

An escape from crowding

Jeremy Freeman

Swarthmore College, Swarthmore, PA, USA



Denis G. Pelli

Psychology and Neural Science, New York University,
New York, NY, USA



Crowding occurs when nearby flankers jumble the appearance of a target object, making it hard to identify. Crowding is feature integration over an inappropriately large region. What determines the size of that region? According to bottom-up proposals, the size is that of an anatomically determined isolation field. According to top-down proposals, the size is that of the spotlight of attention. Intriligator and Cavanagh (2001) proposed the latter, but we show that their conclusion rests on an implausible assumption. Here we investigate the role of attention in crowding using the change blindness paradigm. We measure capacity for widely and narrowly spaced letters during a change detection task, both with and without an interstimulus cue. We find that standard crowding manipulations—reducing spacing and adding flankers—severely impair uncued change detection but have no effect on cued change detection. Because crowded letters look less familiar, we must use longer internal descriptions (less compact representations) to remember them. Thus, fewer fit into working memory. The memory limit does not apply to the cued condition because the observer need remember only the cued letter. Cued performance escapes the effects of crowding, as predicted by a top-down account. However, our most parsimonious account of the results is bottom-up: Cued change detection is so easy that the observer can tolerate feature degradation and letter distortion, making the observer immune to crowding. The change detection task enhances the classic partial report paradigm by making the test easier (same/different instead of identifying one of many possible targets), which increases its sensitivity, so it can reveal degraded memory traces.

Keywords: crowding, attention, change blindness, change detection, feature integration, object recognition, segmentation, visual memory, partial report

Citation: Freeman, J., & Pelli, D. G. (2007). An escape from crowding. *Journal of Vision*, 7(2):22, 1–14, <http://journalofvision.org/7/2/22/>, doi:10.1167/7.2.22.

Introduction

How do we recognize objects? The answer to this question has eluded vision science. We know that the first step is elementary feature detection, and its mechanisms are relatively well understood (Neisser, 1967; Pelli, Burns, Farell, & Moore-Page, 2006; Robson & Graham, 1981; Treisman & Gelade, 1980). Far more puzzling, however, are the processes by which features are integrated to recognize objects. A phenomenon known as “crowding” may provide a window into the feature integration process (Pelli, Palomares, & Majaj, 2004). Crowding occurs when nearby (nonoverlapping) flankers hinder the identification, but not the detection, of a target object, as demonstrated in Figure 1. Crowding can be relieved by increasing the center-to-center spacing between the target and the flankers. Bouma (1970) showed that the spacing required to eliminate crowding, called the *critical spacing*, is proportional to eccentricity (distance from fixation). Crowding is caused by feature integration over an inappropriately large region, but what determines the size of that region? Is attention involved?

Why should we care whether attention is involved? After all, one can add an attentional manipulation to practically any task. Does it matter whether attention is an

explicit part of the model for crowding? Yes, but the reason depends on how you think it will turn out. At one extreme, some have claimed that crowding is a bottom-up phenomenon that may be described mechanistically without invoking attention, except, perhaps, as an overall modulation of performance, independent of the degree of crowding (Banks, Larson, & Prinzmetal, 1979; Pelli et al., 2004; Scolar, Kohlen, Barton, & Awh, 2007). Their aim is parsimony. None of us expect to understand attention anytime soon, so showing that it can be left out greatly increases the prospect of understanding crowding in a reasonable time. At the other extreme, many are drawn by William James’s suggestion that attention selects for awareness, and some have argued that the critical spacing of crowding reflects the size of the spotlight of attention (He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001). This too is parsimonious, merging the two problems of attention and crowding into one. Here we present a new experiment and critique an old one, “The spatial resolution of visual attention,” to determine whether attention is essential to an account of crowding.

Attentional accounts are top-down. William James (1890), considering the “millions of items ... present to my senses which never properly enter into my experience,” concluded that “my experience is what I agree to attend to.” A century later, Treisman updated James’s



Figure 1. Crowding. Fixate the large +. It is easy to identify the r on the left but impossible to identify the r on the right. The flankers, t and y, spoil recognition of the target. If you fixate the small +, the r on the right becomes easy to identify. Reducing the eccentricity of the target reduces your critical spacing, which relieves crowding.

account with her Feature Integration Theory, according to which “features are registered early [and] automatically ... while objects are identified separately and only at a later stage, which requires focused attention” (Treisman & Gelade, 1980). Ned Block (2001) argues that although we might “have phenomenal awareness” of the features of crowded letters, we still lack “the attentional resources to apply [letter] concepts to them.” However, Holcombe and Cavanagh (2001) find that conjunctions of orientation and luminance can be reported for rates of presentation as high as one every 14 ms. Given that attentional mechanisms require at least 80 ms to deploy, these results (among others) provide evidence for feature integration without focused attention (Blaser, Papathomas, & Vidnyánsky, 2005; Driver, Davis, Russell, Turatto, & Freeman, 2001; Lamme, 2004).

According to bottom-up proposals, the visual system has access to variously sized *isolation fields*—elliptical regions defined by the critical spacing within which features are necessarily integrated and outside which features are excluded—but the minimum isolation field size increases with eccentricity and is “independent of everything else, including attention” (Bouma, 1970; Pelli et al., 2004, p. 1160). As a result of this fixed anatomical limitation (dependent on retinal locus and little else), if a peripherally viewed target is too closely flanked, the features of both the target and the flankers are combined in a single isolation field. Integration fails to yield recognition, leaving the observer with a jumble of features.

Top-down proposals similarly assert that crowding is caused by feature integration over an inappropriately large region. However, in these accounts the size of the region is not anatomically fixed and instead varies under attentional control. Arguing for such an explanation, Intriligator and Cavanagh (2001) measured critical spacing for two attentive tasks—tracking and stepping dots—and measured the critical spacing to estimate the attentional selection region within which objects are individuated. As in the classical study of Bouma (1970), they found that the size of the selection region increases with eccentricity with a proportionality of roughly 0.5, but they argued from this similarity that crowding effects measure the spatial resolution of attention. When crowded, the attentional spotlight cannot isolate the correct set of features for integration (Cavanagh, 2004).

Experiments have failed to distinguish between the two accounts. We defer until the [Discussion](#) section our refutation of the attempt by Intriligator and Cavanagh (2001) to prove the attentional account. If crowding effects are due to overly coarse attentional selection, then precueing a target location in a traditional crowding paradigm might focus attention and relieve crowding. Two studies (Nazir, 1992; Wilkinson, Wilson, & Ellemberg, 1997) found no effect of precueing on crowding, but they failed to find any effect of their precue under any condition, leaving open the possibility that attention was not manipulated at all. Several studies do find that precueing improves performance somewhat, although crowding is not eliminated (Huckauf & Heller, 2002; Morgan, Ward, & Castet, 1998; Strasburger, 2005; Van der Lubbe & Keuss, 2001). Scolari et al. (2007) find that precueing improves the identification of oriented T’s surrounded by flankers but does not reduce the critical spacing.

