

Crowding is unlike ordinary masking: Distinguishing feature integration from detection

Denis G. Pelli

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



Melanie Palomares

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



Najib J. Majaj

Center for Neural Science, New York University,
New York, NY, USA



A letter in the peripheral visual field is much harder to identify in the presence of nearby letters. This is “crowding.” Both crowding and ordinary masking are special cases of “masking,” which, in general, refers to any effect of a “mask” pattern on the discriminability of a signal. Here we characterize crowding, and propose a diagnostic test to distinguish it from ordinary masking. In ordinary masking, the signal disappears. In crowding, it remains visible, but is ambiguous, jumbled with its neighbors. Masks are usually effective only if they overlap the signal, but the crowding effect extends over a large region. The width of that region is proportional to signal eccentricity from the fovea and independent of signal size, mask size, mask contrast, signal and mask font, and number of masks. At 4 deg eccentricity, the threshold contrast for identification of a 0.32 deg signal letter is elevated (up to six-fold) by mask letters anywhere in a 2.3 deg region, 7 times wider than the signal. In ordinary masking, threshold contrast rises as a power function of mask contrast, with a shallow log-log slope of 0.5 to 1, whereas, in crowding, threshold is a sigmoidal function of mask contrast, with a steep log-log slope of 2 at close spacing. Most remarkably, although the threshold elevation decreases exponentially with spacing, the threshold and saturation contrasts of crowding are independent of spacing. Finally, ordinary masking is similar for detection and identification, but crowding occurs only for identification, not detection. More precisely, crowding occurs only in tasks that cannot be done based on a single detection by coarsely coded feature detectors. These results (and observers’ introspections) suggest that ordinary masking blocks feature detection, so the signal disappears, while crowding (like “illusory conjunction”) is excessive feature integration — detected features are integrated over an inappropriately large area because there are no smaller integration fields — so the integrated signal is ambiguous, jumbled with the mask. In illusory conjunction, observers see an object that is not there made up of features that are. A survey of the illusory conjunction literature finds that most of the illusory conjunction results are consistent with the spatial crowding described here, which depends on spatial proximity, independent of time pressure. The rest seem to arise through a distinct phenomenon that one might call “temporal crowding,” which depends on time pressure (“overloading attention”), independent of spatial proximity.

Keywords: crowding, masking, peripheral vision, feature integration, illusory conjunction, critical spacing, letter identification, object recognition, isolation field, integration field, second-order mechanisms

1. Introduction

Object identification involves the moderately well understood process of feature detection, followed by a mysterious “integration” process that combines the detected features to produce a classification decision. The purpose of this paper is to characterize “crowding.” Crowding is excessive integration, which spoils identification and reveals the inner workings. With this characterization in hand, one can address some longstanding questions about object identification, such as whether faces are recognized by parts and the roles of letter and word recognition in reading (Martelli, Pelli, & Majaj, [in press](#); Su, Berger, Majaj, & Pelli, 2004).

Crowding and ordinary masking are special cases of masking. In general, “masking” refers to the impairment of

the discriminability of a signal by another pattern. Ordinary masking, such as masking by gratings (Legge & Foley, 1980; Swift & Smith, 1983; Levi, Klein, & Hariharan, 2002) or noise (Stromeyer & Julesz, 1972; Pelli & Farrell, 1999), is usually only effective when the mask overlaps the signal. However, in the normal periphery or the amblyopic fovea, neighboring letters with no overlap severely impair the identification of a signal letter (Korte, 1923; Ehlers, 1936, 1953; Bouma, 1970; Anstis, 1974; Flom, 1991). This particular masking phenomenon is called “crowding” (Stuart & Burian, 1962; for historical review, see Strasburger, Harvey, & Rentschler, 1991, or Strasburger, 2002). Crowding is not specific to letters. We will argue that ordinary masking occurs when signal and mask stimulate the same feature detector and that crowding occurs when signal and mask stimulate different feature detectors that both reach

the same feature integrator (where features are combined to recognize an object).

Despite progress in vision research, we still can only barely begin to answer a simple question like, “How do I recognize the letter A?” The literature on grating detection, with the ideas of spatial frequency channels (feature detectors) and probability summation, offers a good answer to the easier question, “How do I tell whether the screen is blank?” The answer goes under many names, including “channels” and “probability summation.” We follow Graham (1980) in calling it “feature detection.” The observer has many independent units, “feature detectors,” each with a receptive field (linear weighting over space and time, summed to yield one number) followed by a nonlinear process that results in a sharply increasing probability of response with contrast.^{1,2} The image that matches a detector’s receptive field is called its “feature.” All feature detectors operate independently and the observer detects a displayed image if and only if any of the detectors do (Brindley, 1960; Quick, 1974; Graham, 1980, 1989; Robson & Graham, 1981).

In the various relevant papers, the word “feature” sometimes refers, as above, to the elementary component of the visual analysis (e.g., Graham, 1980) and sometimes refers to the labeled value (e.g., “red” or “A” or “triangle”) of a stimulus dimension that the experimenter chose to vary (e.g., Treisman & Schmidt, 1982, p. 139). We will be referring to elementary features, except in the Section 4.7 discussion of illusory conjunctions and Feature Integration Theory.

Elementary-feature detection provides a good account of detecting simple targets (i.e., for which detection of a single feature suffices for a correct response). However, identifying (or detecting a second-order signal) usually requires combining the information from several feature detections to respond correctly (see Chubb, Olzak, & Derrington, 2001). This (nonlinear) assembly process is called “feature integration” (or “binding”). Feature integration may internally represent the combined features as an object, but we will not address that here. We will suggest that crowding is excessive feature integration, integrating over an inappropriately large area that includes the flanking mask as well as the signal.

This Introduction presents a simple intuition (Section 1.1) that brings together ideas about feature detection (1.2) with facts of ordinary masking (1.3) and crowding (1.5). Later, in Discussion, we will review the close connection between crowding and illusory conjunction (4.7).

1.1 Overview

This paper characterizes crowding, distinguishing it from ordinary masking. We believe that the term “crowding” should encompass not just the original task of identifying a letter among letters in the periphery (or amblyopic fovea), but also any other task with similar results: critical spacing proportional to eccentricity and independent of

size. A diagnostic test is proposed in Discussion (Section 4.1).

Past attempts to characterize and explain crowding have each varied a few parameters in similar tasks. In this experimental and theoretical synthesis we have tried to be more comprehensive. As we attempt to put it all together into one story, there are many points of agreement between our proposed explanation and earlier suggestions, but there are also some important differences. What is new here arrived late in the process, forced upon us by the data, after a long period of stumbling in the dark.

Perhaps the most important new fact emerging from this union of old and new results is the effect of which task the observer is assigned. In ordinary masking the signal disappears, so the observer cannot say anything about it, and fails all tasks (Thomas, 1985b). Many investigators have assumed that this would be true of crowding as well (e.g., see Cavanagh, 2001). But, in fact, conditions of crowding that severely impair identification of a letter (reported here) or orientation of a grating (Wilkinson, Wilson, & Ellemberg, 1997) have little or no effect on the detectability of the target. Observers report seeing a jumbled target that incorporates features from the mask. We struggled with this detection/identification dichotomy for a long time, and failed in our attempts to crowd gratings, until we eventually realized that the dichotomy is more subtle than just detection versus identification. All the tasks susceptible to crowding are tasks that, with some plausible assumptions, require more than one feature-detection event (a “conjunction” of several feature detections). Tasks that require only a single feature-detection event are immune, or nearly so. This parallels the dichotomy found in searching for one feature versus a conjunction of features — a feature pops out and a conjunction does not³ — and is strong evidence that crowding interferes with feature integration, not feature detection. The multiple detections must be integrated, and that integration is susceptible to crowding; the single detection doesn’t need to be integrated, so there’s no crowding.

Previous authors, aware that ordinary masking is selective, have shown that crowding too is selective (e.g., Kooi, Toet, Tripathy, & Levi, 1994). Here we compile old and new results showing that the selectivity of crowding is vastly broader than that of ordinary masking. Ordinary masking reveals the narrow selectivity of a feature detector (the first stage), whereas crowding reveals the broad selectivity of a feature integrator (the second stage).

It is more-or-less established that in ordinary masking the same feature detector mediates the effects of mask and signal (Legge & Foley, 1980; Foley & Chen, 1999; Wilson & Kim, 1998).¹ A new finding, the effect of mask contrast as a function of spacing (Section 3.6), provides strong evidence that, in crowding, distinct feature detectors mediate the effects of mask and signal.

We survey the literature on illusory conjunctions at the end of Discussion (Section 4.7), but the only prerequisite for reading that section is the vocabulary established here

in the [Introduction](#). Most of the illusory conjunction papers' results are consistent with crowding, as defined here, but a few papers, including Treisman and Schmidt (1982), describe a different phenomenon that we will call "temporal crowding."

1.2 Feature detection and integration

The familiar notion that the observer detects features (components of the image) independently and then integrates them to perceive an object goes back to Weber's (1834, 1846) and Sherrington's (1906) suggestions, based on their psychophysical evidence, that neural receptive fields mediate the sense of touch. Indeed, simply supposing that independent detection of features is a necessary first stage of vision (i.e., cannot be bypassed) implies that any observer response (e.g., object recognition) that communicates information about a combination of features must be based on an integration (combination) of several detected features (e.g., Selfridge, 1959; Neisser, 1967; Campbell & Robson, 1968; Thomas, Padilla, & Rourke, 1969; Rosch & Lloyd, 1978; Treisman & Gelade, 1980; Sagi & Julesz, 1984; Olzak & Thomas, 1986). Despite its appealing simplicity, feature detection has been hard to establish convincingly. The grating detection literature is convincing (e.g., Campbell & Robson, 1968; Robson & Graham, 1981; Graham, 1989), but that leaves open the possibility that other tasks and targets (e.g., identifying letters) might bypass feature detection. Judging whether or not a screen is blank, as one does in detection experiments, might not be representative of what the visual system can do. Some capabilities might appear only for important highly practiced tasks, like reading faces or text. Part of this concern is allayed by the finding that thresholds for identifying letters, across the entire range of size, font, and alphabet, is accounted for by a slight extension of the standard "probability summation" model of independent feature detection (Pelli, Burns, Farell, & Moore, [in press](#)). Finding, as predicted by feature detection, that efficiency for identification is inversely proportional to complexity (number of features), even when highly practiced, is strong evidence that observers cannot bypass the feature-detection bottleneck (Pelli, Farell, & Moore, 2003).

We all want to know how features are integrated, but findings to date provide only hints as to the nature of this computation. Perception of coherent motion of two-grating plaids is based on a nonlinear combination of the two grating components (Adelson & Movshon, 1982) and some MT neurons actually implement this combination rule (Movshon, Adelson, Gizzi, & Newsome, 1986). Speed discrimination is affected by whether the components are perceived to form an object (Verghese & Stone, 1995, 1996). Applying the classic summation paradigm to motion discrimination and texture segregation reveals the exponent of the nonlinear combination of multiple components (Morrone, Burr, & Vaina, 1995; Graham & Sutter, 1998). Accounts of texture discrimination suppose linear combina-

tion of nonlinearly transformed feature detection signals (for review, see Chubb et al., 2001; Landy & Graham, 2004). Visual search and crowding experiments have also contributed hints, as we will see below. Accounts of the feature integration that underlies identification of objects are more speculative. Much of the debate has distinguished the recognition-by-components approach championed by Biederman (1987) from the alignment approach championed by Ullman and Poggio (see Tarr & Bulthoff, 1998). Alas, putting together the hints from all these studies fails to provide clear guidance as to how to address the larger question of what kind of computation underlies object recognition.

1.3 Ordinary masking

Masking provides an important part of the evidence for feature detection. Masking goes beyond the narrow domain of the question, "Is the screen blank?" to examine the effect of an irrelevant background mask on visibility of the signal. In ordinary masking, it is generally supposed that the mask affects the visibility of the signal only to the extent that the mask stimulates the receptive fields of the feature detectors that pick up the signal. We will argue that crowding cannot be explained as ordinary masking (i.e., mediated by mask stimulation of the feature detector(s) that detect the signal).

Ordinary masking is most effective when the mask has more or less the same spatial frequency, orientation, and location as the signal (Legge & Foley, 1980; Phillips & Wilson, 1984; Levi, Klein, et al., 2002). Critical-band masking experiments have shown that the spatial frequency tuning of grating detection (Greis & Rohler, 1970; Stromeyer & Julesz, 1972; Solomon & Pelli, 1994) and letter identification (Solomon & Pelli, 1994; Majaj, Pelli, Kurshan, & Palomares, 2002; Chung, Levi, & Legge, 2001) is 1.6 octaves wide. And it is independent of eccentricity, having the same tuning in central and peripheral vision (Mullen & Losada, 1999).

Ordinary masking has very similar effects on detection and identification (Thomas, 1985a, 1985b). As we shall see, our results show that crowding affects only identification, not detection. (We would expect crowding to affect detection of second-order signals, but no one has tried it yet.) With no mask, threshold contrasts for identifying a signal are usually higher than for detecting it, but, for a wide range of signal size (Pelli et al., [in press](#)) and viewing eccentricities (Raghavan, 1995; Thomas, 1987), identification and detection thresholds are in a constant ratio (also see Graham, 1985). In critical band masking studies, channel frequencies for detection and discrimination (of letters and gratings) are the same (Majaj et al., 2002). Threshold contrasts for identification and detection have similar dependence on mask contrast (Raghavan, 1995; Pelli, Levi, & Chung, 2004). These characteristics of ordinary masking are evidence for the popular idea that ordinary masking impairs discriminability of the signal by directly stimulating the feature detector that mediates our judgments about the

signal. The very different characteristics of crowding will require a different kind of explanation.

