Combining local and global limitations of visual search

Endel Põder
Institute of Psychology, University of Tartu, Estonia

There are different opinions about the roles of local interactions and central processing capacity in visual search. This study attempts to clarify the problem using a new version of relevant set cueing. A central precue indicates two symmetrical segments (that may contain a target object) within a circular array of objects presented briefly around the fixation point. The number of objects in the relevant segments, and density of objects in the array were varied independently. Three types of search experiments were run: (a) search for a simple visual feature (color, size, and orientation); (b) conjunctions of simple features; and (c) spatial configuration of simple features (rotated Ts). For spatial configuration stimuli, the results were consistent with a fixed global processing capacity and standard crowding zones. For simple features and their conjunctions, the results were different, dependent on the features involved. While color search exhibits virtually no capacity limits or crowding, search for an orientation target was limited by both. Results for conjunctions of features can be partly explained by the results from the respective features. This study shows that visual search is limited by both local interference and global capacity, and the limitations are different for different visual features.

Introduction

Traditionally, visual search has been used to study global/central processing capacity. In early studies, the main idea was to separate parallel and serial processing modes in vision (Treisman & Gelade, 1980; Bergen & Julesz, 1983; Wolfe, Cave, & Franzel, 1989). Usually, the effect of the number of objects (set size) on reaction time has been measured. According to a simple view, this effect should be close to zero with parallel search, and much larger with serial search. Serial processing can also be mediated by eye movements, but the idea of covert movement of the focus of attention and its role in the visual processing has been more interesting.

Gradually, the idea of internal noise and signal detection models were accepted in visual search studies (Kinchla, 1974; Shaw, 1980). In these studies, proportion correct is a usual measure of performance, and the main question is about processing capacity limitations, rather than parallel–serial dichotomy. It was realized that some impairment of performance with increasing set size can be explained by the integration of increasing number of noisy signals, and is expected even without any capacity limitations. In 1990s it was shown by several researchers (Palmer, Ames & Lindsey, 1993; Palmer, 1994; Foley & Schwarz, 1998) that search for a difference in simple features like orientation, size, contrast, or color is consistent with unlimited processing capacity. Eckstein (1998) demonstrated that an unlimited capacity signal detection theory (SDT) model can explain set-size effects in conjunction search as well. Still, there are other combinations of visual features that definitely do not fit this model. Shaw (1984) found that search for letters is not consistent with unlimited capacity. Põder (1999) and Palmer, Fencsik, Flusberg, Horowitz, and Wolfe (2011) have shown that search for a target defined by a relative position (or a spatial configuration) of simple features fits best to a strict limited capacity model. Not surprisingly, search for even more complex objects (words, 3D object categories) is also consistent with fixed capacity models (Scharff, Palmer, & Moore, 2011, 2013). Gilden, Thornton, and Marusich (2010) have argued that search for certain types of relative-position stimuli conforms to a serial model.

The serial model has been traditionally applied to reaction-time (RT) experiments. However, with a brief presentation of a search display, this model also predicts strong set-size effects on proportion correct (Bergen & Julesz, 1983). When relevant set size exceeds the number of objects that can be processed during the exposure, predicted proportion correct (corrected for guessing) is inversely proportional to set size.

Of course, visual search is limited not only by central processing, but also by lower-level factors like masking, crowding, grouping, and drop of spatial resolution in the periphery. Visual search studies have usually either ignored or tried to eliminate or control these factors. Several studies (Hoffman, 1979; Cohen & Ivry, 1991; Xu, 2010) have used “spread” and “clumped” spatial configurations in order to gain some understanding of the role of local interactions. These studies revealed...
several useful results about the extent and importance of the local effects. However, the accurate interpretation is difficult, especially when eye movements were not controlled.

In early studies of visual search, the role of eye movements was frequently overlooked. It was widely believed that covert shifts of attention are nearly equivalent to eye movements and restriction of these does not affect the nature of visual search (e.g., Klein & Farrell, 1989). However, some more recent studies have argued that a large part of regularities of visual search can be explained by accurate modeling of low-level limitations of peripheral vision and regularities of eye movements (Geisler & Chou, 1995).