How else might we distinguish between the two accounts? In the change blindness paradigm, an observer sees an image and then, after a brief pause, sees an altered image. Observers fail to notice even dramatic changes (Simons & Ambinder, 2005). Some have argued from change blindness that awareness is limited by attention (Noë, Pessoa, & Thompson, 2000). However, Landman, Spekreijse, and Lamme (2003) found that an interstimulus cue, presented as much as 1500 ms after the first stimulus, improved change detection. From their partial report data, Landman et al. argued for the existence of a nondurable but high-capacity awareness similar to iconic memory (Sperling, 1960). The interstimulus cue enabled the retrieval of cued items from this awareness, moving them into a more durable form and making them available for comparison. Becker, Pashler, and Anstis (2000) obtained a similar result for change detection with letters.

Here we examine change detection for widely and narrowly spaced letters, using a cue to manipulate attention. Change detection performance without a cue will lack focused attention and should therefore suffer crowding effects according to both bottom-up and top-down accounts of crowding. However, the two accounts make different predictions for the effectiveness of an interstimulus cue. If crowding effects reflect a fixed, preattentive limit, we might not expect precisely focusing attention to help retrieve crowded letters. However, if crowding effects reflect a purely attentional limit, then an interstimulus cue might enable the attentional retrieval of otherwise crowded letters.

Methods

Observers

Two observers (age 20) participated. Both had normal or corrected-to-normal vision. Observer J.F. is an author

and was experienced in the task. Observer S.B. was unaware of the purpose of the experiment.

Stimuli

Stimuli were created by a Macintosh PowerPC computer running MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A video attenuator was used to drive only the green gun of a 19-in. AppleVision monitor with a frame rate of 75 Hz and a resolution of 832×624 pixels, 57 pixels/in. (Pelli & Zhang, 1991). Letters were presented as black text on a background of 35 cd/m^2 .

The stimulus displays consisted of eight target letters and two flanker letters presented around a black fixation point subtending 0.2° . The task was to monitor and to report whether one of those eight letters changed from Frame 1 to Frame 2. Viewing distance was 100 cm. For all but one experiment (extra-narrow spacing), letters were 0.9° high capital letters displayed in the Linotype font HelveticaNeue LT 85 Heavy (x-height 0.6°). For the extra-narrow spacing experiment (all four conditions in Figure 3b), the letters were 0.6° high (x-height 0.4°). For every trial, eight different target letters were randomly selected from among the 19 English letters **A B C D E F G H K N O P Q R S U V Y Z**. The flanker letters were both **X**. [Most of the modern European alphabets are based on the Latin alphabet, but the small differences among them (e.g., Italian does not include J, K, W, X, and Y) are relevant to our observers' familiarity with their own alphabet, so we will refer to our letters as English.] The letters were spaced equally in a circular arrangement centered on fixation with a radius of 6° . To rule out acuity problems, we confirmed that our observers could identify single letters displayed at this size and eccentricity. Note the excellent identification performance with wide spacing in Figure 7a.

In cued trials, a cue was presented for 100 ms during the interstimulus interval to indicate the location of a potential change. The cue was a black dot (0.25° diameter) presented 6.75° from fixation. The cue was 0.75° from the target, center to center. There was no cue–target overlap.

Change detection task

The change detection task is depicted in Figure 2. Each image persists until it is replaced by the next. In each trial, a display of letters (Frame 1) was presented for 500 ms, followed by an interstimulus interval of 1600 ms, after which a second display of letters (Frame 2) appeared for 500 ms. The fixation point remained on the screen throughout each trial.

In 50% of all trials, one randomly chosen letter from Frame 1 changed to a different letter in Frame 2; in all other trials, Frames 1 and 2 were identical. Observers were told that a change would occur in 50% of trials.

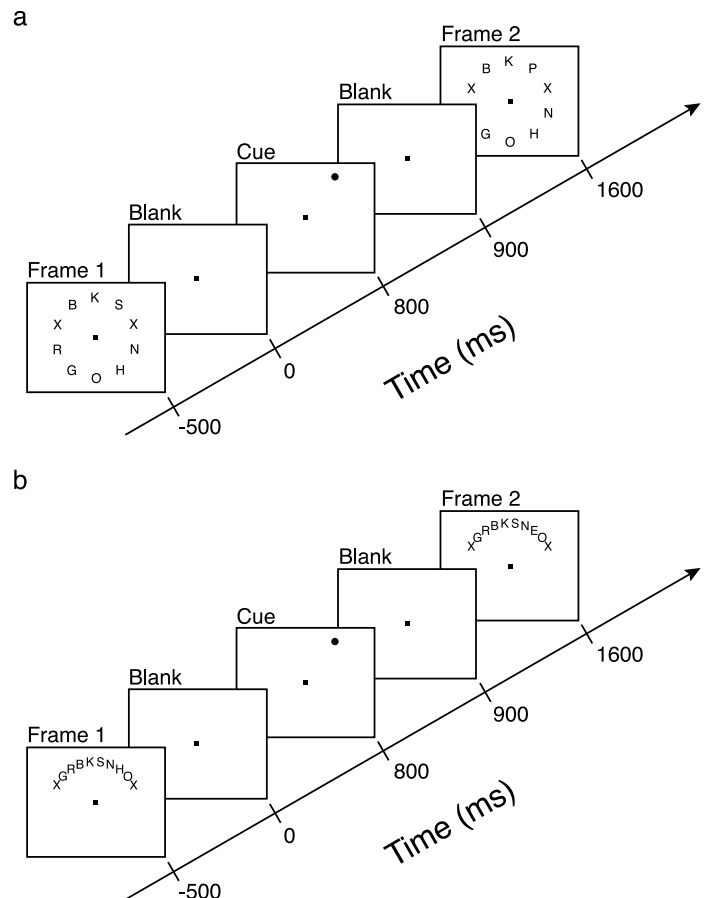


Figure 2. The change detection task. The diagram depicts the onset of each image. Each image persists until the next. (top) Change detection task with wide (3.7°) spacing. (bottom) Change detection task with narrow (1.1°) spacing.

Following each trial, observers were asked to indicate with a key press whether they saw a change; response time was unlimited. After they pressed the key, a high tone indicated a correct response and a low tone indicated an incorrect response. Observers' eyes were monitored to ensure proper fixation, discarding trials where fixation was lost. A blank screen (with fixation) was presented for one second between trials.

With respect to cueing, there were two conditions: cue at 800 ms or no cue. In trials with a change, the cue always appeared at the location of the change. In trials without a change, the cue appeared at a random location from among the locations of the eight target letters. The cue never appeared at a flanker location. For cued trials, observers were informed that the change might or might not occur, but that if it occurred, it would occur at the cued location.

Observers performed blocks of 48 trials. Blocks of different cue, spacing, and flanker conditions were performed separately in random order. The number of change/no-change trials for each condition was balanced, and the order was randomized during the 48 trials.

Observers completed at least 48 practice trials in each condition, the results of which are not reported.