1.4 Scope

We restrict our scope to simultaneous mask and signal, of any duration. A flanker that is delayed or prolonged relative to the signal can produce “metacontrast” or “object substitution” masking (Breitmeyer, 1984; Enns & Di Lollo, 1997, 2000; Tata, 2002; Enns, 2004). These phenomena seem to be closely related to motion perception (Didner & Sperling, 1980; Reeves, 1982; Burr, 1984; Bischof & Di Lollo, 1995), and may be related to what we will call “temporal crowding.” They are not directly relevant to understanding (spatial) crowding, and will not be discussed here (see Huckauf & Heller, 2004).

1.5 Crowding

Our final conclusions rest on objective measurements: thresholds for detection and identification. However, the subjective crowding experience, all by itself, makes a strong case for a key point. Examine the two blocks of letters in this demo while fixating on the central cross:

A	A		B	A
A	A	+	A	B

What you see on the left is a block of four A's. What you see on the right is much harder to describe. It's a block of four letter-like objects. But they aren't clearly A's or B's; they're in-between and unstable. Each letter may seem at times to be an A and sometimes a B, but most of the time it has a confusing hybrid A-B appearance that would be impossible to draw. We usually assume that visual object recognition segments the scene and accounts for each segment by hypothesizing an “object” with appropriate properties. One supposes that all the object's properties are estimated from the same image segment. Surprisingly, this demo shows that a single object's several properties are estimates from various regions, large and small. Each letter is an object. The perceived presence and locations of the letters distinguish four objects, arranged in a square. To resolve four items these properties must each be assessed over a more-or-less one-letter region. Yet each item's shape has a hybrid A-B appearance, incorporating information from a region that includes several letters. (Using your finger to cover other letters in the demo above, you will find that to see one letter clearly you must cover the rest of the letters in the block.) This seriously undermines the notion of object recognition as a unitary process that takes in a region of the image and emits an “object” with properties. Instead, our demo shows that, in this case, the distinct properties of location (where) and shape (what) are estimates from very differently sized regions. Perhaps, despite its unitary appearance, an “object” is just a loose bundle of independ-

ently estimated properties. [This differs from the Wolfe & Bennett (1997) suggestion that loose bundling results from inattention. Our demo of loose bundling occurs with full attention.] This demo, like the rest of this paper, reveals a dichotomy between properties (e.g., presence or location) that may be estimated from a single detected feature and those (e.g., letter identity or shape) that require integration of several features.

It is as if there is a pressure on both sides of the word that tends to compress it. Then the stronger, i.e. the more salient or dominant letters, are preserved and they ‘squash’ the weaker, i.e. the less salient letters, between them. (Korte [1923], translated by Uta Wolfe)

It looks like one big mess. I keep seeing [the letter] ‘A’ even though there is no ‘A’ in the Sloan alphabet. I seem to take features of one letter and mix them up with those of another. (Observer JG)

When it's difficult, I see a unit that is a combination of letters and I can't say how many there are. (Observer MLL)

I know that there are three letters. But for some reason, I can't identify the middle one, which looks like it's being stretched and distorted by the outer flankers. (Observer MCP)

These are observers' descriptions of how they see a letter that is flanked by other letters in the periphery. This was first described by Korte (1923), and was dubbed *crowding* by Stuart & Burian (1962). They and others showed that acuity is greatly impaired by crowding (Ehlers, 1936, 1953; Woodworth, 1938; Flom, Weymouth, & Kahneman, 1963; Bouma, 1970), which backs up the introspective descriptions by objective measurement of impaired form recognition.

For identifying a letter among letters, the spatial extent of crowding is roughly half the eccentricity (Bouma, 1970; Toet & Levi, 1992). For identifying a numeric character among numeric characters, Strasburger et al. (1991) reported a similar proportionality constant, 0.4, independent of character size (0.05 - 1.4 deg). Latham and Whitaker (1996) report similar results for a 3-bar acuity target among four such distractors of random orientation. Tripathy and Cavanagh (2002) report similar results for identifying the orientation of a T among “squared thetas.” Wilkinson et al. (1997), as well, report a proportionality constant of 0.4 for fine discrimination of the contrast or spatial frequency of a grating among gratings. Levi, Hariharan, and Klein (2002, p. 175) report a (center-to-center) proportionality constant of 0.5 for masking of an E by a bar, both made up of grating patches.

This scaling with eccentricity, independent of size, is utterly unlike ordinary masking, where critical spacing

scales with signal size, independent of eccentricity. As we'll see, the most dramatic difference – for us the *defining* difference (Section 4.1) – between crowding and ordinary masking is the complementary effects of signal size and eccentricity.

Many lateral masking studies have varied size and eccentricity, but, unfortunately, typically not in a way that would distinguish crowding from ordinary masking. Under the rationale that acuity scaling would provide a more level playing field for comparing different eccentricities, most studies that varied signal size or eccentricity, varied both together, roughly in proportion (e.g., Andriessen & Bouma 1976; Loomis, 1978; Jacobs, 1979; Santee & Egeth, 1982b; Chung et al., 2001). Alas, proportional increase of the stimulus size and spacing with eccentricity would not be expected to affect either crowding or ordinary masking and thus does not distinguish the two kinds of effect. Chung et al. (2001) studied some of the properties of crowding of letters by letters to compare crowding with ordinary “pattern” masking of gratings by gratings. Filtering target and mask letters to one-octave bands, they identified the most effective mask frequency as a function of target frequency, and found that this agreed with the earlier literature on ordinary masking. At a large, near-critical spacing they found a shallow log-log slope (0.13 – 0.3) for the effect of mask contrast on threshold contrast for identifying the target, which they noted is much shallower than the slopes of 0.5 to 1 generally found in ordinary masking. Using ordinary, unfiltered letters we further investigate the contrast response function here (Figures 9–11, below).

Levi, Klein, et al. (2002) and Levi, Hariharan, et al. (2002) used a tumbling E and a flanking bar that were both made up of grating patches. They separately varied eccentricity, grating frequency, and patch extent. In the fovea, the critical spacing was proportional to signal extent, consistent with ordinary masking. In the periphery, the critical spacing was proportional to eccentricity, consistent with crowding.

Another important difference between crowding and ordinary masking is that ordinary masking blocks both detection and identification – the signal disappears – whereas crowding affects only identification – the signal remains visible, but is jumbled with the mask. This dichotomy has not been spelled out in the earlier literature, although Wilkinson et al. (1997) noted a much weaker effect of crowding on detection than on identification: Their signals were still detectable when they could no longer be identified.

Because the range of crowding is roughly half the eccentricity, it extends only a few minutes of arc for foveal targets (Flom, Weymouth, et al., 1963; Bouma, 1970; Loomis, 1978; Jacobs, 1979; Levi, Klein, & Aitsebaomo, 1985; Toet & Levi, 1992; Wilkinson et al., 1997; Leat, Li, & Epp, 1999; Hess, Dakin, & Kapoor, 2000; Chung et al., 2001). Liu and Arditi (2000) found that letter-string length

is underestimated when observers are asked to judge the number of acuity-sized letters in the fovea. Their descriptions of this foveal effect are similar to those by Korte (1923) and our observers of crowding in the periphery, but with the greatly reduced range, less than 5 arcmin, that one would expect from its proportionality to eccentricity. Thus crowding treats fovea and periphery alike, following one eccentricity rule throughout.

Lateral masking studies with larger signals find no effect of nonoverlapping flankers on foveal targets (Strasburger et al., 1991; Leat et al., 1999). Bondarko and Danilova (1996; 1997) showed that nonoverlapping bars slightly decrease acuity for a Landolt C signal in the fovea. In foveal tasks that do show effects of laterally displaced masks, the spatial extent of the lateral interference scales with the size of the signal: maximum effect at a spacing of 5 times the gap width of a Landolt C (Flom, 1991) and 3 times the wavelength of a grating (Polat & Sagi, 1993; Levi, 2000). Levi, Klein, et al. (2002) found that the critical spacing of a tumbling E and flanking bars (all made up of grating patches) is proportional to signal extent over a 50:1 range, independent of spatial frequency. This scaling with signal size is characteristic of ordinary masking and unlike crowding.

Because our experiments are done mostly with letters, we postpone until Discussion the rest of our review of crowding with other stimuli (Section 4.2). What we have reviewed so far tells us to look at the effects of spacing, eccentricity, contrast, and task. With those results in hand, we will be ready to tackle illusory conjunctions (Section 4.7).

1.6 Our study

We begin by replicating previous results on the spatial extent of crowding as a function of viewing eccentricity. We then explore the effects of varying signal and mask (size, contrast, complexity, and type: letter and grating) and task (identification and detection). (See Table 1.) The effects of spacing, eccentricity, size, contrast, and task distinguish crowding from ordinary masking. The other manipulations help characterize the selectivity of crowding. The selectivity of ordinary masking is that of the feature detector. Our results indicate that the selectivity of crowding is that of the feature integrator.

The experiments were exploratory, trying to characterize the phenomenon, especially as a window into the mysterious feature integration process. The results indicate that the observer's identification response is based on an amalgam of all the features detected in a large region we call the “integration field,” which is approximately centered on the signal (Toet & Levi, 1992). Most relevant to this conclusion are the effect of task and the combined effects of mask contrast and spacing.

Figure	Effect of	Task	Signal and flanker (font or grating)	Signal size (deg)	Flanker size (deg)	Ecc. (deg)	Observer
3	eccentricity	identify	Sloan	1	1	0, 2, 4, 8, 12, 20, 24	MCP, SJR, SSA
4	fovea vs. periphery	identify	Sloan	0.32	0.32	0, 4	MCP, AG, MLL
5	size	identify	Sloan	0.32, 0.5, 1, 2	same as signal	4	MCP, AG, SSA
6	flanker size	identify	Sloan	0.32	0.32, 0.64, 0.96, 1.6, 3.2	4	MCP, AG
7	font	identify	2x3 Checkers, Sloan, Bookman, Outline Sloan	0.25, 0.32, 0.50, 0.32	same as signal	4	MCP, MLL
8	# of flankers	identify	Sloan	0.32	0.32	6	MS, MLM, MCP
9, 11	flanker contrast	identify	Sloan	0.32	0.32	4	MCP, AG, MLL
12	task	identify, detect	Sloan	0.32	0.32	4	MCP, AG, MLL
13	eccentricity	detect	Sloan	0.75	0.75	2, 4, 8	MLM
14	size	detect	Sloan	0.75, 1.5, 3.0	same as signal	8	MCP, MLL
15	extent	identify	1 c/deg grating	2, 4, 8	same as signal	20	MCP, AG
16	letter vs. grating	identify, detect	Sloan, 8 c/deg grating	0.32, 0.52	0.32, 0.52	4	MCP, AG, MLL

Table 1. The experiments. For gratings, "size" is the 1/e radius of the Gaussian envelope, and the observer "identified" the $\pm 45^\circ$ orientation. Regarding Figure 8, observer MCP was tested at 4 instead of 6 deg eccentricity.

2. Methods

2.1 Observers

Seven observers with normal or corrected-to-normal acuity performed these experiments binocularly (see Table 1). One observer (MCP) is an author. The other observers were paid for participating.

2.2 Tasks and stimuli

All experiments were performed on Apple Power Macintosh computers using MATLAB software with the **Psychophysics Toolbox** extensions (Brainard, 1997; Pelli, 1997). The background luminance was set to the middle of the monitor range, about 18 cd/m². Sloan letters were based on Louise Sloan's design specified by the NAS-NRC (1980). (The Sloan font is available from <http://psych.nyu.edu/pelli/software.html>.) Sloan letters were usually 0.32 deg high and wide. Sinewave gratings were 1 or 8 c/deg with a circularly symmetric Gaussian envelope with a 1/e radius that we specify as "size."

Observers viewed a gamma-corrected grayscale monitor (Pelli & Zhang, 1991). The fixation point was a 0.15 deg black square. The position of the fixation point on the screen determined the eccentricity of the signal (always presented at the center of the display). For peripheral viewing conditions, the fixation point was displayed for the entire trial. For foveal viewing, the fixation point was presented for 200 ms, followed by a 200 ms blank and then the signal. The signal, flanked by two horizontally aligned high-contrast masks of either letters or gratings, appeared at the

center of the screen for 200 ms (Figure 1). Signal eccentricity was controlled by varying the position of the fixation point on the screen. Thus the signal was presented at various eccentricities along the horizontal meridian in the right visual field. Letter contrast is defined as the ratio of luminance increment to background. Letter contrast can be greater than 1. Flanker contrast was usually 0.85. Each signal presentation was accompanied by a beep. Mask-to-signal spacing is measured center to center. Usually the signal and each flanking letter were independent random samples from the same alphabet. A response screen followed, showing all the possible signals (usually the 10 letters **CDHKNORSVZ** of the Sloan alphabet) at 80% contrast. Observers identified the signal by using a mouse-controlled cursor to point and click on their answer. Correct identification was rewarded with a beep.

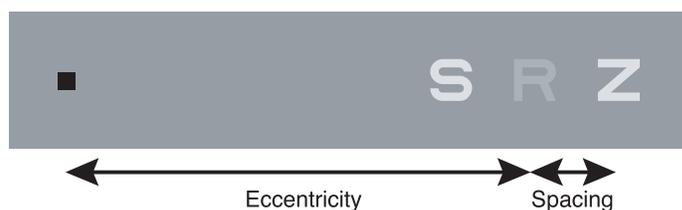


Figure 1. Typical condition for crowding. The black square is a 0.15 deg fixation mark. The signal is a faint 0.32 deg Sloan letter at 4 deg in the right visual field. Two 85%-contrast masks (S, Z) flank a signal letter (R) with a signal-to-mask center-to-center spacing of 0.64 deg. Letter contrast is defined as the ratio of luminance increment to background. Letter contrast can be greater than 1. The signal contrast changes from trial to trial.

The signal duration (200 ms) is too brief for eye movements in response to the signal to help see it. We occasionally watched the observer's eyes while the observer was doing the task to detect anticipatory eye movements, but we never saw any. The results presented in this paper (e.g. Figure 3a) reveal a more-than-tenfold threshold elevation and a steep dependence on spacing. Anticipatory eye movements would reduce the signal eccentricity by an amount that would vary between trials and among observers. The steep dependence of threshold on spacing (e.g., Figure 3a) and the consistent critical spacing among observers (e.g., Figure 3b) indicate that anticipatory eye movements were not a problem.

