace** and

Department of Psychology, University of Southern California, USA

**Bosco S. Tjan**

Neuroscience Graduate Program, University of Southern California, USA

### Abstract

Crowding occurs when stimuli in the peripheral fields become harder to identify when flanked by other items. This phenomenon has been demonstrated extensively with simple patterns (e.g., Gabors and letters). Here, we characterize crowding for everyday objects. We presented three-item arrays of objects and letters, arranged radially and tangentially in the lower visual field. Observers identified the central target, and we measured contrast energy thresholds as a function of target-to-flanker spacing. Object crowding was similar to letter crowding in spatial extent but was much weaker. The average elevation in threshold contrast energy was in the order of 1 log unit for objects as compared to 2 log units for letters and silhouette objects. Furthermore, we examined whether the exterior and interior features of an object are differentially affected by crowding. We used a circular aperture to present or exclude the object interior. Critical spacings for these aperture and “donut” objects were similar to those of intact objects. Taken together, these findings suggest that crowding between letters and objects are essentially due to the same mechanism, which affects equally the interior and exterior features of an object. However, for objects defined with varying shades of gray, it is much easier to overcome crowding by increasing contrast.

### Keywords

spatial vision; object recognition; detection/discrimination

### Introduction

Crowding refers to the phenomenon where stimuli in the peripheral field become harder to identify when flanked by other items (Bouma, 1970; Korte, 1923). A simple example of this is provided in Figure 1. The effect is often attributed to anomalous integration of features, representing an information-processing bottleneck that if understood could provide insight into general mechanisms of form processing and object recognition (Levi, 2008; Pelli & Tillman, 2008). Crowding has been demonstrated extensively with simple visual stimuli such as Gabors and letters and was the focus of a recent special issue in *Journal of Vision* (2007, vol. 7 no. 2); however, there are surprisingly few studies of the effects of crowding on objects. The present paper serves to characterize crowding for everyday objects.

---

© ARVO

Corresponding author: Bosco S. Tjan. btjan@usc.edu. Address: SGM 501 3620 South McClintock Ave., Los Angeles, CA 90089-1061, USA..

Citation: Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision*, 11(6):19, 1–17, <http://www.journalofvision.org/content/11/6/19>, doi:10.1167/11.6.19.

Commercial relationships: none.

The term “crowding” was introduced by Stuart and Burian (1962) to describe the detrimental influence of flankers on target identification, a phenomenon originally reported by Korte (1923; see also Ehlers, 1936). Key properties of the effect that can be considered the hallmarks of crowding have been discovered (Levi, 2008). The distance between target and flanker stimuli required to induce crowding depends upon the eccentricity in the visual field, and this “critical spacing” is approximately a constant fraction of eccentricity (Bouma, 1970, 1973). Bouma (1970, 1973) reported a critical spacing around half the eccentricity, a relation that is commonly referred to as Bouma’s Law, although others find lower scaling values (e.g., Strasburger, Harvey, & Rentschler, 1991). Bouma’s Law suggests that the underlying mechanism of crowding is tied to the cortical mapping of space, for example, the eccentricity scaling could reflect a fixed distance on cortex (Pelli, 2008; Pelli & Tillman, 2008). Further, the spatial extent of crowding has been found to be anisotropic; “interaction regions” or crowding zones are elliptical, with radial elongation, i.e., with the longer axis in the foveal direction (Chambers & Wolford, 1983; Toet & Levi, 1992). Crowding has a detrimental effect upon identification, over and above any effect on simple detection of targets in the periphery (Andriessen & Bouma, 1976), and is specific to peripheral vision—in the fovea, the critical distance between flankers and targets is proportional to target size and can be attributed to simple contrast masking (Levi, Hariharan, & Klein, 2002). The location of flankers has some influence on the strength of the effect. Specifically, a flanker with a greater eccentricity than the target leads to more crowding than a flanker at the same distance from the target but is situated between the target and the fovea (e.g., Banks, Larson, & Prinzmetal, 1979; Chambers & Wolford, 1983; Petrov, Popple, & McKee, 2007). Lastly, crowding does not depend upon the size of the gap between the edges of the target and a flanker. It depends upon the distance between the center of the target and the center of a flanker (Strasburger et al., 1991) or the centroid of multiple flankers (Levi & Carney, 2009).

Crowding is known to occur beyond retinal processing (Flom, Heath, & Takashi, 1963), but the precise mechanism remains to be understood. One account of crowding is lateral masking (Townsend, Taylor, & Brown, 1971; Wolford & Chambers, 1983). Although the presence of flankers impairs the identification of a central target, the effect is distinct from ordinary masking (Chung, Levi, & Legge, 2001; Levi, Hariharan et al., 2002; Levi, Klein, & Hariharan, 2002; Pelli, Palomares, & Majaj, 2004). Another possibility is that the visual system uses only low spatial frequencies to analyze a target when it is crowded; in fact higher than optimal frequencies are used in the periphery, but the shift in spatial frequency tuning is small and insufficient to account for crowding (Chung & Tjan, 2007). The alternative accounts understand crowding in terms of inappropriate integration of features (Levi, Hariharan et al., 2002; Levi, Klein et al., 2002; Nandy & Tjan, 2007; Pelli et al., 2004); examples include defective contour interaction (Flom et al., 1963) and averaging of orientations (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). Some authors argue that proper feature integration requires spatial attention, in which inappropriate integration of features would be a result of improper binding due to a limited spatial resolution of attention in the periphery (He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001; Strasburger et al., 1991; Tripathy & Cavanagh, 2002).

In their review, Pelli and Tillman (2008) provide a convincing demonstration of crowding between individual objects, but previous studies on object crowding have been quite limited. Crowding has been shown to occur between the individual features of a face and can be relieved (at least with line drawings) by moving the individual features further apart (Martelli, Majaj, & Pelli, 2005). Louie, Bressler, and Whitney (2007) found that crowding could occur between faces, specifically between an upright face target and a “crowd” of other upright faces, since replicated by Farzin, Rivera, and Whitney (2009) with Mooney faces. However, face processing may be different from object processing (Kanwisher, McDermott, & Chun, 1997). Indeed, Louie et al. also demonstrated crowding between

houses, but unlike faces this non-face object effect was independent of orientation. Their result was taken at a fixed eccentricity and a fixed target-to-flanker separation, with performance limited by the addition of noise. Such a quantification is incomplete as it has been argued that crowding should be described with two values: one that measures performance (accuracy or threshold) and another, which is probably more fundamental for understanding the mechanism of crowding, that measures the spatial extent of crowding (Chung & Bedell, 1995; Pelli & Tillman, 2008, online supplement). A manipulation can affect performance without changing the spatial extent of crowding. In the present study, we aim to provide a full characterization of the properties of object crowding and compare the results to more frequently studied letters.

As noted by Levi (2008), two approaches have commonly been used to assess crowding. The first method is to measure accuracy as a function of the spacing between the target and flanker. The second is to measure a contrast threshold for identifying the target, i.e., the contrast required for a specific level of performance (Strasburger et al., 1991), and to repeat this at a range of target-to-flanker spacings. One advantage of the latter method is that it provides a measure of the severity of the crowding effect in terms of threshold elevation.

We presented three-item arrays of objects and of letters, arranged radially (along the axis connecting the target and the fixation) and tangentially (orthogonal to the target–fixation axis) in the lower visual field. Subjects identified the central target, and we measured contrast energy thresholds as a function of target-to-flanker spacing (center to center). Similar to Pelli et al. (2004), we fitted a “clipped line” function (Figure 2) to estimate critical spacing and threshold elevation.

To anticipate our results, we found that object crowding is similar to letter crowding in spatial extent but is much weaker when assessed in terms of threshold elevation (less than a factor of 10 in contrast energy for objects vs. over a factor of 100 for letters). Silhouettes of objects, with uniform interior, were found to have comparable threshold elevation to letters. We examined whether the exterior and interior features of an object, operationally defined, are differentially affected by crowding. We used a circular aperture to present either just the interior portion of an object or everything else but the interior portion (a “donut” object). These manipulations had no consistent effect across subjects, except that threshold elevations were mildly larger for intact objects than donuts. To sum up, crowding between objects does not significantly differ from that between letters in terms of spatial extent and the anisotropy along the radial and tangential directions. Nevertheless, crowding-induced threshold elevations for objects (intact, aperture, donut) are much lower than that for letters and object silhouettes. Taken together, these findings suggest that crowding between letters and objects are essentially due to the same mechanism. However, for objects, it is easier to compensate for the loss in performance by increasing contrast. We found that the rich inner features of objects can support, rather than hinder, recognition in crowded conditions.

## Methods

### Stimuli

The letter stimuli were 26 lowercase letters in Arial font (provided on Macintosh OS × 10.5.4). Objects were selected from a commercially available image set of photographs of real objects at [www.photos.com](http://www.photos.com) (now at [www.thinkstockphotos.com](http://www.thinkstockphotos.com)).

One hundred and forty candidate objects were selected from this image base. To control for image size, object images with widths or heights greater than two standard deviations (SDs) of the set were excluded. All the remaining objects were scaled to the average height, and objects with aspect ratios greater than one SD were further excluded. Finally, we excluded

remaining outliers manually based on a measure of complexity (ratio of squared perimeter to area)—we plotted the histogram of complexity across all candidate objects and selected a criterion that removed obvious outliers. This gave us 56 test objects, equal in height and variable in width (SD equal to 0.22 times the mean width) and with a mean complexity of 24.07 (ranging from 13.94 to 44.67).