Letter identification task

In each trial of the letter identification task, a “frame” of letters was presented for 500 ms. The fixation point remained on the screen throughout each trial. There were two possible cue conditions: cue onset 800 ms before the frame onset or cue onset 800 ms after the frame offset. In both cases, the cue was presented for 100 ms. Following each trial, observers were asked to identify the cued letter with a key press (1 of 19 possible). Observers performed blocks of 48 trials. Blocks of different cue, spacing, and flanker conditions were performed separately in random order. The same five spacing and flanker conditions (described below) were used in the change detection and letter identification tasks.

Capacity

Following Landman et al. (2003) and Pashler (1988), performance is expressed as capacity, c . Capacity is a parameter in a model that supposes that the observer monitors a number c of items during a single change detection trial. To estimate capacity, we express the hit rate h as

$$h = \frac{c}{n} + \frac{n-c}{n}f, \quad (1)$$

where c is the capacity, n is the number of items (in this experiment, eight letters), and f is the false alarm rate, the fraction of no-change trials on which the observer incorrectly reports a change. The hit rate is the observer’s probability of reporting a change when 1 of n letters changes. Our model for hit rate assumes that, because of a capacity limitation, the observer monitors a fixed number c of letters. If one of the c monitored letters changes, then the observer reports the change. If one of the $n - c$ unmonitored letters changes, then the observer reports a change with a probability given by the false alarm rate measured on the no-change trials. Solving Equation 1 for c gives

$$c = \frac{h-f}{1-f}n. \quad (2)$$

Using Equation 2, a capacity estimate for each condition was calculated from 48 trials (24 change trials and 24 no-change trials). 1 to 4 capacity measurements were taken for each observer in each condition. Paired t tests were used to compare capacities between conditions across multiple observers and unpaired t tests were used to compare capacities between conditions within single observers.

Crowding

Crowding occurs when nearby letters (flankers) jumble the appearance of a target letter, making it hard to identify (Bouma, 1970). Critical spacing is the target-to-flanker spacing required to relieve crowding. In the radial direction, critical spacing is given by

$$s'_r = b\varphi_r, \quad (3)$$

where φ_r is the radial eccentricity of the target letter (6° in this experiment) and b is Bouma’s (1970) proportionality constant (roughly 0.5). The shape of the integration region defined by critical spacing, however, is elliptical with an ellipticity of roughly two (Toet & Levi, 1992). Thus, critical spacing in the circumferential direction is approximately half of critical spacing in the radial direction and is given by

$$s'_c = \frac{1}{2}b\varphi_r. \quad (4)$$

If the ratio of actual spacing s to critical spacing s' (either radial or circumferential, depending on flanker location) is below one, then the letters should be crowded by the flankers; and if it is well above one, then the letters should escape crowding.

We first attempted to crowd target letters by decreasing center-to-center letter spacing. In an “uncrowded” condition, we used wide (3.7°) letter spacing and the ratio s/s' was 2.5 (using Equation 4). In one “crowded” condition, we used narrow (1.1°) letter spacing and the ratio s/s' was 0.7. In a second “crowded” condition, we used extra-narrow (0.63°) letter spacing, giving a ratio s/s' of 0.4. For this extra-narrow spacing experiment (all four conditions in Figure 3b), letter size was reduced to 0.6° high capital letters (x-height 0.4°).

We also attempted to crowd target letters by adding X’s inside and outside the original 10 letters. The center-to-center spacing between target letters and flankers was 1.1° , giving a ratio s/s' of 0.35 (using Equation 3). These flankers were added to the wide and narrow letter spacing conditions. For the latter, we estimate the ratio s/s' to be roughly $0.25 = 0.7 \times 0.35$, assuming a compounded effect for narrowing spacing ($s/s' = 0.7$) and adding flankers ($s/s' = 0.35$).

Foreign alphabet

To assess whether both the cued and the uncued versions of the change detection task require letter identification, we had observers perform two versions of the change detection task (3.7° and 1.1° letter spacing) with Armenian letters. These were nineteen 0.9° high capital letters displayed in the TrueType font Nork: Հ Ճ Ժ Ի Դ Վ Ը Պ Թ Ռ Շ Ղ Ջ Խ Մ Ա Չ Փ Ն