Threshold contrast was measured by a modified QUEST staircase procedure (Watson & Pelli, 1983; King-Smith, Grigsby, Vingrys, Benes, & Supowit, 1994) using an 82% criterion and β of 3.5 for 40-trial runs. Log thresholds were averaged over two runs for each condition.

In the detection task, the signal letter was randomly presented in one of two consecutive intervals. The flankers were displayed in both intervals, independently randomly selected for each interval. Observers indicated their choice of interval by clicking the mouse once for first and twice for second. Correct responses were rewarded with a beep.

2.3 Clipped line fit

Strasburger et al. (1991) suggested that threshold contrast for target identification is a good way to measure the effect of crowding, and we agree. Most of our data are threshold contrast plotted against spacing, and have a generally sigmoidal shape. We fit a clipped line to the data by eye. This fit has three parts: a horizontal ceiling, a falling slope, and a horizontal floor (Figure 2). *Threshold elevation* (a ratio) is measured from floor to ceiling. *Critical spacing* is

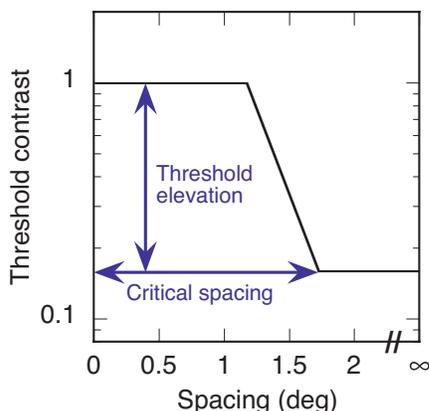


Figure 2. Clipped line fit: threshold contrast as a function of center-to-center spacing of signal and flanker. (At zero spacing the signal and flanker are superimposed, added on top of one another.) We fit a clipped line to each data set by eye. Two parameters of that fit are of interest. *Threshold elevation* is the ratio of thresholds at zero and infinite flanker spacing (i.e., ceiling:floor ratio). *Critical spacing* is the least spacing at which there is no threshold elevation (i.e., edge of the floor).

the least spacing at which there is no threshold elevation in the fit (i.e., edge of the floor).

3. Results

Figures 3–16 present our results. To help the reader make sense of it all, Table 2 presents the nine empirical differences between crowding and ordinary masking. We recommend focusing on the sheer strangeness of crowding. Our intuitions, based on familiarity with ordinary masking, were defied at every turn. The single most important result

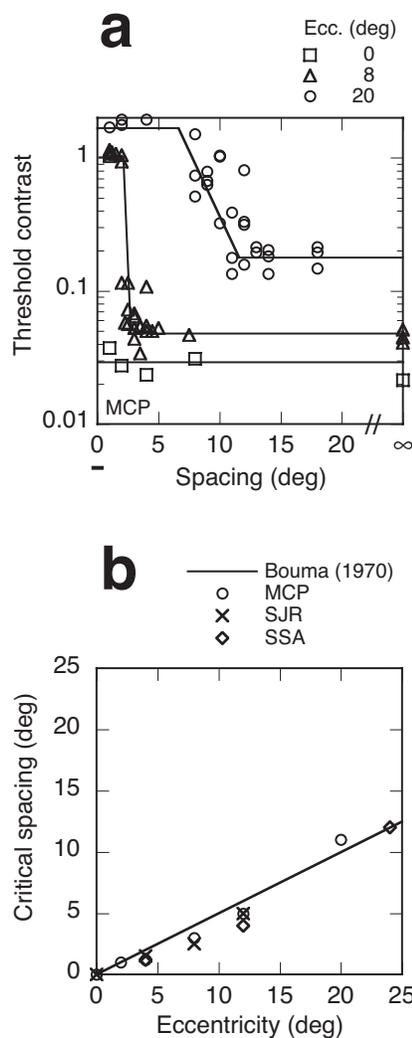


Figure 3. Effect of eccentricity. Each symbol is the geometric average of two threshold estimates. (a). Threshold as a function of flanker spacing for a 1 deg Sloan letter between flankers at various eccentricities. The solid lines are clipped-line fits, as explained in Figure 2. The horizontal line at the bottom left, below the graph, represents the width of the signal in deg. Contrast is the ratio of increment to background, and can exceed 1. (b). Critical spacing plotted against viewing eccentricity for three observers. Like Bouma (1970), we find that critical spacing is roughly half of viewing eccentricity. Observers MCP, SJR, and SSA.

Fact	Ordinary masking	Crowding	Figures
a	Similar in fovea and periphery.	Normally evident only in the periphery (Korte, 1923; Stuart & Burian, 1962; Flom, Weymouth, et al., 1963; Bouma, 1970).	3, 4
b	Signal disappears, suppressed by mask.	Signal is visible but ambiguous, incorporating features from mask (Korte, 1923; Flom, Weymouth, et al., 1963; Andriessen & Bouma, 1976; Wolford & Shum, 1980; Wilkinson et al., 1997; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001).	1, 10
c	Occurs for any task and signal.	So far, specific to identification of letters (Flom, 1991), orientation of tumbling E (Levi, Hariharan, et al., 2002), and fine discrimination of contrast, spatial frequency, and orientation (Andriessen & Bouma, 1976; Wilkinson et al., 1997; Parkes et al., 2001).	12 - 16
d	Similar effect on identification and detection.	Little or no effect on detection (Wilkinson et al., 1997) <i>and coarse discrimination</i> .	12 - 14, 16
e	Narrow critical spacing, little or no effect of nonoverlapping mask.	Wide critical spacing can be more than 10 times bigger than a small signal (Korte, 1923; Stuart & Burian, 1962; Bouma, 1970; Toet & Levi, 1992; Levi, Hariharan, et al., 2002).	3 - 5
f	Critical spacing scales with signal, independent of eccentricity.	Critical spacing is roughly half of viewing eccentricity (Bouma, 1970; Toet & Levi, 1992), independent of signal size (Strasburger et al., 1991; Levi, Hariharan, et al., 2002), <i>mask size, mask contrast, and number of masks</i> .	3 - 11, 13 - 15
g	Spatiotemporal selectivity more or less consistent with a receptive field.	<i>Remarkably unselective, showing equal effect over a wide range of flanker type (letter, black disk, or square; Eriksen & Hoffman, 1973; Loomis, 1978), flanker size (10:1), and flanker number (≥ 2).</i>	6, 8
h	Shallow power-law contrast response (log-log slope of 0.5 to 1).	<i>Steep sigmoidal contrast response. Log-log slope of 2 for close spacing. Log ceiling and log slope fall exponentially with spacing.</i>	10, 11
i	Threshold mask contrast depends on spacing. No saturation.	<i>Threshold and saturation mask contrasts are independent of spacing.</i>	11

Table 2. Facts: summary of the differences between crowding and ordinary masking. We cite the authors of the known facts about crowding, many replicated here, and *italicize our new findings*. We take line f as the defining difference: critical spacing scales with eccentricity, not size. (a). The extremely-short-range foveal effect described by Liu and Arditi (2000) is likely to be crowding. (c). Andriessen and Bouma (1976) show a large crowding-like effect of flanking bars on fine discrimination of bar orientation, and a small effect on detection threshold, too small to account for the effect on orientation discrimination. Illusory conjunction provides evidence for crowding of conjunction of color vs. shape (Treisman & Schmidt, 1982). (d). The critical spacing for detecting a letter among letters can be as large as that for identification, but we call it ordinary masking, not crowding, because it scales with letter size (Figure 14b), not eccentricity (Figure 13b). Despite refutals of the feature vs. conjunction dichotomy in the search literature, we still expect a robust feature vs. conjunction dichotomy in crowding.³ (e). Figures 12, 16a, and 16b are examples of weak effects of nonoverlapping masks in ordinary masking. (f). At 0 deg eccentricity, Levi, Klein, et al. (2002) found that critical spacing is proportional to signal size over a 50:1 range. "Roughly half" can be as low as 0.3, as in Figure 5b. Fine (2003) also reported crowding to be independent of contrast. (g). We say "more or less consistent" because current feature detector models to explain ordinary masking have not just one but several similar receptive fields (to implement divisive inhibition) as noted earlier.¹ The spatiotemporal selectivity found by Chung et al. (2001) with filtered letters is like that for ordinary masking, unlike our summary for crowding, but it is not certain whether their paradigm elicited crowding or ordinary masking (see Section 4.6). Many studies have documented systematic effects of the similarity of target and flanker (e.g., Estes, 1982; Ivry & Prinzmetal, 1991; Nazir, 1992; Kooi et al., 1994; Donk, 1999; Chung et al., 2001).

is Bouma's (1970), greatly extended here, that critical spacing is roughly half the eccentricity (distance from fixation), independent of everything else.

There is a minor caveat to Bouma's rule, but it does not affect the basic intuition. The caveat is that critical spacing is asymmetric, greater in the peripheral than in the central direction from the target (Bouma 1970, 1973; Townsend, Taylor, & Brown, 1971; Banks, Larson, & Prinzmetal, 1979; Chastain & Lawson, 1979; Wolford & Shum, 1980; Toet & Levi, 1992). It is greater in radial directions (peripheral and central) than in circumferential directions (Chambers & Wolford, 1983; Toet & Levi, 1992). It is greater in the upper than in the lower visual

field (Intriligator & Cavanagh, 2001). These details matter when comparing results across conditions, but they reinforce the basic intuition that the extent of crowding depends almost exclusively on the local anatomy of the visual field, independent of the signal, unlike ordinary masking, which is co-extensive with the signal, independent of location in the visual field.

3.1 Effects of spacing and eccentricity

One of the stranger aspects of crowding is Bouma's (1970) finding that the critical spacing is proportional to eccentricity, which we replicate here. We measured the threshold contrast for identifying a 1 deg letter as a func-

tion of signal-to-mask spacing. The signal was at 0 to 24 deg eccentricity in the right visual field (see Table 1). There were two flankers, one to the left and one to the right of the signal. Figure 3a shows that the letter masks have a very strong effect, raising threshold tenfold. For each eccentricity, the clipped-line fit provides an estimate of the critical spacing. Figure 3b shows that critical spacing is proportional to eccentricity. Our data confirm the finding (Bouma, 1970; Strasburger et al., 1991; Toet & Levi, 1992; Levi, Hariharan, et al., 2002, p. 173) that the critical spacing is roughly half of the viewing eccentricity. (Bouma, 1970, was right to say “roughly” 0.5. For some of our data, this value drops as low as 0.3, as we will see below.) Andriessen and Bouma (1976) report a critical spacing of 0.4 of eccentricity for fine discrimination of line orientation. Wilkinson et al. (1997) report a critical spacing of 0.4 of eccentricity for fine discrimination of crowded grating contrast and spatial frequency, and slightly higher for fine discrimination of orientation.

Figure 4 shows threshold for observer AG in the presence of one mask, as a function of horizontal mask offset, for a 0.32 deg signal. The width of the critical region is the sum of the critical spacings, left and right. Separate curves show results at 0 and 4 deg eccentricity. In the fovea, the critical region (i.e., the sum of critical spacings left and right) is about as wide (0.40 deg) as the signal (0.32 deg). In the periphery, the critical spacings are 1.00 deg to the left and 1.25 deg to the right, for a total critical region width (2.25 deg) about 7 times the 0.32 deg width of the signal. This replicates the asymmetry of previous findings that, for a given signal location, crowding extends farther in the peripheral direction than in the central direction (Bouma 1970, 1973; Townsend et al., 1971; Banks et al., 1979;

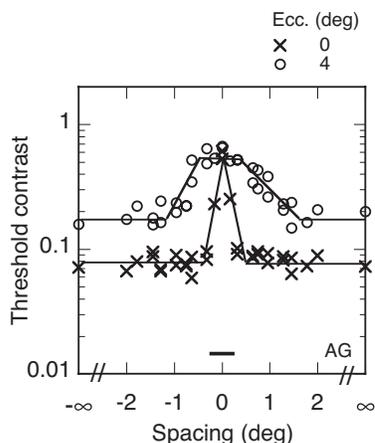


Figure 4. Fovea vs. periphery. Threshold for identifying a 0.32 deg Sloan letter (at 0 or 4 deg in right visual field) in the presence of a single flanker of the same size, as a function of spacing (i.e., flanker position). (Positive values are positions to the right of the signal. Negative values are to the left.) At 4 deg eccentricity, the critical region is 7 times wider than the letter (indicated by the horizontal bar). Observer AG. Not shown: similar results for observers MCP and MLL. Similar results at 6 deg eccentricity appear in Figure 8.

Chastain & Lawson, 1979; Wolford & Shum, 1980; Toet & Levi, 1992).

It may seem surprising that a more peripheral mask is more effective than a more central mask equally distant from the target. However, critical spacing is proportional to eccentricity, suggesting that the relevant cortical representation of visual space is progressively more compressed at greater eccentricity. Thus the more-eccentric mask is effectively closer than the less-eccentric mask (i.e., at a smaller fraction of the ever-increasing critical spacing).

Section 3.5 will show that a single flanker is much less effective than flankers on both sides (Bouma, 1970).

3.2 Effect of size

Ordinary masking would lead one to expect that the critical spacing in crowding would be proportional to signal size, not eccentricity. What is the effect of size on critical spacing? Levi, Klein, et al. (2002) and Levi, Hariharan, et al. (2002) found that, at 0 deg eccentricity, the critical spacing is proportional to size over a 50:1 range, but that, in the periphery, critical spacing is proportional to eccentricity, independent of size. We measured threshold contrast for letters of various sizes at 4 deg viewing eccentricity. Figure 5a shows threshold contrast as a function of spacing for letter sizes of 0.32, 0.5, 1, and 2 deg. For these sizes, threshold is elevated 26-fold (geometric mean). Figure 5b shows that the critical spacing did not change with letter size, instead remaining constant at about 1.2 deg, which replicates the Strasburger et al. (1991) finding, for numerals, that the spatial extent of crowding is 1.2 deg at 4 deg eccentricity, independent of size. Threshold elevation increases as a function of size (Figure 5c) because, as Figure 5a shows, the ceiling remains fixed at about 0.7 while the floor drops with size. This is just the familiar fact that contrast sensitivity for letters depends on size (see Pelli & Farell, 1999).