During the last decade, local adverse interactions, or crowding (Bouma, 1970; Andriessen & Bouma, 1976) have become an important topic in vision research. Crowding is almost absent in the fovea while more peripherally its spatial extent is roughly proportional to the eccentricity of the target (Toet & Levi, 1992; Pelli, Palomares, & Majaj, 2004). Peripheral (crowded) vision is quite similar to preattentive perception of visual textures where certain statistical regularities are seen, but exact spatial relations are lost (e.g., Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). That idea was effectively implemented in Portilla and Simoncelli (2000) texture model. This model has been used in several studies on crowding (Balas, Nakano & Rosenholtz, 2009; Freeman & Simoncelli, 2011).

Some researchers (Wertheim, Hooge, Krikke, & Johnson, 2006; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012) have concluded that crowding has more important role in visual search than previously thought. Rosenholtz, Huang, and Ehinger (2012) and Rosenholtz, Huang, Raj et al. (2012) have proposed that efficiency of visual search may be fully explained by local processes within a classic zone of visual crowding combined with eye movements, and the limited capacity attentional processing is unnecessary.

It is possible that some limitations attributed originally to central processing can be better explained by regularities of low-level vision, local interactions, and eye movements. Still, it seems unlikely that all the effects of attentional processing can be simply explained away.

There are only a few studies that have combined the analysis of set size and crowding effects in the same visual search experiments, and the results are far from conclusive. Pöder (2004) used bisected rectangles as stimuli that could be arranged for both simple feature and relative position search. He varied set size and interobject spacing as independently as possible, in a simple circular configuration. He found large differences in set-size effects: simple feature search was consistent with unlimited capacity, and relative position search was closer to a fixed capacity model. The spatial extent of crowding was also larger for relative position condition, but the difference was not statistically significant.

Neri and Levi (2006) studied search for objects of odd contrast polarity, odd orientation, and odd conjunction of polarity and orientation. They varied spacing of objects together with their size. Number of objects (set size) was also varied, from 16 to 64. They found that, in the visual periphery, both spacing and set size effects were larger for conjunction search as compared to feature search. The results were consistent with their model based on an imperfect registration of two feature maps.

Reddy and VanRullen (2007) used four kinds of search stimuli—two simple and two complex. They found set-size effects consistent with limited capacity for complex stimuli and no set-size effects for simple ones. Spacing (crowding) effect was found for one stimulus set only—search for an upright face among inverted faces. However, the interobject distances in this study were large (in eccentricity units) as compared to usual crowding studies, and backward masking was used to limit performance.

The goal of the present study was to explore more systematically the set-size and spacing effects in visual search. This study uses several classic search stimuli: simple features, their conjunctions, and spatial configuration stimuli. Relevant set size and interobject distance were varied independently using a version of spatial cueing.

Methods

Experiments

There were six types of experiments run in this study: three kinds of feature search (orientation, color, and size); two kinds of conjunction (orientation & color, size & color); and a search for relative position/feature configuration (rotated Ts).

Stimuli

Stimuli for feature and conjunction search experiments were small bright ellipses or rings depicted on a dark gray background (Figure 1). For orientation and color search, ellipses with aspect ratio 2 (size 14 × 7 pixels) were used. From a viewing distance of 60 cm, the size of stimuli was about 0.50° × 0.25°. The orientation was varied relative to vertical, so the ellipses could be tilted either left or right, by the same degree. The color was varied by the ratio of R and B values in RGB color code. Thus, the objects could be either orangish or bluish. In the size search experiments, rings with two
different diameters were used. The smaller object had a fixed diameter of 9 pixels (approx. 0.35°).

In the feature search experiments, a target differed from the distractors by a fixed simple feature. For example, an observer could search for a left-tilted ellipse among right-tilted ones, or for a large ring among small rings. Another (irrelevant) feature was varied independently (for example, both

Figure 1. Examples of stimuli. Maximum (A) and minimum (B) interobject distance; relevant set sizes 8 (C) and 2 (D); search for color–orientation conjunction (A–D); color (E); orientation (F); size (G); and rotated T (H).
small and large rings could be either orangish or bluish).

In a conjunction search experiment, the target was a fixed conjunction of two features. The distractors differed from the target by a single feature. For example, when the target was a left-tilted bluish ellipse, distractors were right-tilted bluish and left-tilted orangish ellipses, in approximately equal proportions.