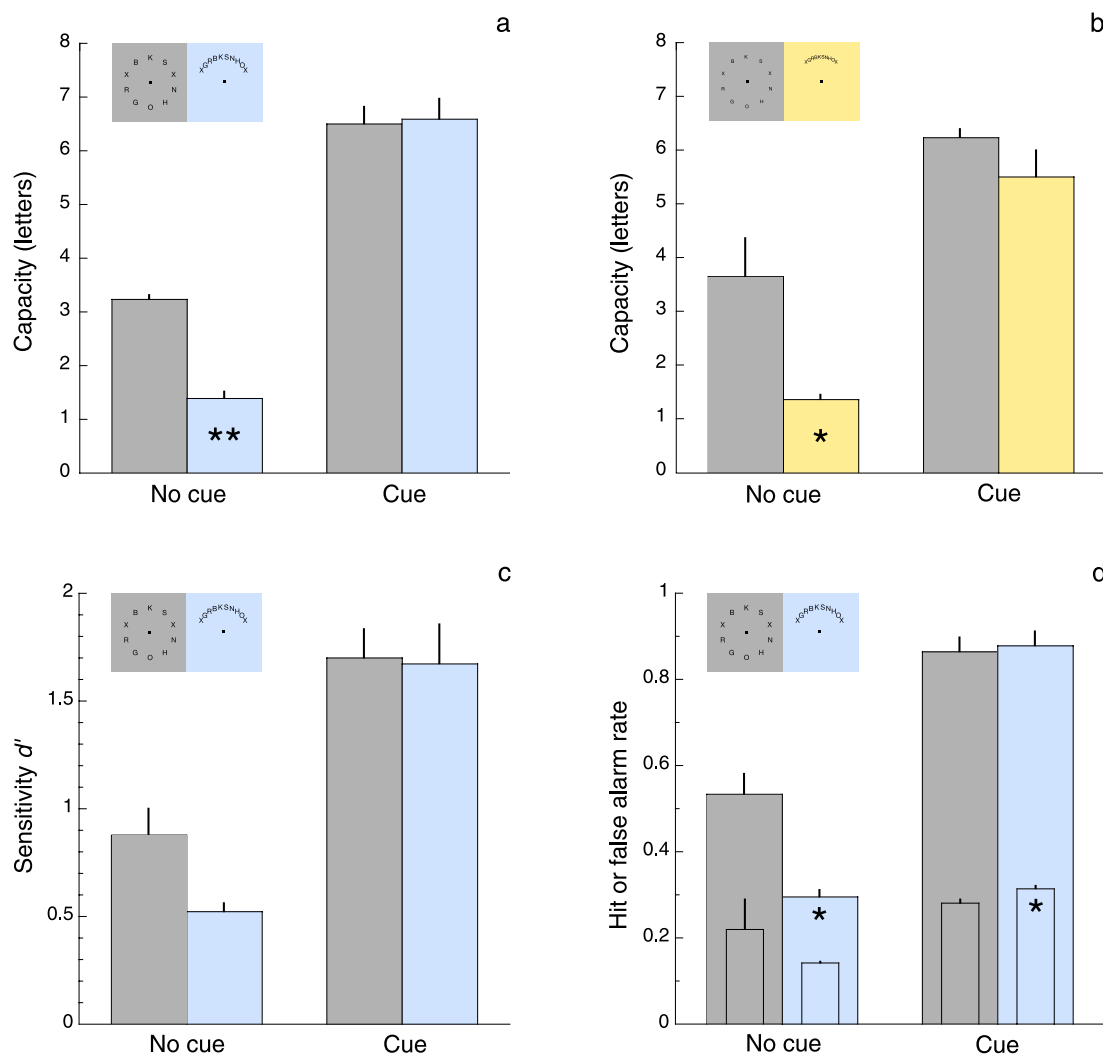


Figure 3. Reducing the spacing. (a) Capacity (the estimated number of letters monitored by the observer) during change detection for letters with wide (3.7°) spacing (gray) and narrow (1.1°) spacing (blue), with and without a cue. Average results for two observers (J.F. and S.B.); error bars show standard error between observers. (b) Capacity during change detection for letters with wide (3.7°) spacing (gray) and extra-narrow (0.63°) spacing (yellow), with and without a cue. Results for observer J.F.; error bars show standard error within observer. (c) Sensitivity d' during change detection for letters with wide spacing and narrow spacing, with and without a cue. Average results for two observers (J.F. and S.B.); error bars show standard error between observers. $d' = z(h) - z(f)$, where h is the hit rate, f is the false alarm rate, and z is the inverse cumulative normal distribution function. (d) Hit rate (wide bars) and false alarm rate (narrow bars) during change detection for letters with wide and narrow spacing. Average results for two observers (J.F. and S.B.); error bars show standard error between observers. * $p < 0.05$. ** $p < 0.01$.

with \uparrow as the flanker letter. Nork is available from Lines, Fonts, and Circles (220 S. Kenwood, Suite 100, Glendale, CA, 91205).

Results

Change detection

Four change detection experiments are presented in four graphs (Figures 3a, 3b, 5a, and 5b), all of which turn out to be very similar. (Figure 6, Armenian, is discussed

separately.) Without a cue, there is a large effect of spacing: The second bar is a small fraction of the first. With a cue, performance is much better and there is no effect of spacing: The third and the fourth bars are tall and practically equal. Let us now examine this finding in more detail.

Reducing the spacing

We first attempted to crowd target letters by reducing the spacing (see Figure 4 for a demonstration). We had observers perform a change detection task with or without

sensitivity by a factor of 1.7 (with marginal significance; $p = 0.13$, one tailed) but did not significantly affect cued sensitivity ($p = 0.95$). In [Figure 3d](#), narrowing spacing significantly reduced uncued hit rate ($p < 0.05$, one tailed) but did not significantly affect cued hit rate ($p = 0.87$; [Figure 3d](#)). Because the alternate ways of plotting the results ([Figures 3c](#) and [3d](#)) provide no new insight, we plot only capacity for the remaining conditions.

With 1.1° letter spacing ([Figure 3a](#)), the ratio of actual to critical spacing is approximately 0.7. In an attempt to crowd letters more, we reduced that ratio to 0.4 by using extra-narrow (0.63°) letter spacing ([Figure 3b](#)). Reducing spacing from wide to extra-narrow reduced uncued capacity by a factor of three ($p < 0.05$, one tailed) but had only a weak insignificant effect on cued capacity ($p = 0.28$).

Adding flankers

We also attempted to crowd target letters by adding flanker X's inside and outside. The effect of adding flankers ([Figure 5a](#)) was very similar to that of narrowing spacing ([Figure 3a](#)). Adding flankers reduced uncued capacity by a factor of three ($p < 0.05$, one tailed) but did not significantly affect cued capacity ($p = 0.26$).

Our last attempt to maximize crowding combined both manipulations. Adding flankers and narrowing spacing together reduced uncued capacity dramatically ([Figure 5b](#)), by a factor of 3.5 ($p < 0.05$), but had no significant effect on cued capacity ($p = 0.35$). Thus, even in the most unfavorable “crowding” conditions, the cue greatly increased capacity and eliminated the effects of spacing.

Foreign alphabet

Does change detection require letter identification? Observer J.F., who is unfamiliar with the Armenian alphabet, did the cued and uncued change detection tasks with Armenian letters, under both wide and narrow spacing conditions ([Figure 6](#)). The results for English letters from [Figure 3a](#) appear as dashed lines in [Figure 6](#) for comparison, showing that changing to an unfamiliar alphabet greatly reduced capacity. The observer could retain only two Armenian letters. Using a partial report paradigm with (lowercase) English and Armenian letters, Pelli et al. (2006) showed that observers can retain five characters in their native alphabet, but only two characters in a foreign alphabet. Thus, the capacities greater than two for English letters may require familiarity. The only caveat is that Pelli et al. demanded accurate identification, whereas here we only demand detection of change. We return to this in the [Discussion](#) section.

Incidentally, note that we used a bold version of Helvetica (the English font) and a plain version of Nork (the Armenian font). We wondered how much of the effect of switching font ([Figure 6](#)) might be attributed simply to the reduction of weight (boldness). The strokes of Nork are about half the thickness of the strokes in our bold Helvetica. We did an extra test, measuring cued capacity with narrow spacing for HelveticaNeue LT 25 Ultralight, for comparison with our standard HelveticaNeue LT 85 Heavy. Switching from **Heavy** to Ultralight reduces the stroke thickness by a factor of five, yet this switch reduced letter capacity hardly at all, about 10%. Thus, boldness accounts for very little, if any, of the effect of changing alphabets in [Figure 6](#).

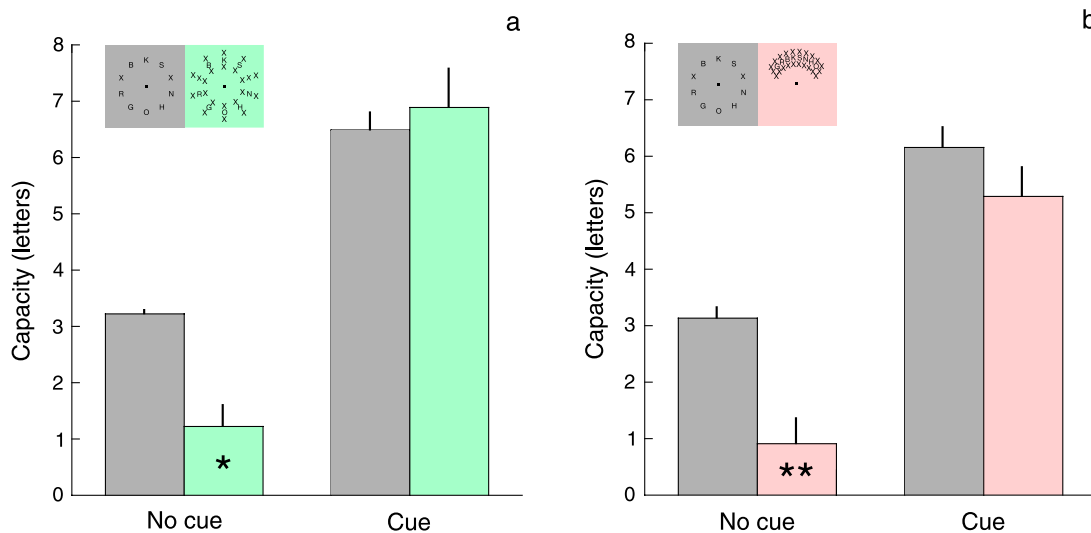


Figure 5. Adding flankers. (a) Capacity during change detection for letters with wide (3.7°) spacing (copied from [Figure 3a](#), gray) and letters with wide spacing and flanker X's inside and outside target letters (green), with and without a cue. Center-to-center distance between target letters and flanker X's was 1.1° . Average results for two observers (J.F. and S.B.); error bars show standard error between observers. (b) Capacity during change detection for letters with wide (3.7°) spacing (gray) and letters with narrow (1.1°) spacing and flankers (pink), with and without a cue. Results for observer J.F.; error bars show standard error within observer.