All the color (RGB) objects were converted to grayscale. The grayscale objects (but not letters) were equalized for root-mean-square (RMS) contrast with respect to the background luminance of the display (25.7 cd/m<sup>2</sup>). The normalized RMS contrast for each object was 0.21. Contrast equalization did not affect the mean luminance of an object. The mean luminance of the objects was 24.4 cd/m<sup>2</sup>, with a standard deviation of 2.0 cd/m<sup>2</sup>. We manipulated the contrast of these normalized objects in the experiments. For objects, we define a “nominal” contrast of 100% when their RMS contrast is 0.21. For letters, nominal contrast is defined as the Weber contrast of the brightest pixel on the stroke of a letter.

Stimuli were displayed on a calibrated and gamma-linearized Dell P1230 19” CRT monitor (resolution: 1024 by 768 at 75 Hz) at a viewing distance of 70 cm and controlled with a MacBook running Mac OS × 10.5.4. Each screen pixel subtended 0.0309° (32.4 pixels per degree). Eleven bits of linearly spaced contrast levels were available by use of a passive video attenuator (Pelli & Zhang, 1991) and custom-built contrast calibration and control software implemented in MATLAB, using only the green channel of the monitor.

## Procedure

In all main experiments, each condition was presented for 3 blocks of 60 trials per block. To distribute conditions evenly throughout the experiment, all possible conditions in an experiment were presented within a superblock that included one block of every condition in random order. Subjects had to fixate on a cross near the top of the screen, such that the target object was presented at an eccentricity of 10 degrees. Contrast was adjusted using the QUEST procedure (Watson & Pelli, 1983) as implemented in the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) to estimate threshold contrast for reaching an accuracy level of 50%, corrected for guessing. Further details are provided in Appendix A.

The temporal sequence of events in a trial was given as follows (see also Figure 3): (a) a fixation screen for 750 ms initiated with an auditory beep, (b) stimulus presentation for 250 ms, (c) subject response period (variable) with positive feedback beep for correct trials or a negative feedback beep for incorrect trials, and (d) a 500-ms delay before onset of the next trial. On each trial, we collected the identity and contrast of the target letter/object and the response of the subject for subsequent data analysis. In Experiments 2 and 3, an additional 250 ms was added on to the fixation and inter-trial delay periods, to further encourage accurate fixation.

Responses were made with a mouse click to select the name of the object or letter from an array. For objects, this was an array of 20 names that constitute the entire set of objects used for that particular block. For letters, this was an array of the 26 letters.

Stimuli were presented at a range of spacings: for the tangential condition, 15 logarithmically spaced values from 0.5 to 10 degrees were used with flankers and target arranged horizontally. In the radial condition, 14 logarithmically spaced values from 0.5 to 8 degrees were used (to prevent overlap of the near flanker with fixation) with flankers and target arranged vertically. In both radial and tangential conditions, an “infinite” (i.e., no flanker) spacing was added, giving 16 spacings in total for tangential conditions and 15 spacings in total for radial conditions. At small flanker spacings, the objects could overlap. In this case, the target object was made to occlude the flankers.

## Acuity

Peripheral object acuity was measured for all subjects prior to the main experiments. Objects were presented at 10 degrees in the lower peripheral field without any flankers, and subjects were asked to identify the object. The object size was varied using QUEST to achieve an identification accuracy of 50%, corrected for guessing (as in the main experiments). Average acuity was taken over 5 blocks of 60 trials. The objects and letters were presented at 1.5 times the subject's object acuity in height.

## Training

All subjects were trained to identify the objects used in the main experiment. In the first training session, a slideshow of the objects was presented foveally. This was followed by a short session of identification in which learned objects were randomly presented foveally, and subjects identified the object by selecting from a response array, over at least 3 blocks of 30 trials, until they reached a criterion level of performance (90% or more on average across blocks). This training was then repeated at 10 degrees in the lower peripheral field. For Experiments 2 and 3, this training was performed for all object types.

Contrast thresholds were measured during training for peripherally presented targets without flankers (objects and letters in Experiment 1 and the different object types in Experiments 2 and 3). Contrast was adjusted using QUEST to estimate threshold contrast over 60 trials for reaching an accuracy level of 50%, corrected for guessing. This continued for a number of blocks until thresholds appeared stable.

## Analysis

The data of threshold contrast energy ( $E$ ) versus center-to-center spacing ( $s$ ) were fit with a clipped line function:

$$\log(E(s)) = \begin{cases} \log(E_{\text{ceiling}}) & \text{if } s \leq s_{\text{sat}} \\ \log(E_{\text{floor}}) & \text{if } s \geq s_{\text{critical}} \\ \text{else:} & \\ \frac{\log(E_{\text{ceiling}}) - \log(E_{\text{floor}})}{\log(s_{\text{critical}}) - \log(s_{\text{sat}})} (\log(s_{\text{sat}}) - \log(s)) + \log(E_{\text{ceiling}}) & \end{cases} \quad (1)$$

This function has four parameters: ceiling ( $E_{\text{ceiling}}$ ), floor ( $E_{\text{floor}}$ ), saturation spacing ( $s_{\text{sat}}$ ), and critical spacing ( $s_{\text{critical}}$ ). It provides an adequate description of the data: the part of the function for  $s > s_{\text{sat}}$  is commonly used to characterize crowding for relatively large target-flanker spacing (Chung et al., 2001; Pelli et al., 2004); for  $s < s_{\text{sat}}$ , we found that over the range we tested,  $s$  had no significant effect on threshold (repeated measures ANOVA,  $F(4, 26) = 0.305$ ,  $p = 0.87$ ). We estimated the parameters by fitting Equation 1 to data using a multi-start procedure, which minimizes the squared residual in  $\log(E)$ .

For each subject and condition, we obtained the mean log contrast energy threshold and corresponding standard error per center-to-center spacing using QUEST. This provided the 14 (radial conditions) or 15 (tangential conditions) data points, not including the “no-flanker” condition, which defined the four parameters. We have sufficient power to make reliable within-subject comparisons by computing the 95% confidence interval ( $CI_{0.95}$ ) for each of the four estimated parameters using bootstrapping. Appendix B provides the details of these procedures. For within-subject comparisons, we declare two estimated quantities to be significantly different at  $\alpha = 0.05$  if neither of the quantities is within the 95% confidence intervals of the other. Given that we had only three subjects per experiment, we shall refrain from making any formal statistical claims at the population level. We rely on cross-subject consistency to draw general conclusions.



## Experiment 1

Objects and letters were presented in radial and tangential arrangements in the lower peripheral field (Figure 4). The center object/letter was always presented at  $10^\circ$  eccentricity. Object flankers were presented at a nominal contrast of 50% (contrast energy of  $0.02 \text{ deg}^2$ ); letter flankers were presented at a nominal contrast of 30% (contrast energy of  $0.03 \text{ deg}^2$ ). These flanker contrasts were significantly above the threshold contrast for single-target identification, thus ensuring a robust crowding effect (Pelli et al., 2004). Radial and tangential conditions were presented within separate superblocks, and subjects completed one superblock of each before continuing. The order was counterbalanced across subjects (JW: T, R, R, T, T, R; AJ: R, T, R, T, R, T; MS: T, R, T, R, T, R). On each block, 20 objects were used for targets and a different set of 20 objects was used for flankers, randomly selected from the set of 56 objects. On each new session, a slideshow of the 56 objects was shown to the subjects.

## Results

Contrast energy thresholds as a function of spacing are shown in Figure 5, for the three subjects. The upper plots are for the letter conditions, and the lower plots are for the object conditions. Equation 1 captures the data well (mean  $R^2 = 0.94$ ). The parameter estimates are presented in Figure 6 and summarized in Table C1 of Appendix C.

For all subjects, critical spacing ( $s_{\text{critical}}$ ) was significantly larger in the radial than tangential arrangement of the flankers, for both letters and objects (Figure 6 and Table C2). For objects, the average critical spacing was  $3.93^\circ$  for radial flankers and  $2.78^\circ$  for tangential flankers. For letters, the critical spacing was  $3.59^\circ$  for radial flankers and  $1.87^\circ$  for tangential flankers. The differences in critical spacings between objects and letters were generally small and inconsistent across subjects.

There were differences in threshold elevations ( $E_{\text{ceiling}}/E_{\text{floor}}$ ) between the radial and tangential conditions for individual subjects; however, these were not consistently in the same direction in the case of objects, while for letters, the radial threshold elevation was consistently larger by a small amount, between 0.10 log unit and 0.13 log unit (Figure 6 and Table C2). Comparing objects with letters, there was a very large difference in threshold elevation between objects and letters in both radial and tangential conditions. This difference was significant for all subjects, with the elevation in threshold contrast energy being 2.05 log units on average for letters but merely 0.80 log unit for objects.

These data demonstrate a clear radial-tangential anisotropy for objects, considered to be a hallmark of crowding. They also demonstrate that there is, in general, no difference in the spatial extent of crowding between letters and objects. However, the results show a larger than ten-fold difference in threshold elevation, even though there is only a three-fold difference between the contrast energy thresholds for identifying a stand-alone target ( $0.0023 \text{ deg}^2$  for letters,  $0.0069 \text{ deg}^2$  for objects).