3.3 Effect of flanker size

We also measured the effect of mask size on critical spacing. We kept signal size at 0.32 deg and varied mask size from 0.32 to 3.2 deg. We didn't know what to expect. On the one hand, increasing the mask's size increases its contrast energy, which we thought might increase the mask's effect. (For a letter, contrast energy is the product of area and squared contrast.) On the other hand, enlarging the mask makes it less similar to the signal, which might lessen its effect. Surprisingly, Figure 6 shows that the threshold curves nearly superimpose, hardly affected by mask size, retaining a critical spacing of about 1.3 deg. (Figure 6a is for one observer; Figure 6b is for another.) Unlike ordinary masking, the crowding effect is not tuned to size. The range (spatial extent) of crowding is independent of signal size (Figure 5b) and mask size (Figure 6), depending solely on eccentricity (Figure 3b).

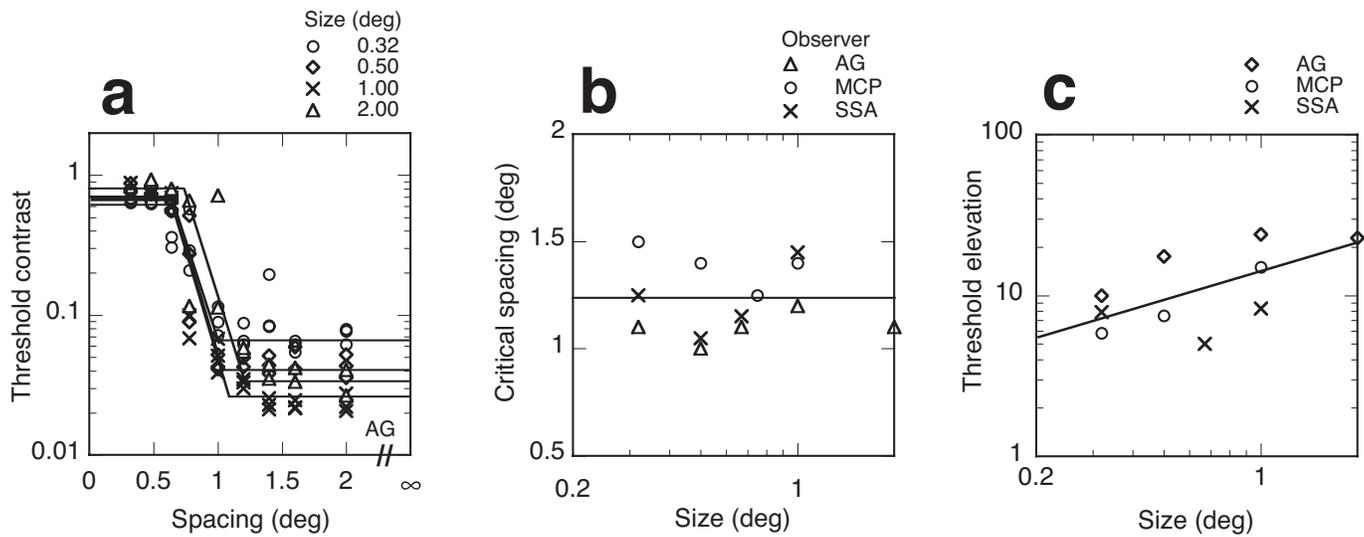


Figure 5. Effect of size. Identification of 0.32 - 2 deg Sloan letter (at 4 deg in the right visual field) between 2 flankers of the same size. (a). Threshold as a function of spacing. For observer AG, the threshold contrasts for all four sizes, 0.32, 0.5, 1, and 2 deg, nearly superimpose at spacings up to 1 deg. (b). Critical spacing vs. size, for three observers, showing no effect. Average (horizontal line) is 1.2 deg. (c). Threshold elevation increases somewhat with size: log-log slope of 0.6. This replicates the familiar finding that threshold contrast depends on size (Pelli & Farell, 1999).

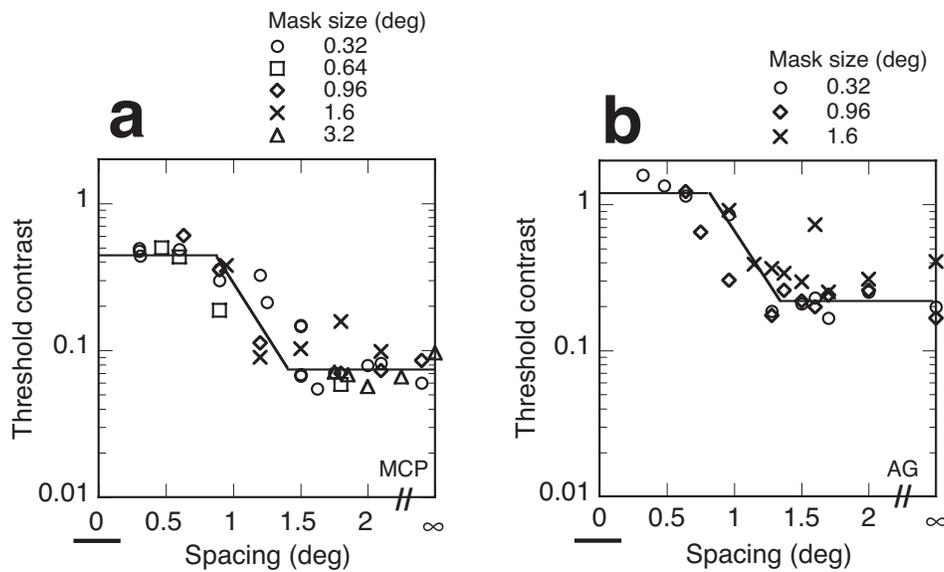


Figure 6. Effect of flanker size. Signal is 0.32 deg Sloan letter at 4 deg in the right visual field. The two flankers were 1 to 10 times the size of the signal. One fit was made to data for all mask sizes. (a). Critical spacing is 1.3 deg for MCP. (b). Critical spacing is 1.2 deg for AG. Horizontal line at the bottom left of the graph represents the width of the signal.

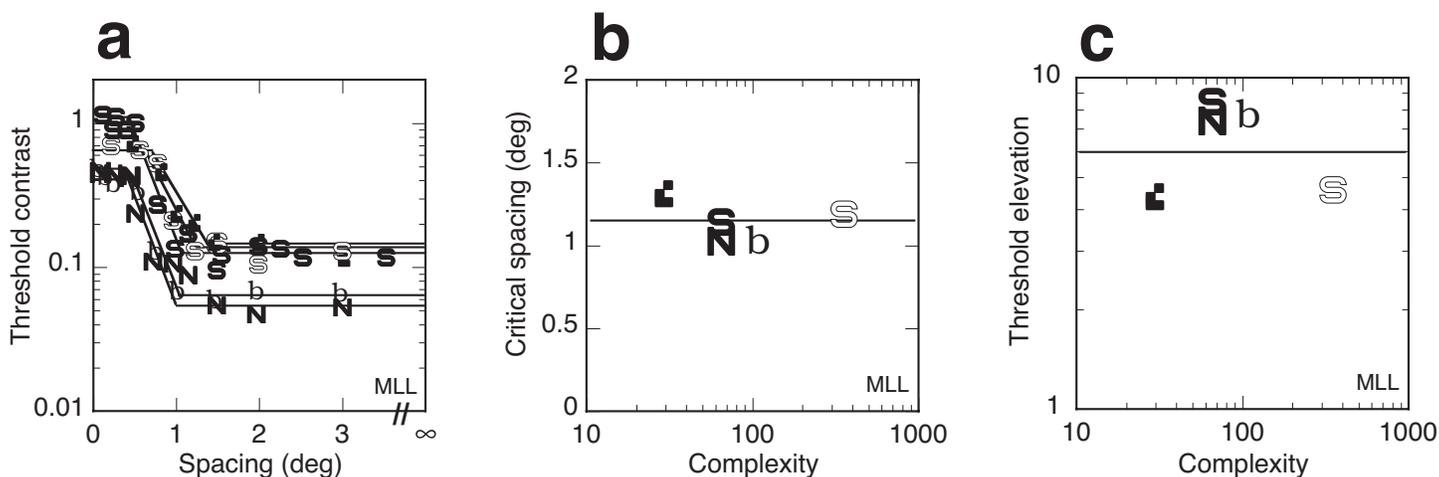


Figure 7. Effect of font. Signal is 0.32 deg letter at 4 deg in the right visual field. (a). Threshold contrast as a function of spacing for various fonts, Bookman **b**, 2×3 Checkers **■**, Sloan **S**, NZ Sloan **N**, and Outline Sloan **S**. (b). Critical spacing is independent of complexity. (c). Threshold elevation as a function of perimetric complexity, perimeter squared divided by area, showing that threshold elevation does not seem to be systematically related to complexity. Observer MLL. Not shown: similar results for observer MCP.

3.4 Effect of font

We wondered whether perimetric complexity (perimeter, squared, over ink area; Pelli et al., [in press](#)) or some other aspect of letter shape is important for crowding. [Figure 7a](#) shows threshold as a function of spacing for several fonts, including a meaningless alphabet of twenty-six 2×3 checkers (e.g., **■**) and another alphabet consisting of just two letters, **N** and **Z**, from the Sloan alphabet. These curves are quite similar to each other, differing from one another by large, but unimportant, vertical translations and small horizontal translations. The vertical shifts track the different threshold contrasts for different fonts, which is not of interest here. The small horizontal shifts are small differences in critical spacing, which ranged from 1 to 1.3. Pelli et al. ([in press](#)) showed that efficiency for letter identification is inversely proportional to perimetric complexity of the font, but complexity seems to be irrelevant to crowding. [Figures 7b](#) and [7c](#) plot critical spacing and threshold elevation as a function of complexity, showing no systematic effect of complexity.

3.5 Effect of number of flankers

Would adding more flankers increase the crowding effect? [Figure 8a](#) plots threshold for letter identification in the periphery with 1, 2, and 4 flankers. The signal and flankers are all Sloan letters, right-side up. [Figure 8b](#) shows that critical spacing is independent of number of flankers. It is about 0.4 of the eccentricity. [Figure 8c](#) shows that threshold elevation increased when flankers were increased from 1 to 2, but threshold was not further elevated when flankers were increased from 2 to 4. Consistent with this, Wilkinson et al. (1997) reported that reducing the number of flanking gratings from 14 down to 2 did not significantly

reduce their effect on the discriminability of the signal. Toet and Levi (1992) report extensive measurements of the effect of two T flankers on judging orientation of a T target, adding that, in pilot measurements, they found no effect of a single flanker. However, Strasburger et al. (1991) did report an increased threshold elevation when they increased the number of flankers from 2 to 4.

It makes sense that a single flanker would be much less effective than multiple flankers that surround the object. One imagines that when there is only one flanker the observer may use a large but offset integration field to pick off the exposed target. This strategy is not available when there are two or more flankers surrounding the target.

For a signal 6 deg to the right of fixation, we find a smaller critical spacing for flankers above and below, **S**
K
R, instead of left and right of the signal, **SKR**, which is consistent with Toet and Levi's (1992) finding that the critical spacing is smaller along the circumference than along a radial ray from the fovea.

3.6 Effect of flanker contrast

The experiments presented above used flankers of a high contrast, 0.85. [Figure 9a](#) shows threshold signal contrast as a function of spacing for several mask contrasts. [Figure 9b](#) shows that critical spacing is independent of mask contrast. There is an outlier, the × representing a critical spacing of 0.5 deg at a mask contrast of 0.1 for observer MLL. This is based on the fit shown in panel a to the 0.1 mask contrast data (solid diamonds). Note that threshold is elevated only when the mask overlaps the signal.

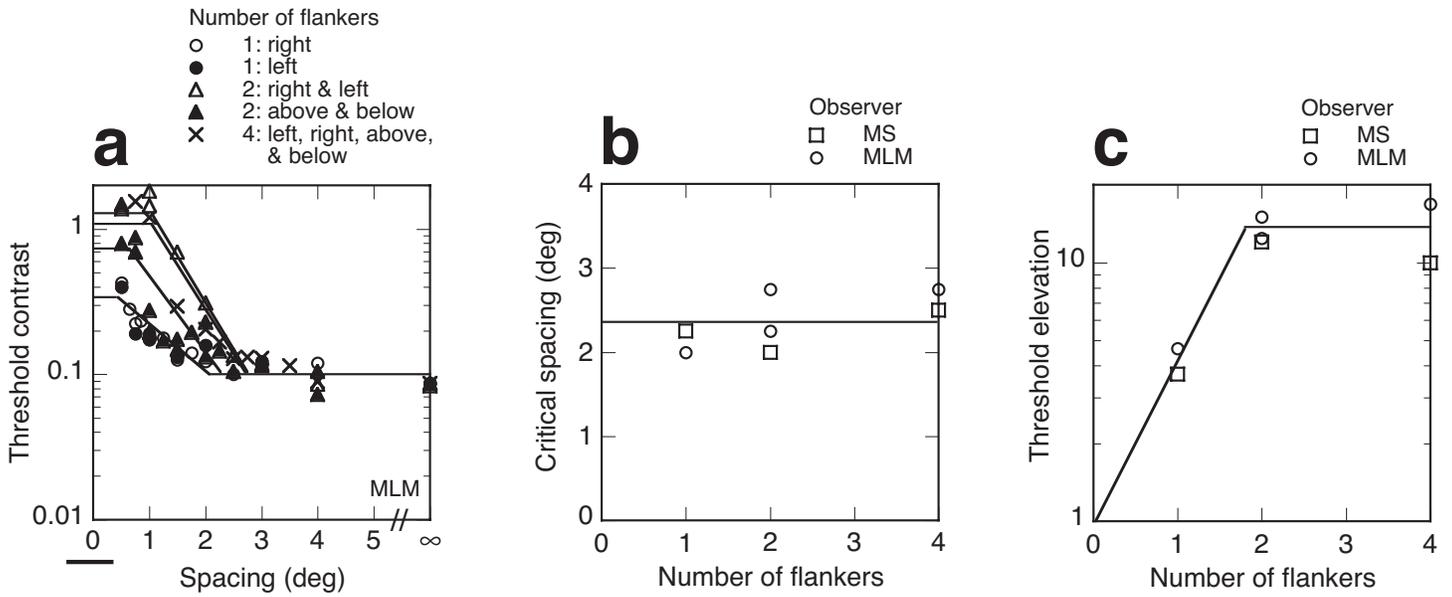


Figure 8. Effect of number of flankers. Signal is 0.32 deg Sloan letter at 6 deg in the right visual field. Flankers are letters, too, also right-side up, but displaced vertically or horizontally. (a). Threshold contrast as a function of spacing, for 1, 2, or 4 flankers. Flanker position (e.g., “right”) is relative to signal position. The horizontal line at the bottom left of the graph represents the width of the signal. Note that, lacking data at zero spacing (because the flankers would have collided), it is not clear whether there is a ceiling at small spacing, so that part of the clipped-line fit is somewhat arbitrary. The one-flanker data shown here, for a signal at 6 deg ecc., are similar to the data shown in Figure 4, for a signal at 4 deg ecc. We replicate the well-established finding that the critical spacing is greater in the peripheral than in the central direction. (b). Critical spacing (estimated separately for each condition, but averaging results for 1-left and 1-right) as a function of number of flankers. (c). Threshold elevation (estimated separately for each condition) as a function of number of flankers. This last graph is tentative because it depends on the somewhat arbitrary ceilings of the clipped-line fits in panel a. Observers MS and MLM. Not shown: similar results for observer MCP at 4 deg eccentricity.