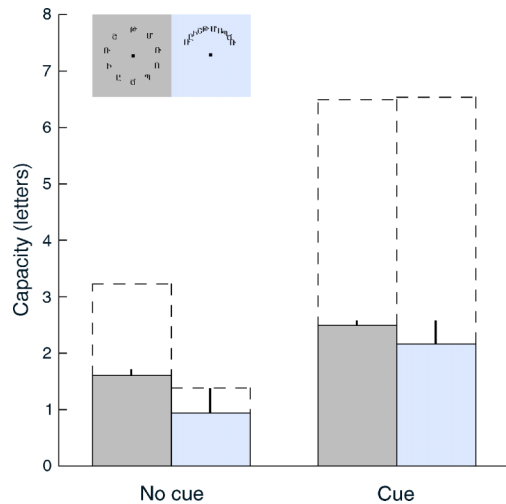


Figure 6. Foreign alphabet. Capacity during change detection for Armenian letters with wide (3.7°) spacing (gray) and narrow (1.1°) spacing (blue), with and without a cue. Results for observer J.F.; error bars show standard error within observer. Results for English letters from Figure 3a are shown as dashed lines for reference.

Letter identification

Baseline crowding measurements from a traditional letter identification task are shown in Figure 7. With a precue, all four spacing manipulations reduced the

proportion correct by a factor of 1.5 or more. With a postcue, 40% of widely spaced letters are identified correctly. All four spacing manipulations reduced performance by a factor of 1.5 or more. Thus, no matter when we cue, all our crowding manipulations impair letter identification.

Discussion

We are interested primarily in the effect of cueing, our attentional manipulation. However, before turning to that, let us consider the uncued results, which are compactly presented in the left half of Figure 6 (“No cue”). The measured capacity, in letters, is 3.2 for widely spaced English letters. Switching to Armenian or reducing the spacing (to narrow) roughly halves the capacity. Doing both reduces capacity still further. These manipulations make the letters less familiar and thus more difficult to describe. We suppose that limited capacity working memory can retain fewer letters when each letter requires a longer description (less compact representation). The finding that familiar faces are processed faster has been attributed to more compact representations of familiar faces (Jackson & Raymond, 2006; Tong & Nakayama, 1999). All our uncued results seem consistent with each letter’s feature integration yielding a letter description, longer for less familiar letters, with performance reflecting

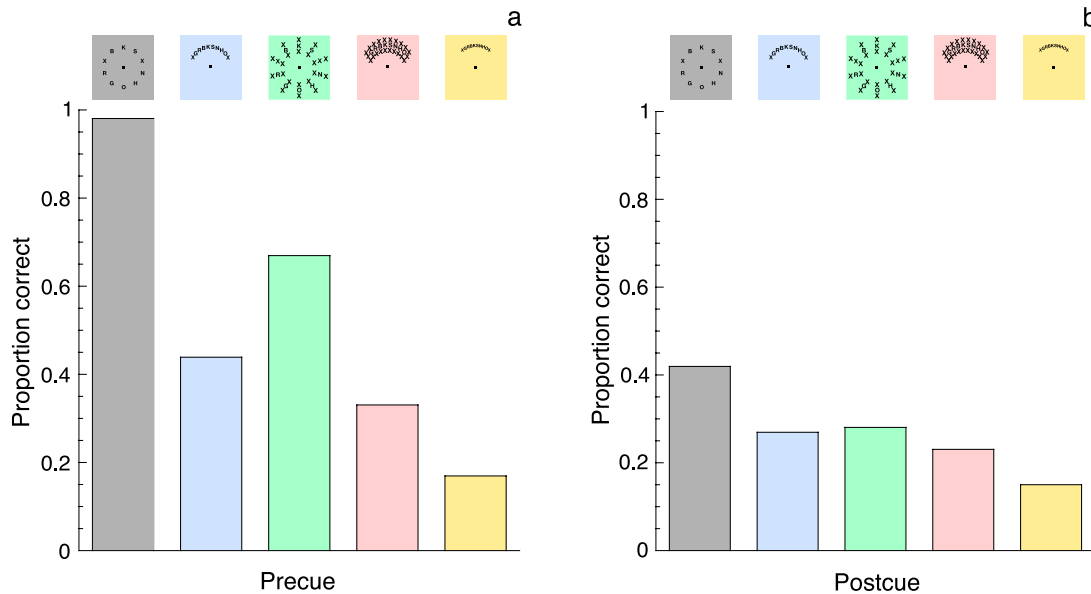


Figure 7. Letter identification. Proportion correct letter identification for cued letters with wide (3.7°) spacing (gray), narrow (1.1°) spacing (blue), wide spacing and flanker X’s inside and outside target letters (green), narrow (1.1°) spacing and flankers (pink), and extra-narrow (0.63°) spacing (yellow). Results for observer J.F. (a) One of eight possible target locations (randomly selected) was cued 800 ms before the frame onset. (b) One of eight possible target locations (randomly selected) was cued 800 ms after the frame offset. The guessing rate for this task, identifying 1 of 19 possible letters, is 0.05. As always, cue duration is 100 ms.

how many working memory can hold. Cueing eliminates this memory limit. Once the cue arrives, the observer needs to remember only the cued letter. The length of its description is now irrelevant because it fits easily within the capacity of working memory. Now we return to the main question.

Can attention relieve crowding? Others have addressed this question by adding an attentional manipulation to a crowding task (Scolari et al., 2007). They find no effect of attention on critical spacing, which favors a bottom-up account. Here, we add a crowding manipulation to an attentional task and find no effect of crowding on cued performance, which seems to favor a top-down account for this particular case. Figure 8 summarizes the effects of our crowding (i.e., spacing) manipulations by plotting the ratio of capacities with and without reduced spacing as a function of the ratio of actual to critical spacing. The dashed horizontal line at 1 represents no effect of spacing. Anything below that line indicates an effect of spacing. With no cue, the ratio starts low, at 0.3, and rises to 1, telling a standard crowding story: Nearby letters hinder performance; increasing the spacing relieves the effect. However, adding the cue eliminates the spacing effect. The ratio hovers near 1, independent of spacing. Every spacing and flanker manipulation reduced uncued capacity

by more than a factor of two, but even the tightest spacing failed to affect cued capacity.