In the relative position (or feature configuration) search, rotated Ts were used as stimuli. One orientation from the possible four was selected as a target. The Ts with remaining three orientations were distractors. An observer ran separate blocks with four different targets.

In order to approximately equate the visibility, the objects were displayed along an imaginary ellipse around the fixation point. Average eccentricity varied from 3.75° to 5.00°, along vertical and horizontal meridians, respectively. A small jitter (amplitude 0.2°) was added to the positions of objects in radial direction, in order to avoid alignment cues. In tangential direction, the objects were located in equal angular steps around the fixation. The number of displayed objects was varied from 8 to 48. This number determines nearest neighbor distance between the objects.

Relevant set size was varied (from 2 to 8) independently from the number of displayed objects. In order to simplify the attentional selection and avoid eye movements, relevant objects were located in two sectors around the horizontal meridian. The angular positions of the relevant objects were indicated by a central cue (black lines with radius of 0.6°). Relevant set size (and the positions of relevant objects) were fixed within blocks of trials and the cue was permanently visible during a block.

Procedure

In all experiments, a simple yes–no task was used. Stimuli were presented for 60 ms. The observer had to determine whether the target object was present or not in a given trial. The probability of the target presence was 0.5.

Before the main experiments, observers ran small staircase experiments searching for the same simple features in the simplest conditions (relevant set size 2, large interobject distance). The results were used to select appropriate target–distractor differences for the main experiments (proportion correct about 0.95 for the simplest condition). For rotated Ts, object size was adjusted. In the main experiments, 100 trials per each combination of relevant set size and interobject distance were run. The experiments were run in blocks of 25 or 50 trials, counterbalanced in order. Set size and interobject distance were fixed within a block.

Observers

Ten observers (five female, five male), including the author, took part in the experiments. They completed different numbers of experiments (from one to six). The number of observers per each experiment (search task) varies from four to seven (indicated in Figure 2).

Modeling

As a first approximation, the data (as given in Figure 2) can be described by two independent, multiplicatively combined effects of crowding and set size, in terms of $d'$ (Pöder, 2004). The crowding effect is described as a Gaussian function of target–flanker distance (Levi, Klein, & Hariharan, 2002), and set-size affects $d'$ according a power function:

$$d' = \frac{d'^1}{n^b \cdot (1 + a \cdot e^{-\frac{d^s}{2\sigma^2}})}, \quad (1)$$

where $d'^1$ is $d'$ for set size 1 and no crowding, $n$ is set size, $b$ is a measure of set-size effect, $a$ is maximum crowding effect (when interobject distance drops to 0), $d^s$ is interobject distance, and $\sigma$ is spatial extent (standard deviation) of Gaussian crowding zone.

This simple descriptive model has a problem with theoretical interpretation, because the observed set-size effect can be caused by at least two factors: limited processing capacity and integration of information from multiple objects; and the effect of integration stage usually cannot be calculated by a simple formula. Therefore, I use a more “theoretical” approach, similar to that of McLean (1999) and Mazyar, Van den Berg, & Ma (2012). I assume that Equation 1 gives a local signal-to-noise ratio (SNR) after the possible capacity limitation and crowding effects, but in order to predict search performance, it should be complemented by a decision model. That model calculates d-prime for yes–no search task ($d'_{\text{no}}$) as a function of local SNR ($d'$) and set size:

$$d'_{\text{no}} = f(d', n). \quad (2)$$

In this study, I used the ideal decision model (e.g., Mazyar et al., 2012). This model calculates the likelihoods of the observed signals under the hypotheses of target present and target absent and selects the one with the higher total likelihood. Assuming equal priors, Gaussian noise, and selecting internal variable for distractor and target, $x_D = -0.5$, and $x_T = 0.5$, log likelihood ratio (target present /target absent) for a single trial is

$$L = \log \left( \frac{1}{n} \sum_{i=1}^{n} e^{x_i} \right), \quad (3)$$

where $x_i$ is noisy internal variable for object $i$, and $\sigma$ is
standard deviation of noise. The ideal model selects “target present” when \( L > 0 \), and “target absent” otherwise.