Interpreting threshold elevation across two different stimulus types has been fraught with difficulties (Tjan, Braje, Legge, & Kersten, 1995). There is also some discrepancy in the literature about effects of flanker contrast on crowding. Some authors report crowding effects that depend on the ratio of target-to-flanker contrast (Chung et al., 2001), while others (Pelli et al., 2004) have shown that flanker contrast, once above what is required to identify a flanker when presented alone, has little effect on threshold elevation in crowding. Flanker contrast may have played a partial role in the difference we observed between threshold elevations. The average contrast energy of the flankers was  $0.03 \text{ deg}^2$  for letters ( $12.89 \times$  the average threshold floor,  $E_{\text{floor}}$ , or 1.11 log units) and  $0.02 \text{ deg}^2$  for objects ( $2.88$

$\times E_{\text{floor}}$  or 0.46 log units). The corresponding threshold elevations were 2.05 log units for letters and 0.80 log units for objects. A 1.5 difference in flanker contrast energy between the two conditions  $\times$  is unlikely to explain an 18 difference in threshold elevations. The observed threshold  $\times$  elevations for both letters and objects are also disproportionately larger than the ratio of the flanker contrast energy to the respective threshold floor ( $E_{\text{floor}}$ ). While flanker contrast may have an influence on threshold elevation, the large difference in threshold elevations between letters and objects cannot be explained by the difference in flanker contrast, which was much smaller both absolutely and also relative to the respective threshold floors. Objects and letters appear inherently different in terms of the threshold elevation caused by crowding. We will return to this point in Experiment 2.

## Experiment 2

In Experiment 1, we found that object crowding has a radial-tangential anisotropy like letter crowding and a similar spatial extent. We also found that threshold elevation was much larger for letters than for objects. One basic difference between letters and objects seems to be that objects can be identified both by the “outer” bounding contour and the “inner” features, while the contour of letters is what defines them; there are no inner features for a letter as there is no contrast variation within the letter. To examine if this difference has any impact on crowding, we presented three different object types: full objects, apertures, and silhouettes (see Figure 7). Full objects have the outer contour and inner features available for identification. Apertures are objects with outer regions cut out by a disk, as if the objects are being viewed through an aperture. This manipulation preserves many inner features but removes most of the bounding contour. Thus, the informative image features are now restricted to a more central region, relative to the flankers. Silhouettes retain the bounding contour, but the inside is filled uniformly—there is no contrast variation inside the object and so this is an object equivalent to letters. We were interested to find out if these different types of rendering led to different critical spacing and threshold elevation.

A subset of 20 objects was selected to be targets from the original 56, and a different subset of 20 objects was selected to be flankers to minimize the possibility of a subject mistaking a flanker for the target. To define object apertures, objects were first aligned by fitting ellipses to the objects and then centered on the ellipse centers. The aperture diameter was set to the smallest width of the set of objects (corresponding to 67% of the average width of the objects)—applying this size of aperture resulted in less than 2% of the bounding contours of the original objects being preserved while retaining 46% of the object area (on average). The objects that were behind the aperture were of the same size as the objects in the object and silhouette conditions, equal to 1.5 times the subject’s intact-object acuity. Objects were presented in radial (vertical) arrangement in the lower field; the target object was presented at 10° eccentricity. Object and aperture flankers were presented at 100% nominal contrast, double that of Experiment 1, and silhouette flankers were presented at 30% nominal contrast (to be comparable with the letter condition of Experiment 1). Before each block, a slideshow of the 20 target objects was shown to the subjects.

## Results

Contrast energy thresholds as a function of spacing are shown in Figure 8, for three subjects. The clipped line function captures the data well (mean =  $R^2$  0.93, Figure 8). There is some overlap between object and aperture functions, while silhouette ceiling is clearly higher. The critical spacings for all conditions are quite similar. The parameters are summarized in Figure 9 and Table C3.

Averaged across the three subjects, the critical spacings are 3.21° for objects, 3.30° for apertures, and 2.41° for silhouettes. Individually, there were significant differences for 2 of 3

subjects for both objects vs. silhouettes and apertures vs. silhouettes, with the silhouettes spacing being smaller and similar to the pattern of results for objects vs. letters in Experiment 1. There was no consistent effect for objects vs. apertures. Further, although the critical spacing for objects ( $3.21^\circ$  on average) appears smaller than that found in Experiment 1 ( $3.93^\circ$  on average), this is likely due to individual differences. For a full summary of within-subject comparisons, see Table C4.

For all subjects, threshold elevation is significantly larger for silhouettes than both objects and apertures (see Tables C3 and C4). The average elevation in threshold contrast energy is 1.26 log units for objects, 1.40 log units for apertures, and 2.28 log units for silhouettes. Threshold elevation for the silhouettes is very similar to that found for letters in Experiment 1. There was no significant difference between objects and apertures. The average flanker contrast energies were  $0.11 \text{ deg}^2$  for intact objects,  $0.05 \text{ deg}^2$  for apertures, and  $0.23 \text{ deg}^2$  for silhouettes. Threshold elevation for the intact objects was higher in Experiment 2 than that for the corresponding radial condition in Experiment 1 (1.26 log units vs. 0.75 log unit). This difference can be attributed to the difference in the flanker contrast energies between these experiments ( $0.11 \text{ deg}^2$  vs.  $0.02 \text{ deg}^2$ ), consistent with Chung et al. (2001).

The average contrast energy for intact-object flankers in Experiment 2 ( $0.11 \text{ deg}^2$ ) exceeded the average contrast energy of letter flankers in Experiment 1 ( $0.03 \text{ deg}^2$ ), both in the absolute and relative to the respective threshold floors. Yet the observed threshold elevation for objects in Experiment 2 remains much lower than the threshold elevation for letters in Experiment 1 (1.26 vs. 2.05 log units). This finding strengthens our conclusion that there is a large difference in crowding-induced threshold elevation between letters and objects that cannot be explained by flanker contrast. Furthermore, between Experiments 1 and 2, we observed that for grayscale objects but not for letters or silhouettes, target contrast does not need to exceed flanker contrast to compensate for crowding at small target-flanker separations.

These results demonstrate that coarse manipulation of object properties, by removing the outer bounding contour or retaining the contour but blanking out the inner features, has little effect on the spatial extent of crowding. The location of the informative image features, whether they be contrast variations present throughout the object (objects), restricted to a central region of the object (apertures), or absent entirely except at the bounding contour (silhouettes), has little influence on the crowding zone. The striking difference is in threshold elevation, between objects and apertures, on the one hand, and silhouettes on the other. In this way, silhouettes behave very much like letters. Threshold elevation, the strength of crowding, appears to depend upon the contrast properties of the stimulus: stimuli with a more uniform distribution of contrast require greater contrast energy to be released from crowding, since the contrasts for both the informative and non-informative pixels are about the same and have to be increased by the same amount. While a relatively small increase in contrast energy is sufficient to aid identification for objects with local variations in contrast, a much larger increase in contrast is required to overcome effects of crowding for letters and silhouette objects.

### Experiment 3

In Experiment 2, the inner features of objects were removed by filling in the contour of the object uniformly. This led to a threshold elevation effect similar to letters but had little influence on the spatial extent of crowding. The lack of an effect on the spatial extent suggests that the interior and exterior features of an object (or its silhouette) are similarly susceptible to crowding. In the present experiment, we further tested this possibility with stimuli of varying local contrast. This was achieved by cutting out a disk from the grayscale



objects, the same size as the aperture used in Experiment 2. These “donut” stimuli are the complement of the aperture stimuli used in Experiment 2 (see Figure 10). The information available for identifying donuts is the outer contour and a small amount of remaining contrast features interior to the outer contour. Thus, the informative image features are constrained to be closer to the flankers, as opposed to the intact objects for which informative image features are available across the entire object. Other aspects of the experiment were identical to those of Experiment 2.

## Results

Contrast energy thresholds as a function of spacing are shown in Figure 11 and are well fit with a clipped line function (mean  $R^2 = 0.92$ ). The estimated values for critical spacing and threshold elevation are summarized in Figure 12 and listed in Table C5. On average, the critical spacing for intact objects was  $2.49^\circ$ , and that for donuts was  $2.72^\circ$ . The average threshold elevation for objects was 1.26 log units, and that for donuts was 0.93 log units. Only one subject showed a significant effect in critical spacing, but for all subjects, there was a mild but significant effect in threshold elevation, with the threshold elevation for the intact objects being higher (see Tables C5 and C6).

The results of this experiment demonstrate that removing a large interior portion of the object (approximately half the object area) has remarkably little impact on the spatial extent of crowding (there was an increase of  $0.23^\circ$  on average or 2.3% of the target eccentricity). The removal of the interior features apparently made the task harder, leading to an increase of  $E_{\text{floor}}$ , with little effect on  $E_{\text{ceiling}}$ . The result is a modest 0.33 log unit reduction in threshold elevation (from 1.26 log units for objects to 0.93 for donuts). In other words, while removing the interior features made the task harder, it did not make the target more susceptible to crowding. The results of Experiments 2 and 3, taken together, suggest that the interior and exterior features of an object are similarly susceptible to crowding.

## Discussion

To summarize, we measured threshold contrast energies for target identification as a function of target-to-flanker spacing at a fixed peripheral eccentricity ( $10^\circ$ ) and estimated critical spacing and threshold elevation for letters and objects over three experiments. In Experiment 1, we found that crowding between objects has a radial–tangential anisotropy like letters, a property considered a hallmark of crowding (Levi, 2008). Crowding between objects occurs across a similar spatial extent as letters, but letters have a much larger threshold elevation compared to objects. In Experiment 2, we found that the critical spacing was similar between intact objects, apertured objects, and silhouettes. However, while threshold elevation was similar for intact objects and apertures, silhouettes had a much larger threshold elevation, similar in magnitude to letters. In Experiment 3, we found that removing a large region of the object’s interior had a negligible effect on critical spacing and threshold elevation compared to the intact objects. These results suggest that the crowding effects we observed for all stimuli, including objects and letters, are due to essentially the same mechanism. Previous findings of crowding that have used letters or letter-like stimuli should, therefore, generalize to objects.