This paper began with two accounts of crowding: one bottom-up and preattentive, the other top-down and attentional. Our attentional manipulation eliminates a large spacing effect (Figure 8). The observer is reliably detecting changes in letters that he cannot identify accurately. This is an escape from crowding, although the crowding itself is unrelieved.

This result is one of the few examples of an attentional manipulation eliminating a crowding effect. (The only other instance we know of is the demonstration by Cavanagh & Holcombe, 2007, that induced motion can eliminate the effects of crowding.) Do these results rule out all bottom-up, preattentive accounts of crowding? No. Consider the possibility that segmentation (isolation) is preattentive and coarse and that feature integration occurs late. By this account, our attentional manipulation (an interstimulus cue) allows our observers to do a second, finer segmentation that results in a more appropriate final integration. Thus, in the presence of the fine segmentation enabled by the interstimulus cue, the early coarse segmentation becomes irrelevant.

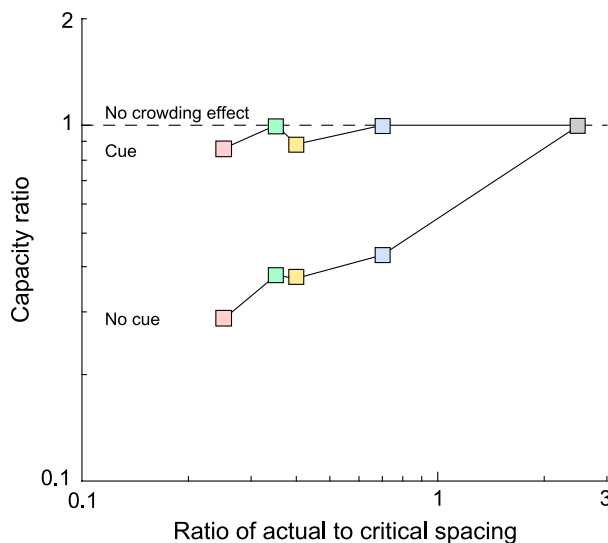


Figure 8. Effect of spacing on capacity. Effect is shown as the ratio of capacity with tight spacing s to capacity with wide spacing as a function of the ratio of actual to critical spacing s/s' for each of our four crowding manipulations and the reference wide-spacing condition (see Methods section): narrow spacing with added flankers ($s/s' = 0.25$, pink, Figure 5b), added flankers ($s/s' = 0.35$, green, Figure 5a), extra-narrow spacing ($s/s' = 0.4$, yellow, Figure 3b), narrow spacing ($s/s' = 0.7$, blue, Figure 3a), and wide spacing ($s/s' = 1$, grey). The dashed horizontal line at 1 represents no crowding effect. Anything below that line indicates a crowding effect.

Top down

Intriligator and Cavanagh (2001) claim to have established a role for attention in crowding, suggesting that critical spacing is the resolution of the “basic ... access of attention to spatial locations.” We accept most of their argument, but we reject their conclusion that the limit is attentional because we cannot accept the extraordinary assumption upon which it rests.

Crowding has mostly been studied with identification tasks because detection is usually unaffected. Intriligator and Cavanagh (2001) applied the crowding paradigm to a new kind of task, individuation: tracking and stepping dots. They found the same critical spacing for individuation as previously found for identification. We essentially agree with their suggestion that the crowding of identification is due to the combining of features in the process of feature integration:

If items are spaced too closely, selection may occasionally pick up an adjacent item in addition to the target. In this case, features may be mixed and degraded *following* selection. If this is the only source of error in the flanker and crowding tasks, critical spacings from these tasks ought to match the critical spacings from our tracking and stepping tasks. To a first approximation, this is what we find

(Intriligator & Cavanagh, 2001, p. 208). Where they say “selection,” we would say “integration.” “Selection” implies both integration and attention, but attention is irrelevant here because the prediction depends solely on

the feature integration. Let us remain agnostic about the role of attention until the data really bear on the issue. To estimate an object property such as shape, the observer must integrate the elementary features. An unavoidable cost of the integration is the mixing in of inappropriate features if the region of integration extends too far and includes neighboring objects.

We find the feature integration account, above, equally applicable to identification and individuation, especially because the two tasks yield the same critical spacing. Any feature integration model of crowding (including the spotlight of attention or the area of selection) will predict that object properties, including shape and position, are disturbed by incorporation of inappropriate neighboring elementary features. That area might be, as they suppose, the area of the attentional spotlight, but this is irrelevant to explaining the results, which depend only on the area's extent. To us, the difference between tasks is unimportant. The same region of integration predicts the same critical spacing.

They, however, assert that the same results from the two tasks require different explanations. They create this complication by assuming that position is privileged. Like us, they seem to believe that the first stage of vision is feature detection (or encoding) and that performance of most tasks, including identification, is feature-based and subject to feature degradation. Unlike us, they seem to believe that judgment of position (location) is not feature-based and is immune to feature degradation.¹

...distortion and mixing of adjacent features may degrade the representation.... This source may contribute to the critical spacing measured in crowding and flanker studies, but it cannot contribute to the critical spacing in our tracking and stepping studies where target features and identity are irrelevant

(Intriligator & Cavanagh, 2001, p. 208). They claim that because individuation is position-based, performance is

¹Intriligator and Cavanagh say that the degradation could be “feature distortion” (mentioned only in passing) or “feature mixing.” “Feature mixing” seems to be the intact migration of nameable high-level object properties (e.g., letter shape) from object to object, reported by Treisman and Schmidt (1982), which is characteristic of “temporal crowding,” not the inappropriate inclusion of neighboring elementary features, which is characteristic of spatial crowding (Pelli et al., 2004, p. 1157; Wolford & Shum, 1980, p. 416). We infer this meaning of “feature” from their assertion that “Our tasks track location in dense arrays of identical items and so are unaffected by preattentive feature interactions that might degrade the recognition measures used in crowding and flanker tasks.” (Intriligator & Cavanagh, 2001, p. 203). Granting their assumption that position is privileged, that statement is true for high-level features (so mixing would preserve object identity) but false for elementary features (whose mixing would change object identity).

limited solely by attentional “access to the location,” so the critical spacing of individuation must be a strictly attentional limit. This is the complication provoked by the assumption that position is privileged. Critical spacing of identification is attributed to feature mixing over the area of selection; critical spacing of individuation, numerically equal, is attributed to limited attentional access to location. This is an extraordinary theory of vision, in which identification and judgment of most object properties are based on the detected features, but individuation and judgment of position are based on some other more reliable source of information and are thus immune to feature degradation.

It is tempting to suppose, as they do, that position is the firmament upon which visual experience is built. But this is an assumption, not a fact. Sagi and Julesz (1985) reported that observers can only say what an object is if they can also say where it is. However, Farell and Pelli (1993), with different stimulus conditions, found the opposite. Observers identified a digit among letters more accurately than they could locate it.