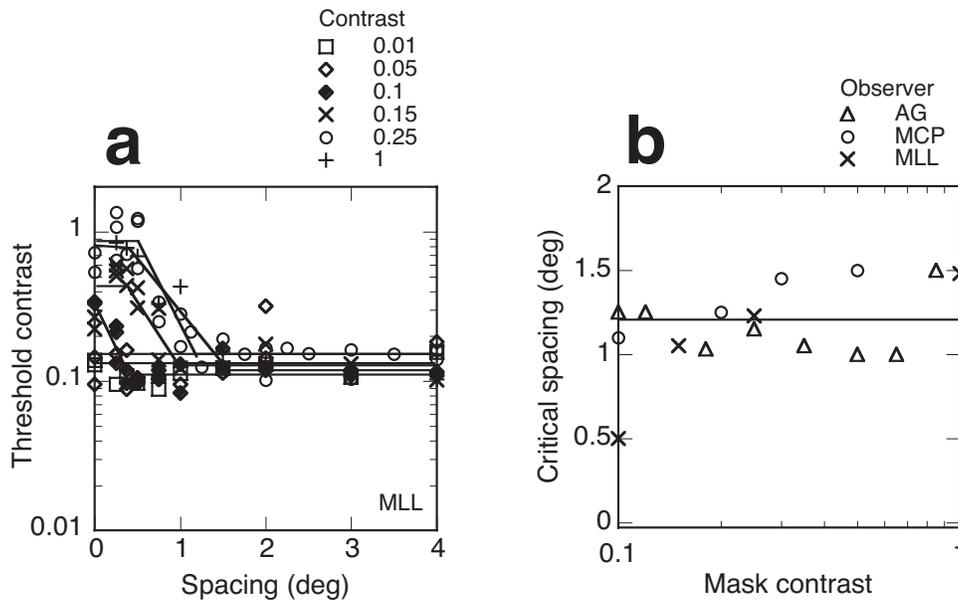


Figure 9. Effect of flanker contrast for three observers identifying a 0.32 deg Sloan letter at 4 deg in the right visual field. (a). Threshold contrast as a function of spacing for several flanker contrasts. (b). Critical spacing as a function of flanker contrast. Mask contrasts below 0.1 did not elevate threshold so they have no critical spacing. Observers MCP, AG, and MLL.

Thus the anomalous point in 9b seems to represent ordinary masking, not crowding. The rest of the data show no consistent effect of mask contrast on critical spacing: For one observer, critical spacing rises slightly with mask contrast, but it falls slightly for the other two observers. Fine (2003), too, reported crowding to be independent of contrast. So far, we have seen that critical spacing is independent of signal size, mask size, mask contrast, signal and mask font, and number of masks.

Figure 10 demonstrates the effect of mask contrast, showing an abrupt transition as mask contrast is increased. Once the mask becomes visible it soon saturates, producing its full effect on the signal.

Based solely on this demo, one might wonder whether the crowding is determined by similarity. The flankers become more similar to the signal as their contrast approaches that of the signal. However, Chung et al. (2001) manipulated signal and flanker contrast to test this hypothesis, finding that, at least in their conditions, more mask contrast always increased masking, even when this made the masks less similar to the signal.

In another view of the same data, Figure 11a shows threshold contrast as a function of mask contrast for several spacings. For a 0.32 deg letter, the contrast response curves show that threshold elevation increases abruptly with mask contrast, going from none to full effect as the mask goes from 0.1, the threshold contrast for identifying an isolated letter, to about 3 times that, saturating at higher contrast. There are two critical mask contrasts. In our clipped-line fit, *mask threshold* is the mask contrast at which threshold contrast of the signal begins to increase (edge of floor). And *mask saturation* is the mask contrast at which threshold contrast of the signal stops increasing (edge of ceiling).

This contrast-response curve is quite unlike what is usually seen in ordinary masking. Here the function rises steeply and hits a hard ceiling, with no further increase over a wide range of high mask contrasts (0.25 - 1). In ordinary masking, the function rises with a log-log slope of 0.5 to 1 and continues to increase relentlessly. The log-log slope of the (clipped line) contrast-response function for crowding is 2 at the closest spacing and falls exponentially with spacing (Figure 11c, right hand scale). The function found here is more reminiscent of the sigmoidal form of a frequency-of-seeing curve, rising suddenly from floor to ceiling over a narrow range of contrast. For comparison, Figure 11b shows the observer's proportion of correct identifications for an unflanked signal at this eccentricity as a function of signal contrast.

Chung et al. (2001) measured the contrast-response function for a bandpass-filtered letter among similar letters at a single separation (2.2 deg) at an eccentricity of 5 deg, obtaining shallow log-log slopes (0.3 and 0.1) that are consistent with the less than 0.4 slope found here at our maximum separation (1.5 deg at an eccentricity of 4 deg). Testing at such large (near-critical) separations (about 0.4 of eccentricity), the threshold elevation and slope are nearly gone.



Figure 10. Effect of flanker contrast. Starting at the top, in each row, fixate the black square, and try to identify the middle letter on the right. As you read down the chart, the contrast of the center letter is always 0.50, while the contrast of the two outer letters increases (0, 0.10, 0.15, 0.25, and 0.50). You'll find that the central letter becomes much harder to identify as soon as the flankers are at all visible.

The series of functions plotted in Figure 11a reveal something quite remarkable. It is hardly surprising that the threshold elevation (on the vertical scale) is reduced at greater spacings, as shown in Figure 11c. But we were surprised to find that the critical mask contrasts (0.1 and 0.25 on the horizontal scale) are unaffected by the spacing. In

Figure 11a every curve (one for each spacing) turns up at a mask contrast of 0.1 and saturates when the mask contrast reaches 0.25, no matter how far away the signal is. Figure 11d shows explicitly that the critical mask contrasts are independent of spacing. We will come back to this in Discussion.

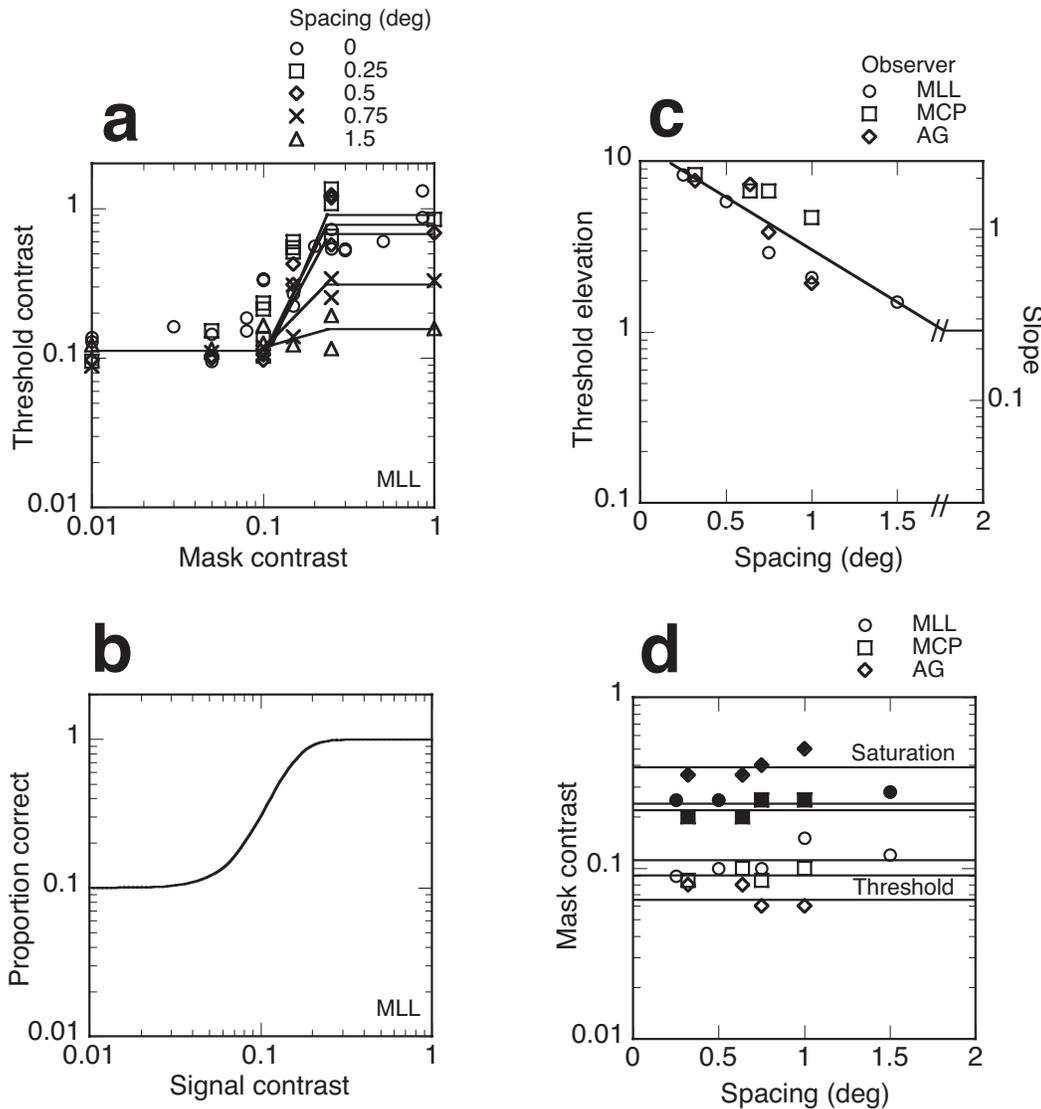


Figure 11. Effect of flanker contrast for three observers identifying a 0.32 deg Sloan letter at 4 deg in the right visual field. Same data as Figure 9. (a). Threshold contrast for identifying the target letter as a function of mask contrast for observer MLL. Clipped lines (shown) are fit to the (roughly sigmoidal) data, constrained to have equal threshold and saturation contrasts of the mask for all conditions. Threshold contrast rises at 0.1 mask contrast and saturates at 0.25 mask contrast for all spacings. Clipped lines (not shown) were also fit independently to the data for each condition for each observer, and the parameters of these fits are plotted in panel d. (b). Psychometric function. Proportion correct identification of a letter as a function of contrast. The knees (critical contrasts) of this psychometric function roughly match those of the contrast response function in panel a. This is a maximum likelihood fit of a Weibull function to the measured proportion correct (not shown) at several contrasts (see Pelli et al., in press; Strasburger, 2001). The lower asymptote is 1/10 because that is the chance of correctly guessing the identity of one of 10 letters. (c). The threshold elevation (left scale of panel c) and log-log slope (right scale) of the fits in panel a (and similar data for observers MCP and AG) are high at small spacings and fall exponentially with increased spacing. (d). Threshold and saturation contrasts of the mask as a function of spacing. Mask threshold is the first knee, where the signal threshold rises. Mask saturation is the second knee, where the signal threshold saturates. Each pair of points (solid and open) is based on an independent clipped-line fit (not shown) to the data for one condition and observer. The threshold contrast for identifying the mask may be estimated from that for the signal (0.1) at low (0.01) mask contrast (panel a). Observers MCP, AG, and MLL.

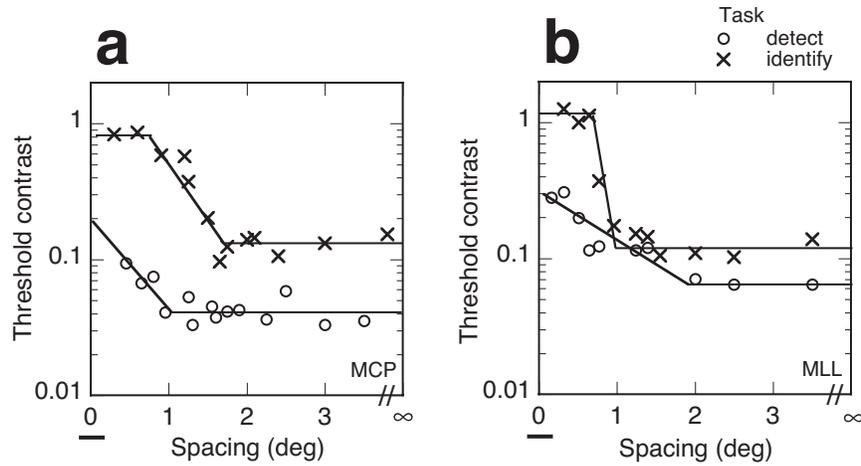


Figure 12. Effect of task. Identification or detection of a letter between letter flankers. Signal is 0.32 deg Sloan letter at 4 deg in the right visual field. Threshold curves for identifying a letter are sigmoidal with an average threshold elevation of about 1 log unit and a critical spacing of 1.2 deg. (a). Observer MCP. (b). Observer MLL. There are some observer differences, but detection threshold is always lower, with less threshold elevation. The average critical spacing for detection is 1.5 deg. Horizontal line at the bottom left of the graph represents the width of the signal. Figure 16a shows similar results for a third observer.