There is an advantage of grayscale objects over letters and silhouettes in a cluttered environment, in that the informative features of objects, defined by local variations in contrast, appear to mitigate the detrimental effects of crowding. Compared to letters or silhouettes, grayscale objects (intact, aperture, donut) require a much smaller increase of contrast in a crowded condition to restore accuracy to the uncrowded level. Furthermore, crowding does not depend on the location of the contrast-defined features in an object: features interior to an object are affected by crowding just as much as the features near and

along the exterior contour of an object. The effects of removing the interior region of the object in the donut condition on both the spatial extent of crowding and crowding-induced threshold elevation were remarkably small and inconsequential. We now consider these results in the context of the prior findings.

### **Crowding as a bottleneck for object recognition**

There has been a surge of interest in crowding in recent years. One aspect of this increased interest is the idea that crowding represents an essential bottleneck for recognizing objects in peripheral vision (Levi, 2008) and that studying crowding in the periphery may lead to insights about the feature integration process that underlies object recognition. However, as noted in the Introduction, there have been few studies that examined crowding with stimuli more complex than letters or Gabors. Those studies have examined crowding for faces, demonstrating crowding between features of line-drawn faces (Martelli et al., 2005) and orientation-dependent face crowding (Farzin et al., 2009; Louie et al., 2007) related to the face inversion effect. Aside from Pelli and Tillman's (2008) convincing demonstration of crowding between individual objects, the only paper to date that has measured crowding between objects is Louie et al. (2007), which showed crowding between images of houses; however, that study was done at a fixed target-to-flanker separation. The present study is the only study to our knowledge that has quantitatively measured crowding for objects (i.e., photographs of real objects) to estimate the spatial extent and amplitude of the effect. Our results demonstrate that crowding does, indeed, occur for objects over a spatial extent similar to that for letters.

If crowding is, indeed, due to inappropriate integration of features, what features are being inappropriately integrated and at what stage of cortical processing does this inappropriate integration occur? Identifying the features that underlie crowding may suggest the stage in cortical visual processing at which crowding first occurs, i.e., the source of the bottleneck and vice versa. The present results indicate that there is no difference in the spatial extent of crowding between objects and letters. This is striking considering how simplistic letters are, with very clear contours of different orientations, compared to the complexity and variety of objects, with potentially many different "features." However, this difference between letters and objects appears irrelevant for the mechanism that determines the spatial extent of crowding. The similarity in results between objects and letters suggests that the type of features that leads to crowding is present in both objects and letters. Since there are large differences in broadband features between objects and letters, narrowband features (each with a narrow range of spatial frequencies) may be underlying the effect. Recent studies have found that phase-scrambled letters cause equivalent crowding to regular letters, and similarly phase-scrambled objects cause equivalent crowding to regular objects (Shin, Wallace, & Tjan, 2010; Tjan & Dang, 2005), which is consistent with the view that the features that lead to crowding are low-level narrowband features before any phase-specific integration occurs. This view would place the cortical bottleneck for crowding at the early stages of visual processing, in V1 or V2, although it does not exclude the possibility that crowding also occurs at subsequent stages of visual processing. Other psychophysical studies have used adaptation to identify the stage at which crowding occurs. He et al. (1996) found no difference in adaptation to a grating whether it was presented alone or with flankers (although observers could not identify the orientation of the grating in that condition), suggesting that crowding occurs beyond V1, while Blake, Tadin, Sobel, Raissian, and Chong (2006) instead found that crowding does affect adaptation when low-contrast gratings are used, suggesting that crowding can occur as early as V1. Chakravarthi and Cavanagh (2009) found that while noise and meta-contrast masks applied to flankers relieved crowding, a "high-level" object substitution mask did not and suggested that the site of crowding falls between V1 and LOC. Liu, Jiang, Sun, and He (2009) presented distractors

(single flankers) to targets near the horizontal or vertical meridian and found that flankers on the same side of the target with respect to the vertical meridian caused stronger crowding, an effect that did not occur across the horizontal meridian. This suggests that crowding occurs in a cortical location in which the cortical representation is discontinuous across the vertical meridian but continuous across the horizontal meridian. This places the site of crowding either higher than V2 and V3 or at V1. Currently, there are few published neuroimaging studies in this direction, but recently Bi, Cai, Zhou, and Fang (2009) found that orientation-selective fMRI adaptation was not affected by crowding in V1 but was in V2 and V3. Taken together, these results strongly favor a low-level cause of crowding without excluding additional deficits in higher level visual processing.

### Crowding in natural scenes

We often encounter cluttered visual scenes and need to identify objects correctly to navigate and interact with the world. These complex visual scenes have no resemblance to the blank backgrounds often used in psychophysics. Indeed, identification of objects can be much worse in some cases when presented on complex backgrounds compared to blank backgrounds (Bravo & Farid, 2006). When viewing a natural scene, crowding could occur not only between objects in close proximity but also between objects and their background, interfering both with object segmentation and identification. The identification of an object on a complex background poses a non-trivial segmentation problem (Marr, 1982), which can be further exacerbated by crowding in peripheral vision. However, it may be that objects do not need to be segmented prior to recognition. Recognition can occur in a bottom-up fashion based on object parts prior to segmentation (Bravo & Farid, 2006; Ullman, Vidal-Naquet, & Sali, 2002). Our results suggest that most natural objects, with non-uniform local contrast and a rich set of features, offer some protection from crowding compared to uniform-contrast stimuli like letters and silhouettes.

Scanning and searching the visual environment is a task we perform frequently. Recent studies have made a link between visual search, visual clutter, and crowding (e.g., Bravo & Farid, 2006; Rosenholtz, Li, & Nakano, 2007; van den Berg, Cornellisen, & Roerdink, 2009). Clutter is a vague term that refers to a visual scene within which objects are not arranged sparsely. Indeed, most natural scenes are cluttered. Clutter behaves similarly to set size in visual search and degrades performance—visual search of a target object is longer for scenes with more clutter (Rosenholtz et al., 2007) and in particular for cluttered backgrounds consisting of compound (multi-part) objects (Bravo & Farid, 2006). Further, clutter can result not only in more identification errors but also a higher confidence about those erroneous responses (Baldassi, Megna, & Burr, 2006). Recently, a model of feature detection that uses eccentricity-dependent integration fields (van den Berg et al., 2009) had been shown to reproduce results of clutter on visual search, suggesting that crowding places a limit on visual search performance in cluttered environments. Further studies in this direction, such as examining visual search for objects with natural scene backgrounds, may prove informative at least as to the extent crowding is a limiting factor when scanning the visual environment.

### Conclusion

Crowding between objects does not significantly differ from that between letters in terms of spatial extent and the anisotropy along the radial and tangential directions. Spatial distribution of informative features within a grayscale object has little effect on crowding. However, crowding-induced threshold elevations for grayscale objects (intact, aperture, donut) are much lower than that for letters or silhouettes of objects. These findings suggest that crowding between letters and crowding between objects are essentially due to the same mechanism; however, for grayscale objects, it is easier to compensate for the loss in

performance by increasing contrast. The rich variations of local contrast of a grayscale object can support, rather than hinder, recognition in crowded conditions.

## Acknowledgments

We thank Rachel Millin, Benjamin Files, Anirvan Nandy, two anonymous reviewers, and Al Ahumada, editor, for their insightful comments. This work supported by NIH/NEI Grants R01-EY017707 and R01-EY016093.

## Appendix A

### QUEST

We used the following parameters for QUEST as implemented in Psychophysics Toolbox: prior log threshold estimate = 0.8; standard deviation of the prior = 3;  $\beta = 3.5$ ;  $\delta = 0.01$ ;  $\gamma_{\text{Objects}} = 1/20$ ;  $\gamma_{\text{Letters}} = 1/26$ . The criterion accuracy was 52.5% uncorrected for 20 objects and 51.9% for 26 letters, yielding an accuracy of 50% after correcting for guessing. The QUEST procedure was reinitialized for each block of 60 trials. The mean and standard error (*SE*) of the log threshold contrast were computed by concatenating blocks of identical condition and refitting the Weibull function that underlies the QUEST procedure by calling PsychToolbox functions QuestRecompute followed by QuestMean and QuestSd. We found that for our data, this procedure yielded similar values as those calculated using per-block log thresholds, but the procedure is more robust in simulation. This finding demonstrates that QuestSd provides an accurate estimate of the true SE.

We also confirmed that the QUEST parameters we used ( $\beta$  in particular) led to veridical estimation of the contrast thresholds at the criterion accuracy of 52.5% for objects and 51.9% letters (i.e., at 50% corrected for guessing). We did so by calculating the mean accuracy of the trials with test contrast within plus and minus one standard error of the estimated log contrast thresholds. For objects (intact, silhouettes, apertures, donuts), the accuracy of these near-threshold trials was  $55.6 \pm 0.3\%$  (from 28,660 trials, which is about 50.7% of the total 57 kilo trials in all

**Table C1**

Experiment 1: Estimated parameters with 95% confidence intervals for subjects AJ, JW, and MS. *S* is the least squares of the fits (Equation B2).