Contrary to Intriligator and Cavanagh's presumption that position is privileged, the facts they present about individuation in particular and position judgment in general all seem consistent with individuation being feature-based, just like identification. They mention “spatial uncertainty” and note that positional accuracy at large separations is proportional to eccentricity. They aptly quote the observation of Landolt (1891, p. 385) that dots become uncountable when too closely spaced, “though still perfectly and distinctly visible.” Noting the crisp distinctness of the dots, one might suppose that one's low-level perception of dot position is correspondingly precise and that the dismaying inability of observers to individuate a dot (to select, track, or count it) is a strictly attentional limitation. After all, these tasks use position coarsely; the observer need only distinguish the currently tracked thing from its neighbors. But appearances can be deceiving. The dots seem distinct, and Intriligator and Cavanagh show that their observers can detect the gap between dots. However, Levi, Klein, and Yap (1987) showed that accuracy of position judgment reflects feature integration. Position is not privileged (Figure 9). Both identification and individuation show similar effects of crowding. Neither seems to be a trivial consequence of the other.² Stripped of the claimed immunity from feature degradation, the crowding of individuation, like that of identification, seems to be the cost of the feature integration that estimates the relevant object property. There is nothing in that result that demands a role for attention in crowding.

²If individuation is a prerequisite for identification, then the critical spacing of identification would be a consequence of that dependence. They assert that individuation is a prerequisite for identification, but this is the old 1960s idea that scene segmentation (then called “edge detection”) precedes object recognition, which, year by year, seems less and less likely to be true.

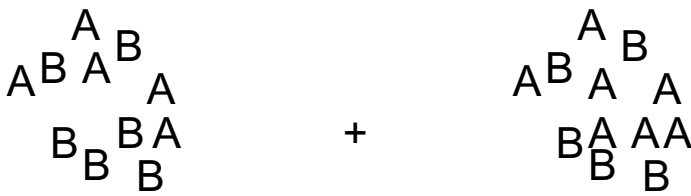


Figure 9. Distinct yet crowded: Position is not privileged. This figure shows that less-than-critically spaced objects may appear distinct, although essential properties—shape and position—are greatly affected by neighboring objects. While fixating the plus, can you see any difference between the left and right letter clusters? In fact the central letters are different in location, identity, and number, but this is undetectable. The lesson is that the properties of these objects—including shape, identity, and position—are all susceptible to crowding. Distinctness is no guide to absence of crowding. These letters seem distinct, yet their apparent shapes and positions are strongly affected by their neighbors.

Our new result provides one instance—interstimulus cueing during change detection—in which cueing allows observers to escape the effects of crowding. As such, our result shows that at least one crowding effect may reflect the want of a finer segmentation triggered by focused attention, as envisioned by Intriligator and Cavanagh (2001).

What is the fine segmentation mechanism that the interstimulus cue enables? We began with the suggestion by Intriligator and Cavanagh (2001) that the spotlight of attention might be focused finely, in this case to isolate a single letter. However, it seems odd that observers benefit from greatly improved resolution only in the particular paradigm used here. Does the supposed high-resolution capability lie latent, unused, the rest of the time? We are reluctant to suppose a powerful mechanism (the attentional spotlight with extra-fine resolution) just to account for the results of our peculiar task.

Bottom up

A key issue is whether the observer's cued change detection is based on feature integration. Considering the English data alone suggests no, but combining our English and Armenian data forces us to say yes. As a whole, our results favor one simple account.

Consider the English data first. When object recognition fails, the observer can still see texture, and texture discrimination may be enough to detect change. At first this seemed an attractive way to account for the absence of any spacing effect in the cued condition. Crowding is feature integration over an inappropriately large area. However, this inappropriate integration is inconsequential if the observer discriminates textures based on raw features (not integrated). By this account, the cue allows

the observer to select and to remember enough target features to detect a change in that letter. Alas, this appealing account cannot cope with the Armenian data. If the change is detected by texture discrimination, without feature integration, then the observer's familiarity with English letters seems irrelevant, and we would not expect much difference in change detection performance for a foreign alphabet, like Armenian letters. In fact, switching from English to Armenian letters halved the cued performance. The great advantage for the familiar alphabet suggests that the observer is integrating features to see the cued letter. The uncued capacity for Armenian letters is one and a half letters, and reducing spacing reduces this by a factor of 1.7 (Figure 6), similar to the factor of 2.3 for English letters (Figure 3a). This effect of spacing indicates that the uncued performance is based on feature integration.

Spurred by the Armenian results, let us now suppose that the observer detects cued change by feature integration. How can there be no effect of spacing for cued change detection, for either English or Armenian letters? Certainly any feature integration must extend to Bouma's critical spacing, and the integrated result must be distorted and jumbled by the inappropriate inclusion of features from the flanking letters. However, the flankers are identical in Frames 1 and 2, so they will have identical effects on identical targets. The observer's task is merely to detect change. The identification experiment (Figure 7) shows that the inclusion of flanker features (at reduced spacing) greatly impairs identification of the target letter. However, the distortion, applied equally to Frames 1 and 2, may have no effect on change detection. The result of the feature integration could take various forms. Perhaps the integrator categorizes the cued, jumbled target as the most similar English letter. The poor identification performance indicates that the letter identification is usually wrong, but it would be the same wrong letter when the target is identical (in the second frame) and presumably usually a different letter when the target is different. Crowded letters are confused with only a few others (Pelli & Tillman, 2007). The observer will correctly detect change as long as different letters look different, even if they look wrong and are misidentified. The observer lacks templates for Armenian letters, so the feature integration cannot yield an Armenian letter, but it might yield the most similar English letter or letters. Using poorly matched templates accounts for the poorer cued performance with Armenian letters.

The cue reduces the memory requirement to just the cued letter, so even a long internal description fits within the capacity of working memory. The cued escape from crowding is enabled by the easy test (same/different) of the change detection paradigm. The same/different test can be applied to distorted letters and is thus immune to crowding, instead of demanding accurate letter identification (1 of 19), which cannot tolerate distortion and is thus susceptible to crowding.

Conclusion

One account of crowding is bottom-up and preattentive. Another is top-down and attentional. Intriligator and Cavanagh (2001) proposed the attentional account, but we show that their conclusion rests on an implausible assumption. Consistent with both accounts, we find that standard crowding manipulations—reducing spacing and adding flankers—impair uncued change detection performance. However, the same crowding manipulations fail to impair cued change detection. The cue eliminates the effects of crowding. This result—cueing to escape crowding—has been sought after but was hard to find. We cannot reject either account, but parsimony favors a bottom-up account. Crowded letters look less familiar, so we must use longer internal descriptions (less compact representations) to remember them. Thus, fewer fit into working memory. The memory limit does not apply to the cued condition because the observer need remember only the cued letter. Cued change detection is so easy that the observer can tolerate feature degradation and letter distortion and is thus immune to crowding.

The cued change detection task enhances the classic partial report paradigm by making the test easier (same/different instead of identifying one of many possible targets), which increases its sensitivity, so it can reveal degraded memory traces. Here it revealed the jumbled traces of crowded letters that could not be accurately identified.