With this model, 100,000 simulated trials were run for a set of SNRs and set sizes, and corresponding \( d_{\text{th}} \) was calculated for the yes–no search. The results follow slightly curved lines in log-log coordinates. The simulation results can be very precisely (\( R^2 > 0.999 \)) approximated by a relatively simple polynomial formula:

\[
d_{\text{th}} = \frac{d'}{\mu^{0.40 - 0.35 \log d' - 0.21\log^2 d'}, \text{ for } d' = 0.25 \ldots 4 \text{ and } n = 1 \ldots 16.
\]

In the model fitting procedure, this approximation was used rather than running simulations at each iteration.

SDT models of visual search (Palmer, Verghese, & Pavel, 2000) predict that set-size effect (in terms of \( d' \)) should be larger for conditions with larger noise (lower

---

**Figure 2.** Experimental results across the search conditions. Mean proportions correct as dependent on interobject distance and relevant set size. Error bars indicate standard errors of the mean. \( N = \) number of observers.
performance). Therefore, assuming that crowding increases noise, the set-size effects should increase with crowding. The present data indicate that there is rather an opposite trend, which may indicate that usual SDT assumptions do not hold in crowding conditions (Palmer, 1994; Palmer et al., 2000). A simple modification may, at least partly, correct the problem.

Crowding has been frequently described as an obligatory pooling of signals within a built-in integration field (Parkes et al., 2001). Thus, it is possible that, aside from a crowd of other objects, signal is considered as having an effect on the process of integration. Moreover, it is possible that, aside from a crowd of other objects, signal is considered as having an effect on the process of integration. This process might be caused by the effective set size, which is increased in proportion to irrelevant objects. For example, when two relevant objects are densely surrounded by irrelevant ones, the effective set size might be 4 or 6, and it does not change much when changing the relevant set size only.

In order to model the pooling effects, relevant set size \( n \) was replaced with an effective set-size \( n_e \). It was supposed that instead of an “ideal” window of spatial attention (with weight of one in the positions of relevant objects, and zero elsewhere), the observer had to use a blurred window—the ideal one convolved with a Gaussian filter with a standard deviation \( s \). Effective set size was increased in proportion to irrelevant objects that were sampled by that window.

\[
\sum_{i=1}^{N} w_i \quad n_e \quad n \quad \frac{\sum_{i=1}^{n} w_i}{\sum_{j=1}^{N} w_j}, \quad (5)
\]

where \( n \) is the number of relevant objects, \( N \) is the number of displayed objects, and \( w_i \) and \( w_j \) are the weights of the attentional window at the positions of objects \( i \) and \( j \). This pooling mechanism does not need additional free parameters as its effect is determined by the spatial parameter \( s \) already present in the original crowding mechanism.

For example, with \( s = 0.2E \), which corresponds to usual crowding zone \( 0.5E \), and minimum interobject distance \( d = 0.13E \), this calculation yields effective set sizes 7.7, 8.5, and 11.6, for relevant set sizes 2, 4, and 8, respectively. When \( d > 2.5s \), the pooling effect virtually disappears (\( n_e \approx n \)).

The main model has four free parameters: one \( b \) for capacity limitation, two \( a \) and \( s \) for crowding effect, and one \( d' \) for overall level of performance. The model predicts \( d' \)'s for yes–no visual search task. The predicted \( d' \)'s can be converted into the predictions of unbiased proportion correct

\[
P_c = \Phi\left(\frac{d_n}{\sqrt{2}}\right), \quad (6)
\]

where \( \Phi \) is standard normal distribution function (Macmillan & Creelman, 1991).

The analysis of hits and false alarms in the experimental data revealed a large variability of decision criteria across observers and experimental conditions. In order to remove the effect of criteria, the empirical hits and false alarms were transformed into the unbiased proportion correct

\[
P_c = \Phi\left[\frac{z(H) - z(F)}{2}\right], \quad (7)
\]

where \( H \) and \( F \) are proportions of hits and false alarms, \( \Phi \) is standard normal distribution function, and \( z \) is the inverse of the normal distribution function. An individual data set consisted usually of 18 proportions correct. MS Excel Solver was used to find maximum likelihood parameters of the model by minimizing the likelihood ratio statistic \( G \).

In order to test the importance of different components of the model, I attempted to fit four simpler versions besides the full four-parameter model:

- unlimited capacity SDT model \( (b = 0) \), three parameters;
- model with no crowding effects \( (a = 0 \) and \( n_e = n ) \), two parameters;
- model without noise effect from crowding \( (a = 0 \) and \( n_e = n ) \), three parameters; and
- model without pooling of distractor objects \( (n_e = n ) \), four parameters.