3.7 Effect of task: identification and detection of letters and gratings

Most crowding studies have used identification tasks, whereas most masking studies have used detection tasks. To determine whether crowding depends on task, Figure 12 shows identification and detection thresholds for a letter among letters as a function of flanker spacing for two observers. (Figure 16a shows similar results for a third observer.) For identification, averaging across the three observers, the threshold elevation is large (ten-fold) and extends out to 1.3 deg (four signal widths). For detection, the

threshold elevation is only three-fold but extends about as far (average is 1.5 deg).

To distinguish crowding from masking, we assessed the effect of eccentricity and size on critical spacing.

We measured the effect of eccentricity (2, 4, and 8 deg in right visual field) on detection thresholds for 0.75 deg Sloan letters. Figure 13a plots threshold as a function of spacing for each eccentricity. Figure 13b shows that the critical spacing for detection is independent of eccentricity, unlike the proportionality found for identification (Figure 3b).

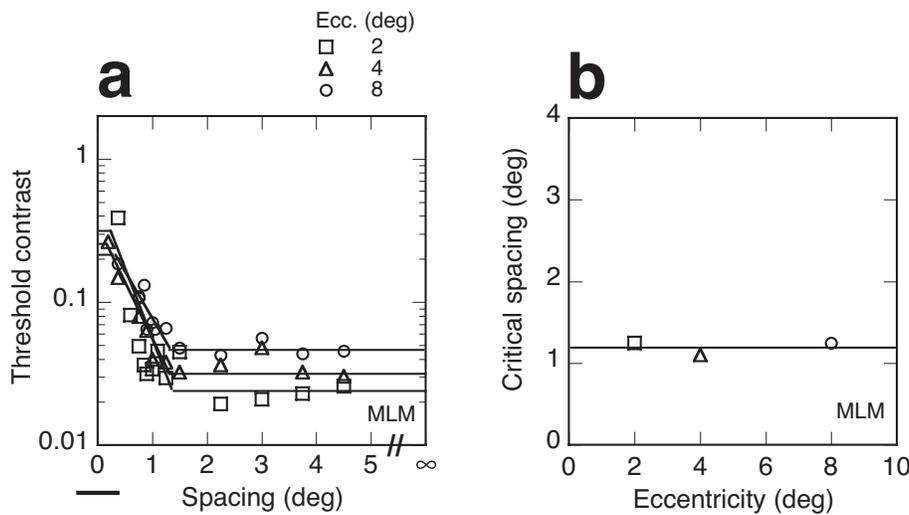


Figure 13. Effect of eccentricity on detection. Detection of a letter among letters at several eccentricities in the right visual field. Signal and flankers are 0.75 deg Sloan. (a). Threshold as a function of spacing. (b). Critical spacing as a function of eccentricity. The critical spacing for letter detection is independent of eccentricity. This is characteristic of ordinary masking, whereas in crowding the critical spacing is proportional to eccentricity, as in Figure 3b. Observer MLM.

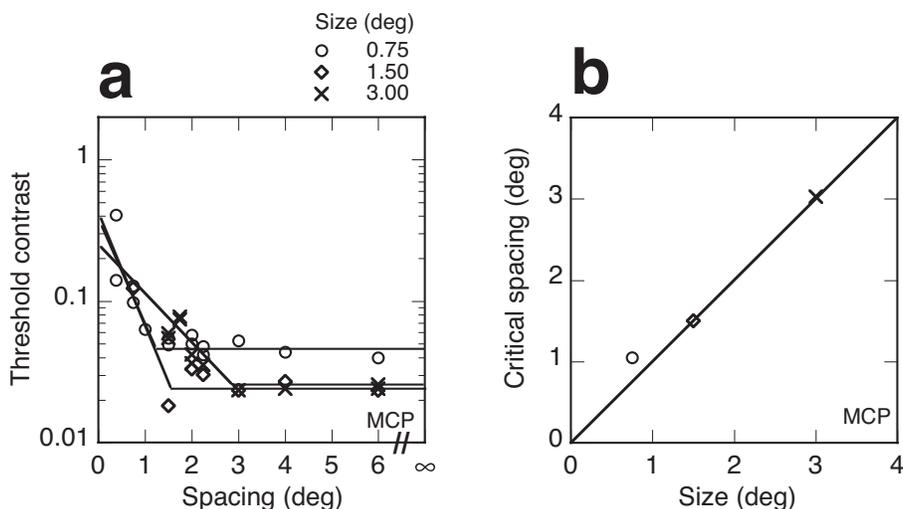


Figure 14. Effect of size on detection. Detection of a letter among letters at several sizes. Signal and flankers have equal size. Signal is Sloan letter at 8 deg in the right visual field. (a). Threshold as a function of spacing. (b). Critical spacing as a function of size. The critical spacing for letter detection is proportional to letter size. This is characteristic of ordinary masking, whereas in crowding the critical spacing is independent of size, as in Figure 5b. Observer MCP. Not shown: similar results for observer MLL.

We measured detection thresholds for Sloan letters of three sizes, 0.75, 1.5, and 3 deg, at 8 deg in the right visual field. The results in Figure 14b show that the critical spacing for letter detection is proportional to size; critical spacing for letter identification is independent of size (Figure 5b).

To us, this is the most telling difference: In ordinary masking (e.g. letter detection), the critical spacing is proportional to signal size (Figure 14b), independent of eccentricity (13b), whereas in crowding (e.g., letter identification)

the critical spacing is proportional to eccentricity (3b), independent of size (5b).

We also changed the envelope size of 1 c/deg gratings in a $\pm 45^\circ$ orientation discrimination task at 20 deg viewing eccentricity (Figure 15a). We saw earlier (Figure 5b) that changing the size of letters did not affect critical spacing. However, for gratings, the critical spacing scales with the size of the envelope (Figure 15b). There is no mystery here: The gratings mask each other only when they overlap; at their critical spacing they are abutting.

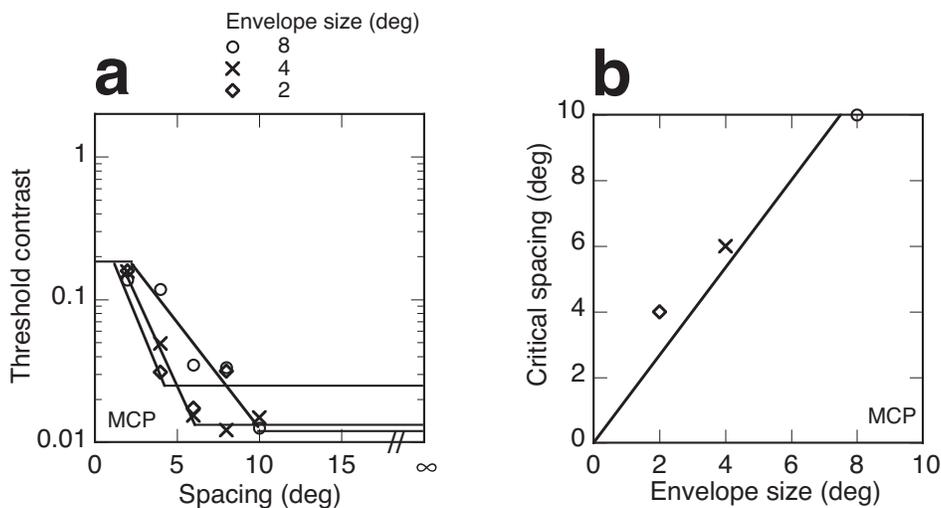


Figure 15. Effect of grating extent. Size is the 1/e radius of the Gaussian envelope. Grating (at 20 deg in the right visual field) flanked by two gratings. (a). Threshold contrast for identification of $\pm 45^\circ$ orientation of 1 c/deg grating between two flanking gratings as a function of spacing, for several envelope sizes. Signal and flanking gratings had same spatial frequency and same size envelope. (b). Critical spacing as a function of envelope size. Observer MCP. Not shown: similar results for observer AG.

3.8 Effect of letter vs. grating

Here we tried letters (0.32 deg Sloan) and gratings (8 c/deg) in every combination of target and flanker. Majaj et al. (2002) show that identification of letters is mediated by a channel with a center frequency determined by the stroke frequency of the letter. For a 0.32 deg Sloan letter the stroke frequency is $1.6/0.32 = 5$ c/deg, and, by their formula, the channel frequency is 6.3 c/deg, which is very close to the 8 c/deg spatial frequency of the grating we used. Thus the identification of letter and grating in this experiment was mediated by channels tuned to similar spatial frequencies.

We measured thresholds for detection and identification of 8 c/deg sinewave gratings. Signal and flanker gratings were each randomly tilted $\pm 45^\circ$ on each trial. In the detection task the observer was required to choose which of

two intervals contained the signal grating (ignoring orientation). In the identification task there was only one interval and the observer was asked, on the response screen, to identify its $+45^\circ$ or -45° orientation.

Figures 16c and 16d show that neither grating nor letter flankers raised the grating signal's threshold unless they overlapped it. (Letter size is 0.32 deg; grating size is 0.52 deg; see Table 1.) Grating threshold elevation at all spacings is similar for both tasks (detection and identification) and flanker types (letter and grating). Compared with identifying a letter among letters, the grating curves show no ceiling and have a small critical spacing (about one signal width). The grating's narrow critical spacing – threshold is elevated only when the flanker overlaps the grating – suggests ordinary masking, not crowding.

When we originally got the grating results reported in Figures 16c and 16d, we were led to think, wrongly as it

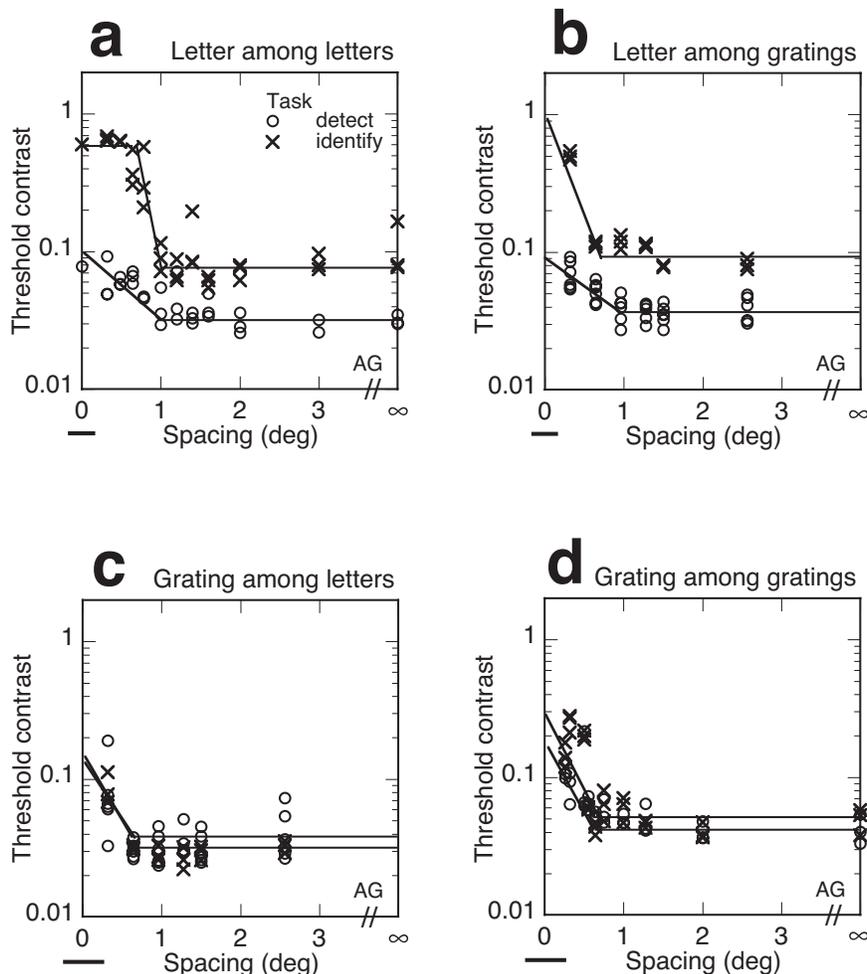


Figure 16. Effect of letter vs. grating. Identification or detection of a letter or grating flanked by letters or gratings. Signal at 4 deg in the right visual field. (a). Letter flanked by letters. Averaging across Figures 12ab and 16a, the critical spacing is about 5 letter widths (1.5 deg) for both identification and detection. (b). Letter flanked by gratings. Critical spacing is 2 times the width of the signal for identification, and 5 times the width of the signal for detection. (c). Grating flanked by letters. (d). Grating flanked by gratings. The results show that threshold is elevated only when the flankers overlap the signal. Sloan letters were 0.32 deg wide and sinewave gratings were 8 c/deg with a 0.52 deg Gaussian window (radius at 1/e). Horizontal line at lower left corner represents the width of the signal. There were always two flankers, to the right and left of the signal. Observer AG. Not shown: similar results for observers MCP and MLL.

turns out, that gratings are immune to crowding. Our identification task was too coarse. We had asked the observer to distinguish orientations 90° apart. Ordinary masking studies have shown that we see gratings by means of feature detectors that have an orientation bandwidth of ±15° to ±30° (Phillips & Wilson, 1984). Thus orthogonal gratings are detected by distinct feature detectors, and we would expect the label of a single feature detector to suffice for identifying the coarse orientation of the grating. Indeed, Thomas and Gille (1979) reported that two gratings differing in orientation by 20° to 30° are identified just as accurately as they are detected. And the thresholds for detection and identification in Figures 16c and 16d seem to be identical. This is the logic that Watson and Robson (1981) applied to frequency identification. When the two signals stimulated different detectors, observers could identify at the threshold for detection. When the same feature detector picks up both signals, then the observer cannot identify based on a single feature detection and requires at least two detectors to be active. The ratio of the detector responses would presumably be a good basis for fine discrimination. (Treisman, 1991, makes the same point for other stimulus dimensions.) Thus a parametric change in the task, from a coarse (>2:1) to a fine (<2:1) frequency discrimination, results in a qualitative change in the observer’s computational algorithm, from single- to multi-feature detection and integration (also see Verghese & Nakayama, 1994, and Discussion, Section 4.3).