Acknowledgments

This (draft 72) is the sixth in a series of papers about crowding and its cure, isolating to recognize (#1 Pelli et al., 2004; #2 Martelli, Majaj, & Pelli, 2005; #3 Pelli & Tillman, 2007; #4 Pelli et al., 2007; #5 Levi, Song, & Pelli, 2007). Author Jeremy Freeman is a student at Swarthmore College, and this project was completed while he was a fellow of the 2006 Summer Undergraduate Research Program at the NYU Center for Neural Science. We thank Ed Awh, Ned Block, Marisa Carrasco, Bart Farrell, Brad Motter, Jordan Suchow, and Bosco Tjan for helpful discussion. Special thanks go to Patrick Cavanagh, James Intriligator, and Katharine Tillman for helping us tell the story. Some of these results were presented at the Leadership Alliance National Symposium, Chantilly, VA, July 28, 2006, the annual meeting of the Visual Sciences Society, Sarasota, FL, May 11–16, 2007, and the annual meeting of the Association for the Scientific Study of Consciousness, Las Vegas, NV, June 22–25, 2007. We thank Chiye Aoki (Director of the SURP program) and the Leadership Alliance for their support.

Also supported by National Institutes of Health Grant EY04432 to Denis Pelli.

Commercial relationships: none.

Corresponding author: Denis G. Pelli.

Email: denis.pelli@nyu.edu.

Address: 6 Washington Place, New York, NY 10003, USA.

References

- Banks, W. P., Larson, D. W., & Prinzmetal, W. (1979). Asymmetry of visual interference. *Perception & Psychophysics*, *25*, 447–456. [PubMed]
- Becker, M. W., Pashler, H., & Anstis, S. M. (2000). The role of iconic memory in change-detection tasks. *Perception*, *29*, 273–286. [PubMed]
- Blaser, E., Pappathomas, T., & Vidnyánszky, Z. (2005). Binding of motion and colour is early and automatic. *European Journal of Neuroscience*, *21*, 2040–2044. [PubMed]
- Block, N. (2001). Paradox and cross purposes in recent work on consciousness. *Cognition*, *79*, 197–219. [PubMed]
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. [PubMed]
- Cavanagh, P. (2004). Attention routines and the architecture of selection. In M. I. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 13–28). New York: Guilford Press.
- Cavanagh, P., & Holcombe, A. O. (2007). Non-retinotopic crowding [Abstract]. *Journal of Vision*, *7*(9):338, 338a, <http://journalofvision.org/7/9/338/>, doi:10.1167/7.9.338.
- Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition*, *80*, 61–95. [PubMed]
- Farrell, B., & Pelli, D. G. (1993). Can we attend to large and small at the same time? *Vision Research*, *33*, 2757–2772. [PubMed]
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, *383*, 334–337. [PubMed]
- Holcombe, A. O., & Cavanagh, P. (2001). Early binding of feature pairs for visual perception. *Nature Neuroscience*, *4*, 127–128. [PubMed] [Article]
- Huckauf, A., & Heller, D. (2002). Spatial selection in peripheral letter recognition: In search of boundary

- conditions. *Acta Psychologica*, *111*, 101–123. [[PubMed](#)]
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, *43*, 171–216. [[PubMed](#)]
- Jackson, M. C., & Raymond, J. E. (2006). The role of attention and familiarity in face identification. *Perception & Psychophysics*, *68*, 543–557. [[PubMed](#)]
- James, W. (1890). *The principles of psychology: Part 1* (p. 402). New York: Dover Publications.
- Lamme, V. A. (2004). Separate neural definitions of visual consciousness and visual attention; a case for phenomenal awareness. *Neural Networks*, *17*, 861–872. [[PubMed](#)]
- Landman, R., Spekreijse, H., & Lamme, V. A. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research*, *43*, 149–164. [[PubMed](#)]
- Landolt, E. (1891). Nouvelles recherches sur la physiologie des mouvements des yeux. *Archives d'ophtalmologie*, *11*, 385–395.
- Levi, D. M., Klein, S. A., & Yap, Y. L. (1987). Positional uncertainty in peripheral and amblyopic vision. *Vision Research*, *27*, 581–597. [[PubMed](#)]
- Levi, D. M., Song, S., & Pelli, D. G. (2007). Amblyopic reading is crowded. *Journal of Vision*, *7*(2):21, <http://journalofvision.org/7/2/21/>, doi:10.1167/7.2.21. [[PubMed](#)] [[Article](#)]
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. [[PubMed](#)]
- Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, *5*(1):6, 58–70, <http://journalofvision.org/5/1/6/>, doi:10.1167/5.1.6. [[PubMed](#)] [[Article](#)]
- Morgan, M. J., Ward, R. M., & Castet, E. (1998). Visual search for a tilted target: Tests of spatial uncertainty models. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *51*, 347–370. [[PubMed](#)]
- Nazir, T. A. (1992). Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Research*, *32*, 771–777. [[PubMed](#)]
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Noe, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, *7*, 93–106.
- Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics*, *44*, 369–378.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. [[PubMed](#)]
- Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research*, *46*, 4646–4674. [[PubMed](#)]
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12):12, 1136–1169, <http://journalofvision.org/4/12/12/>, doi:10.1167/4.12.12. [[PubMed](#)] [[Article](#)]
- Pelli, D. G., & Tillman, K. A. (2007). Parts, wholes, and context in reading: A triple dissociation. *PLoS One*, *2*, e680. [[PubMed](#)] [[Article](#)]
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, *7*(2):20, <http://journalofvision.org/7/2/20/>, doi:10.1167/7.2.20. [[PubMed](#)] [[Article](#)]
- Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research*, *31*, 1337–1350. [[PubMed](#)]
- Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, *21*, 409–418. [[PubMed](#)]
- Sagi, D., & Julesz, B. (1985). “Where” and “what” in vision. *Science*, *228*, 1217–1219. [[PubMed](#)]
- Scolari, M., Kohnen, A., Barton, B., & Awh, E. (2007). Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays? *Journal of Vision*, *7*(2):7, 1–23, <http://journalofvision.org/7/2/7/>, doi:10.1167/7.2.7. [[Article](#)]
- Simons, D. J., & Ambinder, M. S. (2005). Change blindness—Theory and consequences. *Current Directions in Psychological Science*, *14*, 44–48.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, *74*, 1–29.
- Strasburger, H. (2005). Unfocused spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision*, *5*(11):8, 1024–1037, <http://journalofvision.org/5/11/8/>, doi:10.1167/5.11.8. [[PubMed](#)] [[Article](#)]
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*, 1349–1357. [[PubMed](#)]
- Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 25, 1016–1035. [[PubMed](#)]
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107–141. [[PubMed](#)]
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136. [[PubMed](#)]
- Van der Lubbe, R. H., & Keuss, P. J. (2001). Focused attention reduces the effect of lateral interference in multi-element arrays. *Psychological Research*, 65, 107–118. [[PubMed](#)]
- Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A, Optics, image science, and vision*, 14, 2057–2068. [[PubMed](#)]
- Wolford, G., & Shum, K. H. (1980). Evidence for feature perturbations. *Perception & Psychophysics*, 27, 409–420. [[PubMed](#)]