		$\log(E_{\text{ceiling}})$	$\log(E_{\text{floor}})$	$\log(E_{\text{ceiling}} / E_{\text{floor}})$	$\log(S_{\text{critical}})$	$\log(S_{\text{sat}})$	$\log(S_{\text{critical}} / S_{\text{sat}})$	<i>S</i>
<i>AJ</i>								
Object	Radial	-1.44 (-1.47, -1.42)	-2.11 (-2.15, -2.07)	0.66 (0.62, 0.71)	0.51 (0.48, 0.56)	0.37 (0.35, 0.40)	0.15 (0.08, 0.21)	199.38
	Tangential	-1.37 (-1.40, -1.33)	-2.25 (-2.29, -2.21)	0.88 (0.83, 0.93)	0.45 (0.40, 0.48)	0.23 (0.20, 0.26)	0.23 (0.15, 0.28)	81.74
Letter	Radial	-0.49 (-0.52, -0.47)	2.65 (2.69, 2.62)	2.16 (2.11, 2.21)	0.61 (0.59, 0.65)	0.30 (0.28, 0.31)	0.31 (0.28, 0.36)	98.90
	Tangential	-0.64 (-0.68, -0.60)	2.70 (2.74, 2.68)	2.06 (2.02, 2.12)	0.33 (0.31, 0.37)	0.02 (0.02, 0.04)	0.30 (0.27, 0.39)	42.62
<i>JW</i>								
Object	Radial	1.36 (1.39, -1.33)	2.29 (2.34, 2.26)	0.93 (0.89, 0.98)	0.60 (0.57, 0.66)	0.16 (0.16, 0.19)	0.44 (0.39, 0.50)	104.92
	Tangential	-1.64 (-1.67, -1.61)	-2.29 (-2.31, -2.27)	0.65 (0.62, 0.69)	0.35 (0.27, 0.36)	0.10 (0.06, 0.14)	0.25 (0.14, 0.29)	154.59
Letter	Radial	0.60 (0.63, 0.58)	-2.65 (-2.70, -2.61)	2.05 (2.00, 2.10)	0.51 (0.46, 0.58)	0.15 (0.11, 0.20)	0.36 (0.27, 0.46)	55.43
	Tangential	-0.67 (-0.80, -0.61)	2.58 (2.63, 2.56)	1.91 (1.78, 2.00)	0.29 (0.24, 0.40)	-0.16 (-0.22, -0.06)	0.45 (0.30, 0.62)	140.60
<i>MS</i>								
Object	Radial	-1.18 (-1.22, -1.16)	-2.08 (-2.12, -2.04)	0.90 (0.84, 0.94)	0.67 (0.53, 0.72)	0.19 (0.16, 0.32)	0.48 (0.22, 0.54)	221.44
	Tangential	-1.15 (-1.18, -1.12)	-1.93 (-1.96, -1.91)	0.78 (0.74, 0.83)	0.53 (0.53, 0.53)	0.14 (0.11, 0.16)	0.40 (0.37, 0.43)	108.85
Letter	Radial	0.46 (0.49, 0.43)	2.59 (2.64, 2.52)	2.13 (2.06, 2.19)	0.55 (0.46, 0.56)	0.12 (0.10, 0.14)	0.43 (0.32, 0.45)	35.85
	Tangential	0.61 (0.65, 0.57)	2.62 (2.65, 2.60)	2.02 (1.97, 2.07)	0.20 (0.19, 0.20)	0.06 (0.05, 0.06)	0.14 (0.13, 0.15)	199.08

object conditions of all experiments). For letters, the accuracy of these near-threshold trials was  $53.8 \pm 0.6\%$  (from 7758 trials, which is about 48% of the total 16 kilo trials in the letter conditions of Experiment 1).

## Appendix B

### Data fitting

The contrast energy ( $E$ ) of an object (or a letter) is the integral of the squared Weber contrast over the entire object (Tjan et al., 1995):

$$E = \sum_j \left( \frac{l_j - l_0}{l_0} \right)^2 \Delta h \Delta w, \quad (\text{B1})$$

where  $j$  ranges over all pixels that belong to the object,  $\Delta h$  and  $\Delta w$  are the height and width of a pixel, respectively, in degrees,  $l_j$  is the luminance of the pixel  $j$ , and  $l_0$  is the background luminance.

To estimate the parameters of the clipped line function, Equation 1 of the main text was fitted to data by minimizing the squared residual in  $\log(E)$ :

$$S(\theta) = \sum_{i=1}^N \left( \frac{\log(E_i) - \log(E(s_i; \theta))}{\sigma_i} \right)^2, \quad (\text{B2})$$

where  $\theta = [s_{\text{critical}}, s_{\text{sat}}, E_{\text{floor}}, E_{\text{ceiling}}]$  is the vector of parameters to be estimated,  $E_i$  is the measured threshold contrast energy at spacing  $s_i$ , and  $\sigma_i$  is the standard error of  $\log(E_i)$ . Because errors have an approximately log-normal distribution, the actual fitting algorithm operates on  $\log(\theta)$  as opposed to  $\theta$ . Each data set was fit using the

**Table C2**

Experiment 1: Within-subject differences in critical spacings and threshold elevations between conditions at  $\alpha = 0.05$  (N.S. = not significant).

Critical spacing				
Subject	Objects: Radial (R) vs. tangential (T)	Letters: R vs. T	Radial: Objects (O) vs. letters (L)	Tangential: O vs. L
AJ	>	>	<	>
JW	>	>	>	N.S.
MS	>	>	N.S.	>
Log threshold elevation				
Subject	Objects: Radial (R) vs. tangential (T)	Letters: R vs. T	Radial: Objects (O) vs. letters (L)	Tangential: O vs. L
AJ	<	>	<	<
JW	>	>	<	<
MS	>	>	<	<



**Table C3**

Experiment 2: Estimated parameters with 95% confidence intervals for subjects AR, JD, and TC.  $S$  is the least squares of the fits (Equation B2).

	$\log(E_{\text{ceiling}})$	$\log(E_{\text{floor}})$	$\log(E_{\text{ceiling}} / E_{\text{floor}})$	$\log(S_{\text{critical}})$	$\log(S_{\text{sat}})$	$\log(S_{\text{critical}} / S_{\text{sat}})$	$S$
<i>AR</i>							
Objects	-1.49 (-1.52, -1.41)	-2.47 ( 2.50, 2.44)	0.98 (0.93, 1.09)	0.49 (0.46, 0.58)	0.36 (0.18, 0.40)	0.13 (0.07, 0.39)	102.24
Apertures	1.00 ( 1.06, 0.96)	-2.38 ( 2.43, 2.33)	1.38 (1.29, 1.44)	0.48 (0.39, 0.56)	0.04 (0.00, 0.13)	0.44 (0.26, 0.54)	82.33
Silhouettes	-0.09 (-0.13, -0.03)	-2.29 ( 2.33, 2.26)	2.20 (2.15, 2.27)	0.48 (0.47, 0.51)	0.20 (0.14, 0.21)	0.28 (0.26, 0.37)	61.89
<i>JD</i>							
Objects	-1.21 (-1.25, -1.17)	-2.46 ( 2.50, 2.42)	1.25 (1.20, 1.31)	0.46 (0.44, 0.47)	0.25 (0.23, 0.26)	0.21 (0.19, 0.24)	57.69
Apertures	0.88 ( 0.91, 0.85)	-2.22 (-2.25, -2.19)	1.34 (1.30, 1.39)	0.44 (0.44, 0.44)	0.19 (0.14, 0.21)	0.25 (0.23, 0.30)	270.85
Silhouettes	0.18 (0.10, 0.21)	-2.19 (-2.22, -2.16)	2.37 (2.28, 2.41)	0.39 (0.36, 0.40)	0.13 (0.12, 0.21)	0.26 (0.15, 0.28)	448.48
<i>TC</i>							
Objects	1.33 ( 1.36, 1.26)	-2.88 (-2.91, -2.80)	1.55 (1.50, 1.60)	0.58 (0.47, 0.60)	0.03 ( 0.04, 0.03)	0.61 (0.50, 0.64)	149.46
Apertures	1.36 ( 1.41, 1.31)	2.83 ( 2.89, 2.76)	1.47 (1.38, 1.55)	0.64 (0.56, 0.66)	0.28 (0.24, 0.42)	0.36 (0.14, 0.41)	12.74
Silhouettes	0.45 ( 0.49, 0.41)	2.72 ( 2.75, 2.69)	2.27 (2.22, 2.32)	0.28 (0.27, 0.29)	0.15 (0.14, 0.15)	0.13 (0.12, 0.14)	73.73

fmincon function in Matlab with a multi-start approach. One thousand independent attempts of fitting a data set were made, each with an initial guess of the log of the parameters uniformly sampled from the domain:  $\log(0.5^\circ) < \log(s_{\text{sat}}) < \log(s_{\text{critical}}) < \log(10^\circ)$ , and  $\min(\log(E_i)) < \max(\sigma_i) \sigma \log(E_{\text{floor}}) < \log(E_{\text{ceiling}}) < \max(\log(E_i)) + \max(\sigma_i)$ . The parameters obtained from the attempt with the least residual were taken as the estimated parameters. This multi-start method led to identical results when we repeated the procedure with different initial guesses, indicating that we were able to reach the global minimum. In fact, we found empirically that we could reach the global minimum with as little as 100 initial guesses.