4. Discussion

Our seven theoretical conclusions about the difference between crowding and ordinary masking are listed in Table 3 and discussed in Sections 4.1 – 4.7. The discussion of illusory conjunctions comes last (Section 4.7), but its only prerequisite is the vocabulary established in the Introduction (Section 1). We begin the discussion by proposing a definition.

Using published and new results, we have established that the original crowding phenomenon – impaired identification of a letter among letters in the periphery – is unlike ordinary masking. We suggest that the term “crowding” be applied to any phenomenon that exhibits the critical-spacing dependence reported by Bouma (1970).

When defining a term already in use, the desire to sharpen must be tempered by the need to respect established usage. Crowding was discovered in the course of measuring letter acuity in patients with central field loss (Korte, 1923) or amblyopia (Ehlers, 1936).⁴ Stuart and Burian (1962) coined the term “crowding” for the impairment of identification of a peripheral letter by neighboring letters. Since then the term has been used primarily, but not exclusively, to refer to lateral masking of letters by letters. Most writings on crowding – and this manuscript is no exception – grant center stage to Bouma’s (1970) still-amazing discovery, reported in a two-page letter to *Nature*, that critical spacing is roughly half the eccentricity. Our proposed definition, below, looks for Bouma’s finding as

Theory	Ordinary masking	Crowding	Facts (Table 2)	Section
a	Critical spacing is proportional to size and independent of eccentricity.	Critical spacing is proportional to eccentricity (Bouma, 1970) and independent of size (Strasburger et al., 1991; Levi, Hariharan, et al., 2002).	f	4.1
b	Occurs for any task.	<i>Specific to tasks that could not be performed based on a single detection by coarsely coded feature detectors.</i>	b - d	4.2, 4.3
c	Same feature detector mediates the effects of mask and signal.	<i>Distinct feature detectors mediate the effects of mask and signal.</i>	g - i	4.4
d	Eccentricity doesn’t matter.	In the periphery, the observer uses an inappropriately large integration field <i>because smaller fields are absent.</i>	a - i	4.5
e	Impairs feature detection.	Impairs feature integration (Flom, Weymouth, et al., 1963; Wolford & Shum, 1980; He et al., 1996; Parkes et al., 2001; Chung et al., 2001; Levi, Hariharan, et al., 2002).	a - i	4.6
f	Selectivity is that of the feature detector.	<i>Selectivity is that of the feature integrator.</i>	g	4.6
g	No signal feature is detected, so the signal is invisible.	Features of both signal and mask are detected and combined, so the signal is visible, but jumbled with the mask (Korte, 1923; Wolford & Shum, 1980; Parkes et al., 2001; Levi, Hariharan, et al., 2002).	b - d	4.7

Table 3. Theory: summary of the differences between crowding and ordinary masking. We cite the authors of existing theories about crowding, and *italicize our new ideas*. (b). Treisman (1991) makes a similar suggestion for illusory conjunctions. (c). This idea is implicit in the models that Wolford and Shum (1980), Treisman and Schmidt (1982), Wilkinson et al. (1997), and Parkes et al. (2001) use to explain their results. (f). Current feature detector models have several receptive fields, to implement divisive inhibition, but the differences in selectivity of these various fields are too small to matter here. (g). Treisman and Schmidt (1982) make a similar suggestion for illusory conjunction.

the signature of crowding, assessed by a practical test applicable to any object and task, not just letter identification.

4.1 A diagnostic test for crowding

When observing a new masking phenomenon, one may wish to classify it as ordinary masking, crowding, or neither. First, when the mask obscures the signal's identity: in crowding the signal is perceived as an ambiguous amalgam; in ordinary masking the signal disappears. Second, the definitive criterion is that *the critical spacing for crowding scales with eccentricity, independent of signal size, whereas for ordinary masking it scales with signal size, independent of eccentricity.*

When away from our own labs we usually cannot make arbitrary changes to stimuli, but we usually can control our viewing distance and fixation to control image size and eccentricity. If the phenomenon is ordinary masking, then changing eccentricity by fixating near or far from the signal should make no difference to the effect. If it is crowding, then the effect should occur when fixating far away, and cease when fixating closer than twice the center-to-center spacing between signal and mask.

If fixation is maintained on a point in the image as we approach/enlarge it, neither ordinary masking nor crowding will be affected. Signal size and eccentricity will both increase proportionally with spacing of signal and mask. Any phenomenon that depends on viewing distance (within the limits of visibility) is neither ordinary masking nor crowding.

4.2 What crowds?

Over the past twenty years, some investigators have tentatively applied the term “crowding” more broadly than Stuart and Burian (1962) did, to describe the effect of laterally displaced masks on other kinds of target, such as the orientation discrimination of a grating among gratings (He, Cavanagh, & Intriligator, 1996; Parkes et al., 2001) and vernier acuity (Levi & Klein, 1983, 1985; Levi et al., 1985) and Landolt C acuity (Hess, Dakin, Kapoor, & Tewfik, 2000) among flanking bars. However, only some of these results seem to describe the same kind of interference as found for identifying a letter among letters in the periphery. As discussed below, the grating orientation task satisfies at least part of our test for crowding, but we suspect that the masking of vernier and Landolt C acuity by bars is not crowding. (The key tests have not been done.) For letters flanked by letters – the prototype for crowding – no facilitation is found at any spacing, and the masks lose their effect only at far spacing, exceeding half the eccentricity (Bouma, 1970; Toet & Levi, 1992). The flanking bars in Landolt C and vernier tasks mask the signal only at intermediate spacing (5 times the gap width of the C), losing effectiveness at close and far spacing (Flom, Weymouth, et al., 1963; Flom, 1991; Jacobs, 1979; Levi et al., 1985). The lateral masking in the vernier and Landolt C tasks may be due to direct stimulation of the same feature detector by

signal and mask, as in ordinary masking. In these two tasks the discriminandum is tiny – a thin dipole for vernier or the square that would fill the gap in the Landolt C – but one would expect it to be detected by a feature detector with a larger receptive field, perhaps comparable in scale to the “perceptive field” defined by the measured masking function (Klein, Casson, & Carney, 1990; Levi et al., 1985).

Similarly, Polat and Sagi (1993) found for grating detection that flanking gratings impair detection of the target grating at close spacing, facilitate at intermediate spacing (three periods of the grating), and lose their effect at far spacing. This may or may not be ordinary masking (Solomon & Morgan, 2000; Sagi, 1990), but it seems unlike crowding and we suspect it will fail our test (Section 4.1).

Besides identification of a letter among letters, what other conditions yield crowding? In Figures 15 and 16 the task was identification (i.e., discrimination) of a grating's $\pm 45^\circ$ orientation. Neighboring gratings produced ordinary masking, not crowding. However, this result, for *coarse* discrimination, is unlike the findings of Wilkinson et al. (1997) for several kinds of *fine* discrimination. They report the effect of eccentricity at fixed signal size, finding threshold elevation for fine contrast discrimination only when the signal eccentricity is more than twice the spacing to the nearest flanker. They estimate a critical spacing of 0.4 eccentricity for fine contrast and spatial-frequency discrimination, and slightly higher for orientation discrimination. Similarly, the lateral masking of gratings by gratings demonstrated in the periphery by He et al. (1996) and Parkes et al. (2001) seems to be crowding: Observers still see the central grating, but its apparent orientation is an amalgam of all the grating orientations.

In summary, the effect of a flanking bar on vernier and Landolt C targets does not seem to be crowding. A letter among letters and a grating among gratings can yield either crowding or ordinary masking, depending on the task. Detection and coarse grating discrimination tasks yield ordinary masking. Letter identification and fine grating discrimination tasks yield crowding.

The difference between fine and coarse discrimination suggests that the coarse discriminations are like detection in that the observer can respond correctly based on a single internal detection event (a unit tuned to $0^\circ \pm 30^\circ$ orientation), whereas a fine discrimination response would require integration over several detection events.

4.3 Task-specific: no crowding of detection

Both detection and identification of a letter are impaired by the presence of neighboring letters. The critical spacing for detecting a letter among letters can be as large as that for identification, but we call it ordinary masking, not crowding, because it scales with letter size, not eccentricity (Figures 13 and 14).

As we said at the end of Section 4.2, our limited survey of tasks distinguishes a few that are susceptible to crowding

(letter identification and fine discrimination of orientation, contrast, or frequency) from other tasks that seem to be immune, though they await definitive testing. So far, this difference corresponds to the dichotomy between tasks that can be performed based on a single coarsely coded feature detection and those that cannot. One-feature-detection-event tasks (like detecting a grating or a letter, or reporting its coarse orientation) are unaffected by crowding, because the signal triggers feature detections as usual, despite the flanker. If we accept the feature detection model, because of its excellent empirical support (e.g., Campbell & Robson, 1968; Graham, 1980, 1989; Watson & Robson, 1981), and the evidence that letter identification is mediated by feature detection (Pelli et al., in press; Solomon & Pelli, 1994; Majaj et al., 2002), then there must be a second stage that assembles the detected features to identify the letter. Similarly, if we suppose coarse coding of orientation among the feature detectors, then decisions that achieve fine discrimination must be based on more than a single feature detection, as Watson and Robson (1981) noted for frequency coding (Verghese & Nakayama, 1994). Our new dichotomy merges two old dichotomies. Watson and Robson (and, later, Treisman, 1991) suggested that a single feature detector's label suffices for coarse discrimination, allowing identification of sufficiently different signals at detection threshold. Treisman and Gelade (1980) suggested, based on their Feature Integration Theory, that detecting a conjunction of two features would be qualitatively different from detecting a single feature.

4.4 Mask and signal detected separately

To characterize the computation that is impaired by crowding, perhaps the most revealing of our results is the effect of mask contrast. Threshold elevation of the signal represents an interaction of the mask and signal, so it is hardly surprising that it falls, exponentially, with increased mask-signal spacing. However, we were surprised to discover that the threshold and saturation contrasts of the mask, to affect the signal, are independent of spacing (Figure 11d, Table 2i). If we supposed that the mask directly impaired the sensor (feature detector) mediating identification of the signal, then this result would be an exception to the usual finding in visual psychophysics and physiology that sensitivity wanes with distance in the visual field between the stimulus and sensor. Rejecting that implausibility, we conclude, instead, that the fixed threshold and saturation contrasts of the mask are determined not at the variously distant sensor that detects the signal, but, instead, at a sensor local to the mask. In other words, the effects of signal and mask are mediated by separate feature detections.

The contrast response functions are extremely nonlinear, with a graded response only in the narrow 3:1 interval between mask threshold and saturation (Figure 11a; Table 2h). It looks like a frequency of seeing curve, the increasing probability of an all-or-none event (like saying "yes"). The

local nonlinear processing of the mask, responsible for the mask's threshold and saturation, could be a probabilistic all-or-none event. The graded increase in signal threshold between these two mask contrasts could represent averaging across trials, which individually had or lacked the all-or-none event. It seems reasonable to call the all-or-none event "feature detection." The threshold elevation produced by the mask feature detection on the signal depends on spacing, falling exponentially, with a space constant proportional to eccentricity. For comparison, Figure 11b shows the proportion correct identification of an isolated letter. The curves in Figures 11a and 11b are remarkably similar in shape and position.

The minimal summation (in threshold elevation) among feature detections (Table 2g) is surprising, but consistent with this model. The absence of summation showed up in two ways: no systematic effect of mask complexity (presumed to be proportional to number of features; Figure 7c) and no effect of number of masks, beyond two (Figure 8c).

4.5 Integration field

Why does the visual system do something so silly as to integrate the features of a remote flanker into the signal letter? People rarely experience crowding in the fovea, so integration must normally have the right range for an object in the fovea, integrating over the entire region of the object, and not beyond.⁵ So why extend it perniciously for signals in the periphery? Our guess is that the visual system has many *integration fields* of various sizes, overlapping one another, and distributed across the visual field. When possible, the visual system uses an integration field of the same size and location as the object to be identified, and this is what normally happens in the fovea. But in the periphery we lack small integration fields, so we use what we have, which may be inappropriately large. The large ones are cheap, because it takes only a few to tile the visual field. Smaller ones are progressively more expensive, because tiling requires more of them, so they exist only in the central visual field. (Allowing overlap in the tiling will increase the number of fields by the overlap factor, without changing the argument.)

Thus it seems that the observer doing a simple detection task can elect to monitor many feature detectors and base a response on one. Alternatively, if a more complex judgment is required, the observer may monitor the output of a feature integrator whose integration field has a minimum size defined by the critical spacing, allowing a response based on a combination of the features in that integration field. This may seem at once familiar and fanciful.

It is familiar because it overlaps with popular existing theories, specifically feature detection (Graham, 1980) and feature integration (Treisman & Gelade, 1980). It differs from Treisman (and He et al., 1996) in attributing the peripheral deficit to an absence of small integration fields rather than a lack of "focal attention." Our suggestion that

crowding results from *absence* (of small integration fields) also differs from past explanations that attributed crowding to the *presence* of perturbing or inhibiting mechanisms in the periphery (e.g., Wolford, 1975; Andriessen & Bouma, 1976; Wilkinson et al., 1997; Hazeltine, Prinzmetal, & Elliott, 1997).