Also, I tried a version of serial search model. This model assumes that an observer can attend a fixed number of objects \( k \) in each trial. In order to account for the imperfect performance for small set sizes, this capacity limitation was combined with noisy percepts and ideal decision making. When relevant set size does not exceed the capacity limit \( (n \leq k) \), the model behaves like unlimited capacity SDT model \( (b = 0) \). With \( n > k \), proportion correct (corrected for guessing) drops in proportion to \( k/n \). Thus, the proportion correct is

\[
P_c = 0.5 + \frac{k}{n} \left[ \Phi\left(\frac{d_n}{2}\right) - 0.5\right], \quad \text{for } n > k, \quad (8)
\]

\[
P_c = \Phi\left(\frac{d_n}{2}\right), \quad \text{for } n \leq k. \quad (9)
\]

This model also has four free parameters. It also included both crowding mechanisms.

**Results**

Average proportions correct for different search tasks, as dependent on interobject distance and relevant set size, are shown in Figure 2. With exception of color search, the graphs are qualitatively similar; however, it
is quite clear that the effects of the two independent variables are different across the search tasks.

Figure 3 depicts examples of individual unbiased proportions correct, together with fits of the limited capacity SDT model. Overall, the fits were reasonably good. For 19 out of 34 data sets, there were no significant differences between the data and the model, 15 had differences with significance $p < 0.05$ (seven had $p$-values less than 0.01).

In order to test importance of different components of the model, I also attempted to fit four simpler versions (unlimited capacity, no crowding, no noise crowding, no spatial pooling) as well. Average goodness-of-fit statistics for different tasks and different models are shown in Table 1.

The results show that both limited capacity and crowding are necessary: the unlimited capacity model can marginally fit the data for a size-based search only, and without crowding effects, the fit is very poor for all search conditions. Inclusion of either one or another crowding mechanism improves the fit dramatically. The noise mechanism alone explains the data better than the pooling mechanism, but the latter still makes a small additional contribution.

Three simpler models (unlimited capacity, no crowding, and no noise crowding) are nested within the limited capacity SDT model and they can be compared using a likelihood ratio test. According to the test, all of these models are significantly inferior, as compared to the limited capacity model ($p < 10^{-6}$ for unlimited capacity, $p < 10^{-12}$ for no crowding, and $p < 0.05$ for no noise crowding model). The limited capacity SDT model, the model without spatial pooling, and the serial model have equal numbers of parameters. For these models, differences in $G$-statistic are equivalent to the differences in Akaike information criterion (AIC). Although the limited capacity SDT model has the minimum $G$-statistic, the observed differences (1.7 and 2.4) are too small to exclude the other two models.

The optimal values of the two most informative parameters from the limited capacity SDT model are shown in Figure 4A and B. Parameter $b$ is a measure of processing capacity limitations. It shows how much the relevant set size affects precision of individual object representations: $b = 0$ for independent processing (unlimited capacity), $b = 1$ for the fixed capacity (sample size) model, where variance of noise of the internal representations is proportional to set size, and $b = 2.0$ for the serial search model. Crowding is

<table>
<thead>
<tr>
<th>Search conditions</th>
<th>Limited capacity SDT model</th>
<th>Unlimited capacity</th>
<th>No crowding</th>
<th>No noise crowding</th>
<th>No spatial pooling</th>
<th>“Serial” model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orientation</td>
<td>19.9</td>
<td>43.7**</td>
<td>124.9**</td>
<td>29.0*</td>
<td>19.7</td>
<td>27.4*</td>
</tr>
<tr>
<td>Size</td>
<td>20.5</td>
<td>26.3*</td>
<td>110.6**</td>
<td>33.7**</td>
<td>24.6</td>
<td>21.2</td>
</tr>
<tr>
<td>Color</td>
<td>26.3*</td>
<td>35.0**</td>
<td>36.2**</td>
<td>26.2*</td>
<td>30.8**</td>
<td>29.3**</td>
</tr>
<tr>
<td>Orientation &amp; color</td>
<td>22.2</td>
<td>63.9**</td>
<td>96.5**</td>
<td>27.6*</td>
<td>22.2</td>
<td>22.3</td>
</tr>
<tr>
<td>Size &amp; color</td>
<td>24.5*</td>
<td>42.7***</td>
<td>71.2**</td>
<td>29.2*</td>
<td>25.5*</td>
<td>26.7*</td>
</tr>
<tr>
<td>Rotated Ts</td>
<td>20.6*</td>
<td>77.8**</td>
<td>106.5**</td>
<td>24.5*</td>
<td>22.0*</td>
<td>21.1*</td>
</tr>
<tr>
<td>Total average</td>
<td>22.5</td>
<td>47.6**</td>
<td>89.7**</td>
<td>28.4*</td>
<td>24.2*</td>
<td>24.9*</td>
</tr>
</tbody>
</table>