A common practice to estimate the error bounds of the fitting parameters is to use the Hessian of the chi-square (Press, 1992, chapter 15). Specifically, we could express the marginal error bounds of the estimated parameters in terms of the Hessian of the chi-square evaluated at the estimated parameters ( $\hat{\theta}$ ):

$$\sigma(i) = \sqrt{2[H(S)|_{\hat{\theta}}]^{-1}(i, i)}, \quad (\text{B3})$$

where  $\sigma(i)$  is the (marginal) standard error of the parameter  $\theta(i)$ ,  $H$  is the Hessian operator,  $[\cdot]^{-1}$  denotes matrix inverse, and  $(i, i)$  indexes the  $i$ th diagonal element.

This Hessian approach assumes that the residual is locally smooth and symmetric about the estimated parameters. While the smoothness assumption is met as long as the estimated  $s_{\text{sat}}$  or  $s_{\text{critical}}$  does not coincide with a tested spacing, the assumption of symmetry cannot be ensured. Hence, instead of using the Hessian approach, we estimated error bounds by bootstrapping. Specifically, we resampled the threshold data such that the quantity  $\log(E_i)$  of Equation B2 is replaced by a random draw from a Gaussian distribution with mean  $\log(E_i)$  and standard deviation  $\sigma_i$ . We generated 1000 resampled data sets and fit the clipped line function to each of these data sets with the same multi-start algorithm described above for minimizing Equation B2 but with only 100 initial guesses per resampled data set to save computing time. We take the 2.5th and 97.5th percentiles of the resulting marginal distribution of a parameter to be the lower and upper values of the 95% confidence interval of that parameter, respectively. For critical spacing, the resulting distribution

**Table C4**

Experiment 2: Within-subject differences in critical spacings and threshold elevations between conditions at  $\alpha = 0.05$  (N.S. = not significant).

Critical spacing			
Subject	Objects vs. silhouettes	Objects vs. apertures	Apertures vs. silhouettes
AR	N.S.	N.S.	N.S.
JD	>	>	>
TC	>	N.S.	>
Log threshold elevation			
Subject	Objects vs. silhouettes	Objects vs. apertures	Apertures vs. silhouettes
AR	<	<	<
JD	<	<	<
TC	<	>	<

**Table C5**

Experiment 3: Estimated parameters with 95% confidence intervals for subjects JW, KS, and RM.  $S$  is the least squares of the fits (Equation B2).

	$\log(E_{\text{ceiling}})$	$\log(E_{\text{floor}})$	$\log(E_{\text{ceiling}} / E_{\text{floor}})$	$\log(S_{\text{critical}})$	$\log(S_{\text{sat}})$	$\log(S_{\text{critical}} / S_{\text{sat}})$	$S$
<i>JW</i>							
Objects	-1.14 (-1.17, -1.11)	-2.56 (-2.60, -2.53)	1.42 (1.38, 1.47)	0.36 (0.31, 0.38)	0.16 (0.15, 0.18)	0.20 (0.13, 0.23)	159.76
Donuts	-1.02 (-1.05, -1.00)	2.28 (2.33, 2.23)	1.25 (1.19, 1.31)	0.39 (0.37, 0.42)	0.15 (0.13, 0.19)	0.25 (0.18, 0.28)	66.01
<i>KS</i>							
Objects	-1.25 (-1.28, -1.22)	2.49 (2.52, 2.44)	1.24 (1.19, 1.28)	0.46 (0.39, 0.48)	0.24 (0.20, 0.27)	0.22 (0.13, 0.27)	239.88
Donuts	-1.15 (-1.20, -1.12)	2.11 (2.27, 2.07)	0.96 (0.89, 1.12)	0.52 (0.44, 0.76)	0.19 (0.14, 0.33)	0.33 (0.12, 0.62)	102.29
<i>RM</i>							
Objects	-1.24 (-1.29, -1.11)	2.35 (2.39, 2.32)	1.10 (1.04, 1.26)	0.36 (0.36, 0.48)	0.28 (0.01, 0.29)	0.08 (0.07, 0.47)	89.99
Donuts	-1.58 (-1.61, -1.55)	-2.17 (-2.21, -2.14)	0.59 (0.54, 0.63)	0.39 (0.37, 0.47)	0.22 (0.13, 0.25)	0.17 (0.13, 0.32)	64.16

was notably asymmetric in some cases, with long tails toward large critical spacings. We note that with the exception of the few cases of high asymmetry, the error bounds computed with the bootstrapping are similar to those obtained with Equation B3.

For descriptive (as opposed to inferential) purposes, we display the across-subject average of the estimated parameters. No error bounds are computed for the averages since we will not make any formal statistical claims on the averages with the small number of subjects.

## Appendix C

**Table C6**

Experiment 3: Within-subject differences in critical spacings and threshold elevations between conditions at  $\alpha = 0.05$  (N.S. = not significant).

<u>Critical spacing</u>	
<u>Subject</u>	<u>Objects vs. donuts</u>
JW	<
KS	N.S.
RM	N.S.
<u>Log threshold elevation</u>	
<u>Subject</u>	<u>Objects vs. donuts</u>
JW	>
KS	>
RM	>

## References

- Andriessen J, Bouma H. Eccentric vision: Adverse interactions between line segments. *Vision Research*. 1976; 16:71–78. [PubMed: 1258390]
- Baldassi S, Megna N, Burr D. Visual clutter causes high-magnitude errors. *PLoS Biology*. 2006; 4:e56. [PubMed: 16494527]
- Banks W, Larson D, Prinzmetal W. Asymmetry of visual interference. *Perception & Psychophysics*. 1979; 25:447–456. [PubMed: 492909]
- Bi T, Cai P, Zhou T, Fang F. The effect of crowding on orientation-selective adaptation in human early visual cortex. *Journal of Vision*. 2009; 9(11):1–10. 13. <http://www.journalofvision.org/content/9/11/13>, doi:10.1167/9.11.13. [PubMed] [Article]. [PubMed: 20053076]
- Blake R, Tadin D, Sobel K, Raissian T, Chong S. Strength of early visual adaptation depends on visual awareness. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:4783–4788. [PubMed: 16537384]
- Bouma H. Interaction effects in parafoveal letter recognition. *Nature*. 1970; 226:177–178. [PubMed: 5437004]
- Bouma H. Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*. 1973; 13:767–782. [PubMed: 4706350]
- Brainard D. The Psychophysics Toolbox. *Spatial Vision*. 1997; 10:433–436. [PubMed: 9176952]
- Bravo M, Farid H. Object recognition in dense clutter. *Perception & Psychophysics*. 2006; 68:911–918. [PubMed: 17153187]
- Chakravarthi R, Cavanagh P. Recovery of a crowded object by masking the flankers: Determining the locus of feature integration. *Journal of Vision*. 2009; 9(10):1–9. 4. <http://www.journalofvision.org/content/9/10/4>, doi:10.1167/9.10.4. [PubMed] [Article].
- Chambers L, Wolford G. Lateral masking vertically and horizontally. *Bulletin of the Psychonomic Society*. 1983; 21:459–461.
- Chung S, Bedell H. Effect of retinal image motion on visual acuity and contour interaction in congenital nystagmus. *Vision Research*. 1995; 35:3071–3082. [PubMed: 8533343]
- Chung S, Levi D, Legge G. Spatial frequency and contrast properties of crowding. *Vision Research*. 2001; 41:1833–1850. [PubMed: 11369047]
- Chung S, Tjan B. Shift in spatial scale in identifying crowded letters. *Vision Research*. 2007; 47:437–451. [PubMed: 17223153]
- Ehlers H. V: The movements of the eyes during reading. *Acta Ophthalmologica*. 1936; 14:56–63.

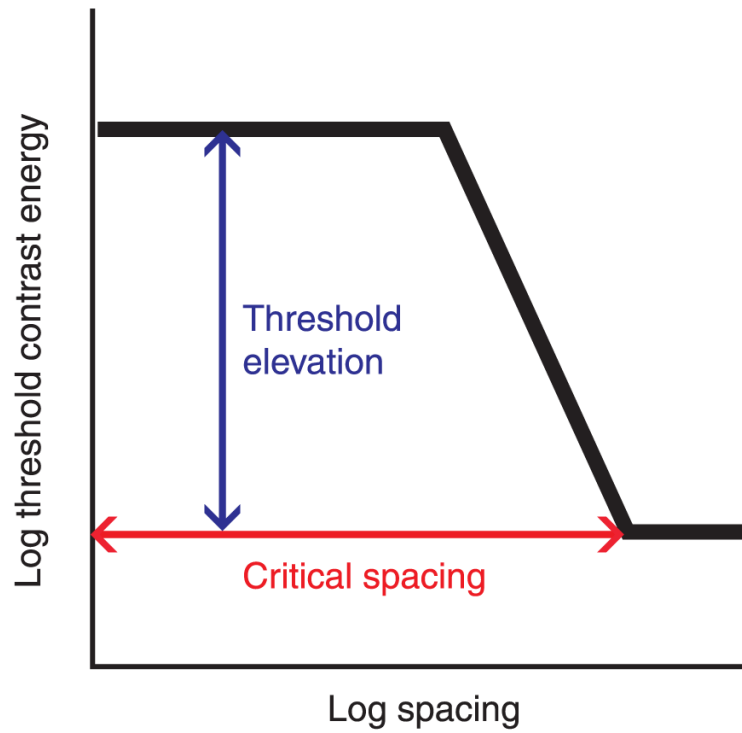
- Farzin F, Rivera S, Whitney D. Holistic crowding of Mooney faces. *Journal of Vision*. 2009; 9(6):1–15. 18. <http://www.journalofvision.org/content/9/6/18>, doi:10.1167/9.6.18. [PubMed] [Article]. [PubMed: 19761309]
- Flom M, Heath G, Takahashi E. Contour interaction and visual resolution: Contralateral effects. *Science*. 1963; 142:979–980. [PubMed: 14069233]
- He S, Cavanagh P, Intriligator J. Attentional resolution and the locus of visual awareness. *Nature*. 1996; 383:334–337. [PubMed: 8848045]
- Intriligator J, Cavanagh P. The spatial resolution of visual attention. *Cognitive Psychology*. 2001; 43:171–216. [PubMed: 11689021]
- Kanwisher N, McDermott J, Chun M. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*. 1997; 17:4302–4311. [PubMed: 9151747]
- Korte W. Über die Gestaltauffassung im indirecten Sehen. *Zeitschrift für Psychologie*. 1923; 93:17–82. 1923.
- Levi D. Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*. 2008; 48:635–654. [PubMed: 18226828]
- Levi D, Carney T. Crowding in peripheral vision: Why bigger is better. *Current Biology*. 2009; 19:1988–1993. [PubMed: 19853450]
- Levi D, Hariharan S, Klein S. Suppressing and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*. 2002; 2(2):167–177. 3. <http://www.journalofvision.org/content/2/2/3>, doi:10.1167/2.2.3. [PubMed] [Article]. [PubMed: 12678590]
- Levi D, Klein S, Hariharan S. Suppressing and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking. *Journal of Vision*. 2002; 2(2):140–166. 2. <http://www.journalofvision.org/content/2/2/2>, doi:10.1167/2.2.2. [PubMed] [Article]. [PubMed: 12678589]
- Liu T, Jiang Y, Sun X, He S. Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli. *Current Biology*. 2009; 19:127–132. [PubMed: 19135367]
- Louie E, Bressler D, Whitney D. Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*. 2007; 7(2):1–11. 24. <http://www.journalofvision.org/content/7/2/24>, doi:10.1167/7.2.24. [PubMed] [Article]. [PubMed: 18217839]
- Marr, D. *Vision*. The MIT Press; 1982. ISBN-10: 0-262-51462-1 ISBN-13: 978-0-262-51462-0
- Martelli M, Majaj N, Pelli D. Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*. 2005; 5(1):58–70. 6. <http://www.journalofvision.org/content/5/1/6>, doi: 10.1167/5.1.6. [PubMed] [Article]. [PubMed: 15831067]
- Nandy AS, Tjan BS. The nature of letter crowding as revealed by first- and second-order classification images. *Journal of Vision*. 2007; 7(2):1–26. 5. <http://www.journalofvision.org/content/7/2/5>, doi: 10.1167/7.2.5. [PubMed] [Article]. [PubMed: 18217820]
- Parkes L, Lund J, Angelucci A, Solomon J, Morgan M. Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*. 2001; 4:739–744.
- Pelli D. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
- Pelli D. Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*. 2008; 18:445–451. [PubMed: 18835355]
- Pelli D, Palomares M, Majaj N. Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*. 2004; 4(12):36–69. 11. <http://www.journalofvision.org/content/4/12/12>, doi:10.1167/4.12.12. [PubMed] [Article].
- Pelli D, Tillman K. The uncrowded window of object recognition. *Nature Neuroscience*. 2008; 11:1129–1135.
- Pelli D, Zhang L. Accurate control of contrast on microcomputer displays. *Vision Research*. 1991; 31:1337–1350. [PubMed: 1891822]

- Petrov Y, Popple A, McKee S. Crowding and surround suppression: Not to be confused. *Journal of Vision*. 2007; 7(2):1–9. 12. <http://www.journalofvision.org/content/7/2/12>, doi:10.1167/7.2.12. [PubMed] [Article].
- Press, WH. Numerical recipes in C: The art of scientific computing (vol. 2). Cambridge University Press; 1992.
- Rosenholtz R, Li Y, Nakano L. Measuring visual clutter. *Journal of Vision*. 2007; 7(2):1–22. 17. <http://www.journalofvision.org/content/7/2/17>, doi:10.1167/7.2.17. [PubMed] [Article]. [PubMed: 18217832]
- Shin K, Wallace JM, Tjan BS. Objects crowded by noise flankers [Abstract]. *Journal of Vision*. 2010; 10(7):1336, 1336a. <http://www.journalofvision.org/content/10/7/1336>, doi:10.1167/10.7.1336.
- Strasburger H, Harvey L, Rentschler I. Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*. 1991; 49:495–508. [PubMed: 1857623]
- Stuart J, Burian H. A study of separation difficulty. Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology*. 1962; 53:471–477. [PubMed: 13917936]
- Tjan BS, Braje WL, Legge GE, Kersten D. Human efficiency for recognizing 3-D objects in luminance noise. *Vision Research*. 1995; 35:3053–3069. [PubMed: 8533342]
- Tjan BS, Dang S. The spatial interaction zone of a shapeless noise flanker [Abstract]. *Journal of Vision*. 2005; 5(8):227, 227a. <http://www.journalofvision.org/content/5/8/227>, doi: 10.1167/5.8.227.
- Toet A, Levi D. The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*. 1992; 32:1349–1357. [PubMed: 1455707]
- Townsend J, Taylor S, Brown D. Lateral masking for letters with unlimited viewing time. *Perception & Psychophysics*. 1971; 10:375–378.
- Tripathy S, Cavanagh P. The extent of crowding in peripheral vision does not scale with target size. *Vision Research*. 2002; 42:2357–2369. [PubMed: 12350424]
- Ullman S, Vidal-Naquet M, Sali E. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*. 2002; 5:682–687.
- van den Berg R, Cornelissen F, Roerdink J. A crowding model of visual clutter. *Journal of Vision*. 2009; 9(4):1–11. 24. <http://www.journalofvision.org/content/9/4/24>, doi:10.1167/9.4.24. [PubMed] [Article].
- Watson A, Pelli D. QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*. 1983; 33:113–120. [PubMed: 6844102]
- Wolford G, Chambers L. Lateral masking as a function of spacing. *Perception & Psychophysics*. 1983; 33:129–138. [PubMed: 6844104]

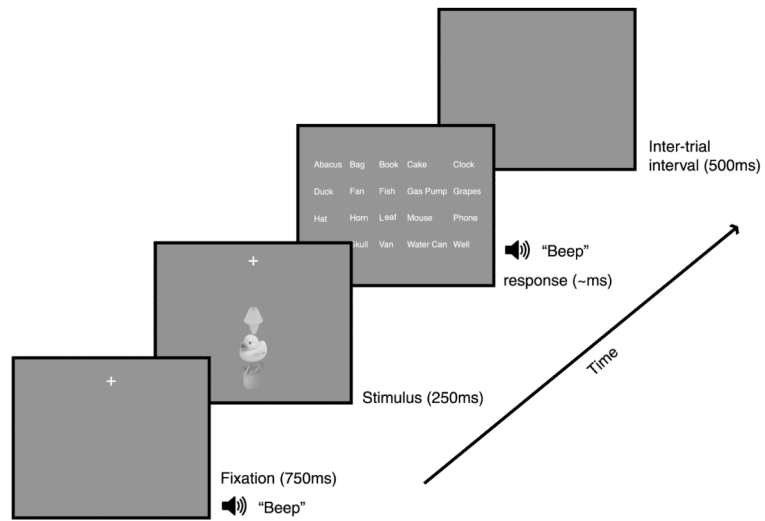


Y + XYZ

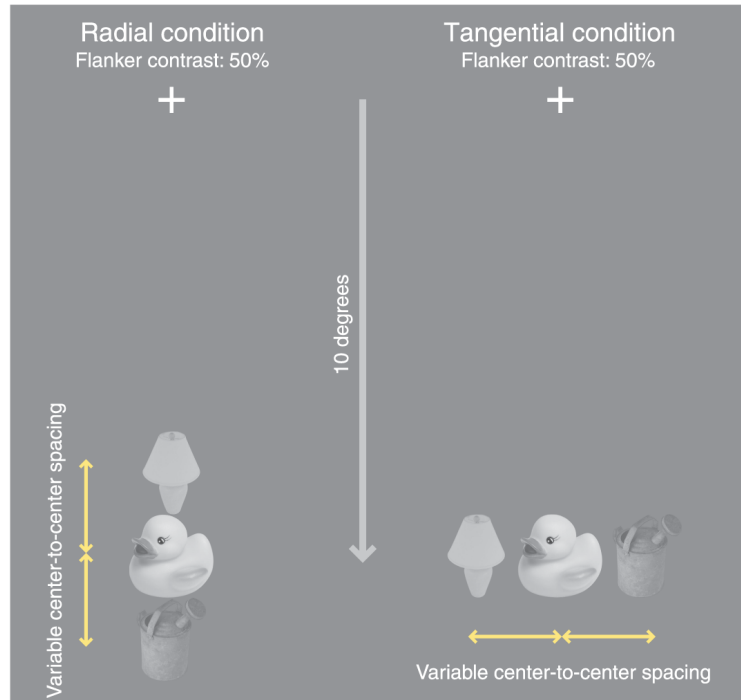
**Figure 1.**  
Demonstration of peripheral crowding. While fixating on the central fixation cross, the Y on the left can be identified easily, but the Y on the right is harder to identify due to the presence of nearby letters, even though both Ys are at the same distance from fixation.