Table 1. Goodness of fit of different models for different search conditions (average values of likelihood ratio statistic $G$). Statistically significant deviations from the data: **$p < 0.01$, *$p < 0.05$. 

Figure 3. Examples of individual data and model fits. Symbols depict data (unbiased proportions correct), and lines are model fits. Set sizes: 2 (blue circles), 4 (red rectangles), and 8 (green triangles).
traditionally characterized by critical distance—the maximum extent of crowding zone, in eccentricity units. Parameter \( s \) is standard deviation of the assumed crowding zone with Gaussian profile. In order to simplify comparison with crowding studies, it is scaled by factor 2.5 that results in a critical distance measure, where crowding effect drops below 0.05 of its maximum value.

There are considerable differences across observers, but several regularities are still visible. The spatial configuration stimuli (rotated Ts) apparently conform to strong capacity limitations and standard crowding zones. For simple features and their conjunctions, the results are different, dependent on the features involved. While color search exhibits virtually no capacity limits or crowding, search for an orientation target appears to be limited by both. Search for conjunctions tends to have stronger capacity limitations, as compared to corresponding simple features.

I used Tukey’s honest significant difference (HSD) test for pair-wise comparison of different search conditions (Figure 4B). With respect to capacity limitations, both rotated Ts and orientation–color conjunctions were significantly different from color \( (p < 0.01) \) and size \( (p < 0.01) \); rotated Ts differed also from orientation \( (p < 0.05) \) and color–size conjunction \( (p < 0.01) \). For crowding distance, the only significant difference was between orientation and color searches \( (p < 0.01) \). There appears to be a positive correlation between capacity limitations and crowding at the level of search conditions (Figure 4B), which is, however, statistically not significant.
Overall, the parameter of capacity limitation $b$ was larger than the expected range of 0 to 1. For the simple feature search, it was mostly larger than 0, and for an example of complex search, rotated Ts, it was about 1.5. One possible explanation might be a different efficiency of the relevant set selection for different set sizes. If spatial selection was very efficient with two relevant objects and not so efficient with eight objects, the data should exhibit some extra set-size effect. A slightly better performance in the condition without irrelevant objects, when spatial selection is not needed (set size 8, maximum interobject distance), seems to support this hypothesis. These data also suggest that selection explains relatively small part of the total set-size effect.

The interpretation of parameter $a$ in the model with two crowding mechanisms would be problematic because it represents only a part of crowding effect associated with increase of noise. It is also affected by a trade-off between the two mechanisms that is not well understood. Therefore, the parameters of crowding from the simpler model, with only the noise mechanism, are shown in Figure 4C. The crowding amplitude $a$ varies from 0.15 to 19.0, with a median of 4.3. This median effect corresponds to about fivefold drop of $d_0$ as compared to the uncrowded conditions. There were no significant differences in this parameter across the search tasks.

Figure 5 depicts the fit parameters of the serial model. Overall, this figure resembles the corresponding graphs from Figure 4A and B, but upside down. Thus, both models of limited capacity may produce essentially similar results. It appears that the data in Figure 5A form two clusters along the capacity limit axis. Capacity limit is two to three objects for rotated Ts and orientation–color conjunctions, but predominantly around the maximum set size 8 for simple size or color search. Other search conditions (orientation, size–color conjunctions) are distributed between the two clusters.

None of the models could well approximate the color search data. These data exhibit some nonmonotonic effects of interobject distance that simple crowding models cannot capture.

**Discussion**

The data from the present study confirm that visual search is limited by both local interference and global capacity. The results are broadly consistent with multiplicative effects of local interactions and global attentional processing.

The limitations are different for different visual features and feature combinations. This confirms some results from earlier studies (Verghese & Nakayama, 1994; Hannus, van den Berg, Bekkering, Roerdink, & Cornelissen, 2006). For spatial configuration stimuli, the results were consistent with a strictly limited processing capacity and standard (0.5E) crowding...
zones. For simple features and their conjunctions, the results were different, depending on the features involved. While color search exhibits virtually no capacity limits or crowding, search for an orientation target was limited by both.

Results for conjunctions of features can be at least partly explained by the results from the simple search of respective features. For example, search for size is more efficient than search for orientation and, similarly, search for conjunction of size and color is more efficient than search for conjunction of orientation and color. Still, in both cases, set-size effects were larger for the conjunctions. These results are broadly consistent with what was found by Treisman and Sato (1990).

More reliable differences along the capacity-limitation (set-size) axis indicate that it is unlikely that search performance can be explained by local crowding-like interactions only. Distribution of covert attention affects visual processing in a fundamental way that is different for different stimuli.

Of course, all of that processing must occur in the form of interactions in a neural network and earlier metaphors like “binding” or “gluing” may be misleading. It is possible that global “capacity limitation” is a kind of “crowding” at the highest level of visual network where receptive fields cover the whole visual field. Differently from “normal” crowding, it can be modified by attentional gating at lower levels of processing, exactly like normal crowding is modified by even more lower gating effects of bottom-up saliency (e.g., Pöder, 2006).

The tendency for larger set-size effects for conjunction compared to feature search appears to contradict Eckstein (1998) and Eckstein, Thomas, Palmer, and Shimozaki (2000) findings. At present, I have no good explanation for the difference in the results. Cueing of relevant set size was different in these studies, smaller set sizes (starting at 2) were included in my experiments, and the targets might be more salient compared to distractors in Eckstein’s (1998) experiments (high vs. low contrast, tilted vs. vertical orientation). There are some earlier studies (McLean, 1999) that have found somewhat larger set-size effects for conjunction search. The interpretation is complicated by relatively large differences across visual features and across observers.

The two clusters of capacity limits revealed by serial model resemble the old story of parallel–serial dichotomy. However, there are some discrepancies from this view as well. According to the present results, serial search may be required for orientation–color but not for size–color conjunctions, and aside from the task and stimuli, individual differences and/or strategies of observers may play a role.

Crowding in the present model has two components: a compulsory pooling of flanking objects in a decision model, and a drop of signal-to-noise ratio for individual representations of the objects. Although several datasets can be well approximated with only one of these components, their combination provides a slightly better fit overall.

It is possible that some aspects of the results were caused by the method of cueing the set of relevant objects. According to some studies (e.g., Gobell, Tseng, & Sperling, 2004), a group-wise selection should be more efficient than selection of separate relevant objects among irrelevant ones. Other studies have found that spatial attention has a suppressive surround that may impede attending of a group of adjacent objects (Bahcall & Kowler, 1999; Cutzu & Tsotsos, 2003). Regardless of the particular method, a decreased efficiency of selection with increasing set size is expected, and that may explain “too large” set-size effects found in this study. However, it seems unlikely that the spatial cueing could have very different effects on the search for different targets.

Conclusions

A complete model of visual search must include both local interference and global capacity limitations. Both local and global limitations are different for different visual features and feature combinations.

**Keywords:** visual search, crowding, capacity limitations, features, conjunctions, modeling

Acknowledgments

This study was supported by Estonian Research Council Grant PUT663. I thank Ronald van den Berg for his useful comments on the modeling.

Commercial relationships: none.
Corresponding author: Endel Pöder.
Email: endel.poder@ut.ee.
Address: Institute of Psychology, University of Tartu, Estonia.

Footnotes

1Sample-size model, a classic of SDT-based fixed capacity, predicts a drop of \( d' \) according to the inverse square root of the set size. Thus, the predicted set-size effect in terms of log-log slope is roughly two times stronger for a serial model as compared to a sample-size model, at least within some linear range of \( d' \).

2A similar model was tested by Eckstein (1998) under the name of hybrid model.
References


psychophysics of visual search. *Vision Research* 40, 1227–1268.