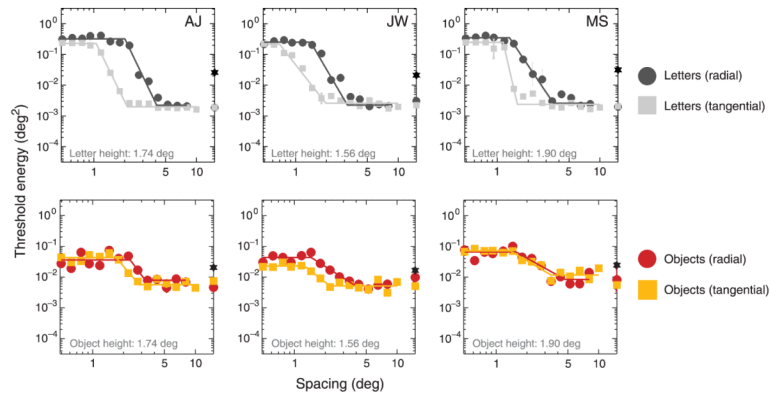
**Figure 2.** Quantifying crowding. Log threshold contrast energy is expressed as a function of the log center-to-center spacing between the target and a flanker. Two parameters of this function are of interest to us: Threshold elevation, which describes the ratio of contrast energy thresholds at ceiling and floor, and critical spacing, which is the smallest spacing with negligible threshold elevation.



**Figure 3.** Timeline of a trial. After viewing the stimulus, subjects are presented with a list of the names of possible target objects and make a mouse click on the name of the object they identified as the target.



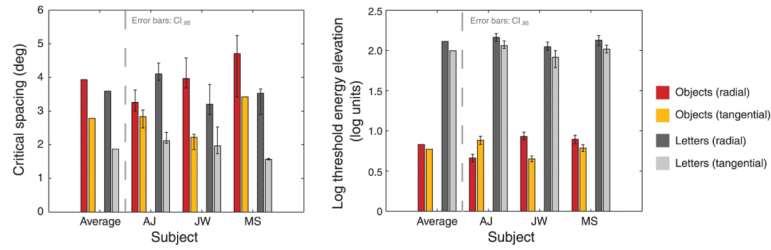
**Figure 4.** Example of the stimuli and stimulus arrangements used in object conditions of Experiment 1.



**Figure 5.**

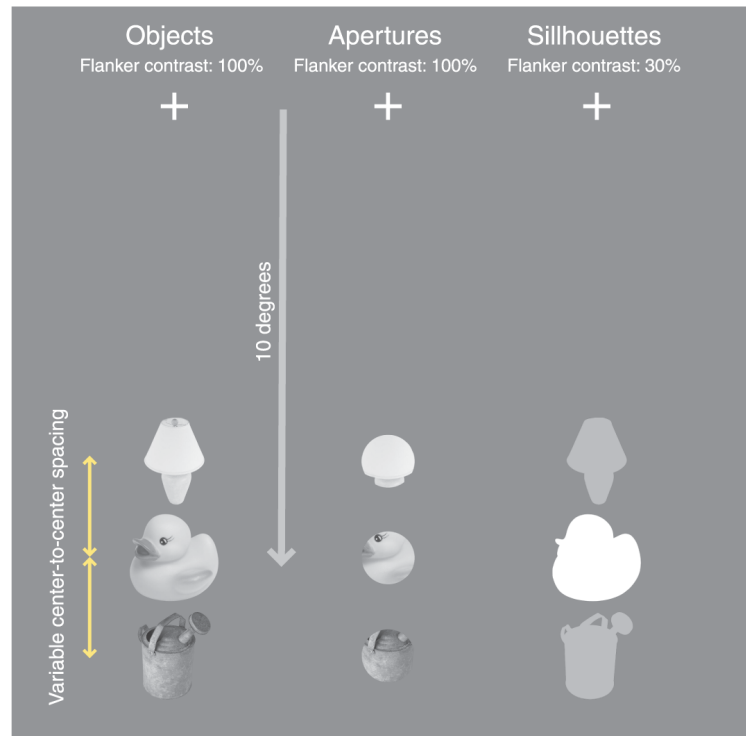
Results for Experiment 1. Threshold contrast energy as a function of the center-to-center spacing between targets and flankers. Results for the different observers are presented in different columns. On the right ordinate are data points for the unflanked target (infinite spacing). The star symbol indicates the flanker energy. Letter thresholds are presented in the upper plots, while object thresholds are presented in the lower plots. Error bars are  $\pm 1$  SE and some are smaller than the plot symbols.



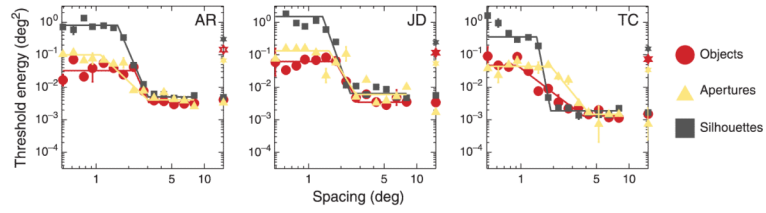


**Figure 6.**

Results for Experiment 1. The critical spacing is plotted on the left, while log threshold energy elevation is plotted on the right (a factor of two increase in threshold corresponds to 0.3 log unit). Asymmetric error bars represent 95% confidence intervals. There are no error bars on the averages, which are provided for descriptive purposes only. For every subject, critical spacing is larger in radial than tangential conditions, for both objects and letters. Threshold elevation is larger for letters than objects.

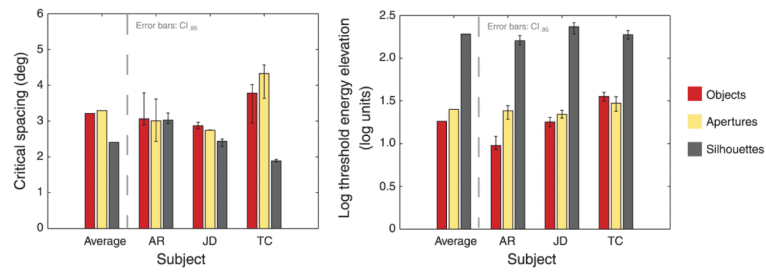


**Figure 7.** Example of stimuli used in Experiment 2. Only the radial stimulus arrangement is used in this experiment.

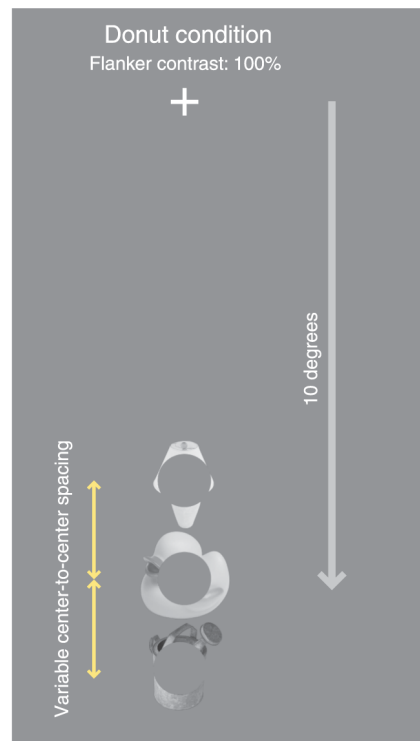


**Figure 8.**

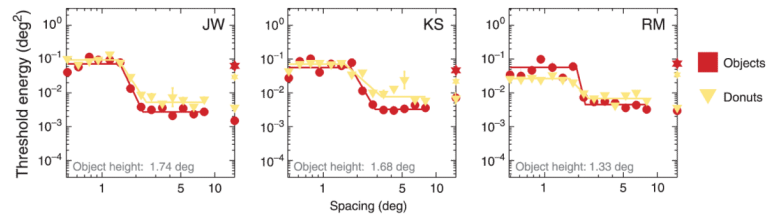
Results for Experiment 2. Threshold contrast energy is plotted as a function of center-to-center spacing between targets and flankers, for three subjects. The different object conditions are plotted in different colors on each plot. On the right ordinate are data points for the unflanked (infinite spacing) targets, for the different conditions. The star symbols indicate the flanker energy for the different conditions. The functions are similar between objects and apertures, while the ceiling for silhouette stimuli is elevated relative to the other conditions.



**Figure 9.** Results for Experiment 2. Critical spacings are plotted on the left, while log threshold elevations are plotted on the right. Critical spacings are very similar for all the conditions. Threshold elevations, while similar for objects and apertures, are much higher for silhouettes.

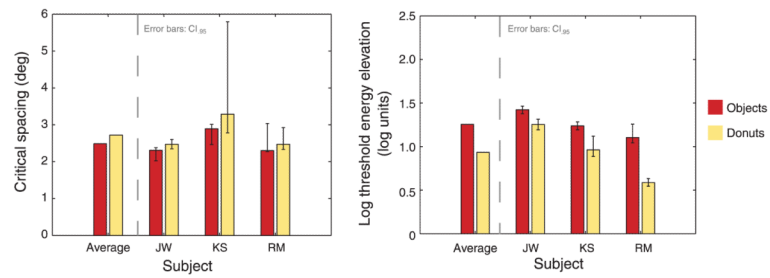


**Figure 10.** Example of stimulus used for Experiment 3. Only the radial stimulus arrangement was used. In addition to the donut stimulus, the intact object condition of Experiment 2 was also used.



**Figure 11.**

Results for Experiment 3. Threshold contrast energy is plotted as a function of center-to-center spacing between targets and flankers, for 3 subjects. On the right ordinate are data points for the unflanked (infinite spacing) targets for the different conditions. The star symbols indicate the flanker energy for the different conditions. The functions for objects and donuts are similar.



**Figure 12.** Results for Experiment 3. Critical spacings are plotted on the left, while log threshold energy elevations are plotted on the right. Both critical spacing and log threshold elevation are quite similar between objects and donuts.