It might seem fanciful to make strong assertions about “integration fields” that are so vaguely specified. But the data support us well here. The “integration field” is just a name for the area circumscribed by the measured critical spacing around the signal (Toet & Levi, 1992). (Levi et al., 1985, called it the “perceptive field” or “perceptive hypercolumn”.) That this area is determined by signal eccentricity, independent of signal and mask size, seems to warrant calling it a “field.” Yet this field is of another kind from that of a receptive field used to detect a feature (see Section 1). Receptive fields detect features. Such feature-detection events are subsequently combined by integration fields.

A grating crowded by other gratings, so that the observer cannot accurately report its orientation, is nevertheless fully effective in producing orientation-specific adaptation (He et al., 1996). The apparent orientation is affected by the neighboring gratings but the orientation at which sensitivity is most reduced is unaffected by the neighboring gratings, indicating that the adaptation occurs at a site earlier in the cortical processing than crowding. This suggests that adaptation occurs before or within the feature detector and that crowding occurs after the feature detector and before or within the feature integrator.

One wonders what the computational capabilities of this integrator might be. Parkes et al. (2001) show that in their task the integrator computes average orientation over the integration field. In our task the integrator computes letter identity, or a precursor to that.

It is remarkable that all the tasks that exhibit crowding yield similar estimates of the critical spacing, even though the computational demands of the tasks (e.g., discriminating letter shape and grating orientation, are very different). In Section 1.4 we demonstrated that the various perceived properties of the same object may be based on regions of very different size. It seems that the size of the region used to assess an object property depends crucially on whether it requires feature integration, and can be independent of the specific object (letter or grating) and task (shape identification or fine discrimination).

Dichoptic presentation of the mask to one eye and the signal to the other shows that crowding is a cortical, not a retinal phenomenon (Flom, Heath, & Takahashi, 1963; Taylor & Brown, 1972; Tripathy & Levi, 1994). As eccentricity increases, each square degree of visual field is represented by fewer cortical neurons. Tripathy and Levi (1994) note that since the critical spacing of crowding is proportional to eccentricity (i.e., half), it corresponds to a constant number of millimeters at the cortex, which roughly matches the length of horizontal connections in V1 (Gilbert, Ito, Kapadia, & Westheimer, 2000). However, the half-the-eccentricity critical spacing is just as good a match

to the receptive field radius of V4 neurons (Desimone & Schein, 1987; Desimone, Schein, Moran, & Ungerleider, 1985; Piñon, Gattass, & Sousa, 1998; Motter, 1994a, 1994b, 2002). Thus, crowding occurs somewhere in the visual cortex, but it’s hard to be more precise than that.

The reader, like us, may be wishing that we could say more precisely what the feature integrator does. How are features bound to recognize an object? We do not know, but the data are very clear on the crucial role played by the integration field in excluding what is outside it. Perhaps it should be called an “isolation field.”

4.6 Selectivity

It seems that ordinary masking has the selectivity of a feature detector, and that crowding has the selectivity of a feature integrator. The spatiotemporal selectivity of ordinary masking is more or less consistent with that of a receptive field (or a few similar receptive fields; Foley & Chen, 1999). The selectivity of crowding seems to be broader in many ways; crowding is equally effective over a wide range of flanker type (letter, black square; Loomis, 1978), flanker size (10:1; Figure 6), and flanker number (≥ 2 ; Figure 8). For letter identification, Loomis (1978) found identical crowding by flanking letters or black squares (at same spacing), and little or no effect of square size. Chastain (1981) and Nazir (1992) found progressively more effect of surrounding flankers that were more similar to the target letter. Banks et al. (1979), Chastain (1983), and Baylis and Driver (1992) found lessened interference of letter identification by flanking letters that are grouped apart (e.g., by color or contrast polarity). Kooi et al. (1994) measured crowding of a T among Ts, all randomly oriented (0° , 90° , 180° , or 270°). Crowding was lessened when the flankers differed from the signal in contrast polarity, color (for most observers), or depth. However, the general rule that more similar flankers produce more crowding has two exceptions. Higher-contrast flankers are more effective than equal-contrast flankers (Chung et al., 2002). A target surrounded by identical flankers appears normal, as shown for gratings by Parkes et al. (2001) and for letters here (Section 1.5).

When we mask a letter with noise, which produces ordinary masking, thresholds for detection and identification are increased by the same factor (Pelli et al., *in press*). Majaj et al. (2002) did critical-band noise masking of letters and gratings, finding that we use the same spatial-frequency channel (feature detector) for detection and identification. This is ordinary masking. The selectivity revealed is that of the feature detector.

Chung et al. (2001) used bandpass filtered letters to measure the critical band for crowding. They measured threshold elevation of the signal letter as a function of center frequency of the mask letter, for a range of center frequencies of the signal. They were surprised to find that their results are similar to earlier results for ordinary masking of grating by gratings (measured bandwidth was about an octave broader, as one would expect from the one-octave

bandwidth of their mask). This is quite different from the absence of size tuning for unfiltered letters (tested over a 10:1 range) shown in Figure 6. The experiments differ in important ways, and it is hard to know which differences account for the qualitative difference in selectivity of masking. First, their letter size (before filtering) was fixed as they varied the frequency of the passband, whereas we varied letter size. So they varied frequency, not size, whereas we varied frequency and size together. Second, their letters were filtered, so their spatial envelopes fell off gradually, unlike the abrupt edges of our unfiltered letters. Majaj et al. (2002) find a qualitative difference in ordinary masking for filtered (soft-edged) versus unfiltered (hard edged) letters. It is conceivable that the Chung et al. (2001) critical band masking results reflect ordinary masking, not crowding. Admittedly, their contrast-response results (at one size, 2.2 deg, and spacing, 2.2 deg) are consistent with our results for crowding, as noted in Section 3.7. However, we also found ordinary masking for letter detection at approximately that size and spacing (Figure 14). (This was at a larger eccentricity, but eccentricity has no effect on ordinary masking, Figure 13.) One would like to know whether the critical spacing for filtered letters is proportional to eccentricity (as in crowding) or to letter size (as in ordinary masking).

Further measurements of the selectivity of crowding could help characterize the computation performed by the feature integrator.

4.7 Illusory conjunctions

Treisman and Schmidt (1982) coined the term “illusory conjunction” to describe the impression of seeing a non-existent object that combines features from several real objects in the stimulus (for antecedents, see Snyder, 1972; Wolford, 1975; Estes, Allmeyer, & Reder, 1976; Bjork & Murray, 1977; Krumhansl & Thomas, 1977). This strikes us as a good description of the effect of crowding. Inappropriate combination of signal and mask features would explain why letter identification (in which the observer presumably must use several features) is affected, while letter detection is unaffected. In this we echo the conclusion of Banks and White (1984) “that a laterally masked letter is one that has suffered illusory conjunction.”

Studies of illusory conjunction have primarily assessed the effect of attention-related factors (e.g., grouping and cueing), especially to determine whether accurate feature integration requires attention, as claimed by Treisman’s Feature Integration Theory. Most studies manipulated attention (which was theoretically important to them) and fixed the spacing and eccentricity (which seemed secondary). Our emphasis complements theirs. We explored a wide range of spacings and eccentricities, but we made no attempt to divert the observer’s attention, and we always presented the target at the same place on every trial within a block. Presumably these conditions encouraged our observers to devote their full attention to the target location.

Illusory conjunction seems to be an apt description of the effect of crowding, but spatial crowding is not the only way to produce illusory conjunction. In addition to the spatial crowding that is the topic of this article, we will now show that there seems to be a distinct phenomenon, which one might call “temporal crowding,” that also produces illusory conjunctions.

A wide variety of stimuli and tasks have been used to produce illusory conjunctions. Most cases conform to Bouma’s (1970) bound: objects interact only at center-to-center separations less than roughly half the eccentricity. This seems to be crowding. However, some experimental results exceed Bouma’s bound. To survey this, Table 4 extracts an estimate of the critical spacing (as a fraction of target eccentricity) from each paper’s experiments. This is a representative list of papers on illusory conjunction among letters, plus antecedents before the term was coined in 1982.

Looking over the table, we are heartened to see the large number of papers (22, above the line) that are consistent with Bouma’s bound. Here crowding and illusory conjunction agree, so parsimony invites us to treat the two as one. However, we are dismayed to discover a second group of 9 papers, below the line, including the one that introduced the term “illusory conjunction” (Treisman & Schmidt, 1982), which obviously exceed Bouma’s bound and presumably describe some other phenomenon, not the spatial crowding we have been discussing here. There is a hint of a difference in the introspective reports. Recall that the descriptions of crowding (including the Section 1.5 demo) report a messy ambiguous jumbled target. Contrast that with these tidy unambiguous reports from Treisman and Schmidt (1982),

A friend walking in a busy street ‘saw’ a colleague and was about to address him, when he realized that the black beard belonged to one passerby and the bald head and spectacles to another.

Having clearly seen a pink **T**, it was hard to accept the evidence on the card, which showed a pink **X** and a green **T**.

Treisman and Schmidt (1982) seem to be describing the intact migration of nameable high-level object properties (e.g., letter shape), whereas our descriptions of crowding seem to involve migration of elementary features. In the same vein, Wolford and Shum (1980, p. 416) used conditions that may have induced both kinds of illusory conjunction, and found that migration of a tick mark (plausibly an elementary feature) was affected differently by stimulus and task than was migration of a whole symbol: “Feature migrations were sensitive to visual field, occurring primarily in the direction of the fovea, and were not sensitive to report order. Whole-symbol movement was not sensitive to visual field but was affected by report order, with the movement occurring toward the beginning of the instructed report order.” Wolford and Chambers (1983, p. 130) go on to

make the prescient remark, “we were struck by the fact that the studies cited in this [Attention and Perceptual Grouping] section tend to use rather wide spacing between characters . . . , while the experiments in support of feature models tend to use close spacing between characters. The possibility exists, then, that qualitatively different processes are involved at different spacings.”

Treisman and Schmidt (1982) present an impressive array of stimuli and tasks that produce illusory conjunctions with probabilities that are independent of target-flanker spacing, out to spacings much larger than Bouma’s bound. This seems not to be spatial crowding, but one

might call it “temporal crowding.” In these experiments, Treisman and Schmidt loaded attention by asking the observer to report on five objects (e.g., digits and colored letters) presented in a single brief display. All the experiments in Table 4 that exceed Bouma’s bound loaded attention by demanding that the observer examine and report on many hard-to-see objects in a single glimpse.

Temporal migration

The conditions in Table 4 that exceed Bouma’s bound seem closely related to the phenomenon of “temporal migration” (James, 1890; Lawrence, 1971; Sperling & Reeves,

Experiment	Critical spacing (re ecc.)	Ecc. (deg)
Prinzmetal, Hendersen, & Ivry 1995	>0.1	9.60
Estes & Wolford 1971	>0.2	0.75
Wolford & Chambers 1983	>0.2	5.00
Wolford & Hollingsworth 1974a	>0.2	0.95
Townsend et al., 1971	>0.2	1.50
Shaw 1969	>0.3	1.80
Taylor & Brown 1972	>0.3	1.25
Donk 1999 Exps. 1,3-6	>0.3	4.57
Prinzmetal & Millis-Wright 1984 Exps. 2-5	>0.3	1.10
Krumhansl & Thomas 1977	>0.3	1.13
Cohen & Ivry 1989	=0.4	2.50
Eriksen & Rohrbaugh 1970	≈0.4	1.10
Prinzmetal, Presti, & Posner 1986 Exp. 3	>0.4	1.36
Banks, Larson, & Prinzmetal 1979 Exp. 2	=0.5	2.00
Bouma, 1970	=0.5	1 to 12
Wolford & Shum 1980	>0.5	0.73
Wolford & Hollingsworth 1974b	>0.5	0.43
Eriksen & Hoffman 1972	>0.5	1.00
Eriksen & Hoffman 1973	>0.5	1.00
Snyder 1972	>0.5	7.50
Estes et al., 1976	>0.5	1.10
Keele, Cohen, Ivry, Liotti, & Yee 1988	>0.6	1.57
Donk 1999 Exp. 2	>0.7	2.42
Woodrow 1938	>0.9	4.89
Navon & Ehrlich 1995	>1.0	1.20
Santee & Egeth 1982b	>1.7	1.08
Bjork & Murray 1977	>2.0	0.66
Santee & Egeth 1982a	>2.0	0.13
Treisman & Schmidt 1982 Pilot Exp., p. 117	>2.0	2.18
Cohen & Ivry 1989, p. 656	>2.0	2.18
Ivry & Prinzmetal 1991	>2.7	0.20

Table 4. A representative list of papers on illusory conjunction among letters, plus antecedent work, before 1982. (Two papers, Donk, 1999, and Cohen & Ivry, 1989, appear twice, above and below the line. Thus there are 22 papers above the line, 9 below the line, and 29 all told.) Bouma (1970) said that the critical spacing is “roughly 0.5” of the eccentricity, which we here take to mean as low as 0.3 and as high as 0.7. Experimental results above the line are consistent with Bouma’s bound (and thus seem to be crowding); those below the line are inconsistent (and thus seem not to be crowding). Based on each paper’s experimental results, we estimate the critical spacing (see Figure 2), beyond which the flanker no longer affects identification of the target. In most papers the flanker was always effective, and we can infer only that the critical spacing, if one exists, is greater than the largest spacing used, “> x”. A few papers systematically increased spacing until they either found the critical value, which we then list as “= x”, or they reached the limit of their paradigm (e.g., double the eccentricity, “> 2”). Eriksen and Rohrbaugh (1970) sampled spacing coarsely, constraining the critical spacing (as a fraction of eccentricity) to be in the range 0.3 – 0.5, which is designated in the table as “≈0.4”. Note that all the “=” entries are roughly 0.5. Some of the presently consistent cases may turn out to be inconsistent if tested at larger spacings. We predict that all the inconsistent cases will fail the proposed diagnostic test for crowding (Section 4.1). The critical spacing is expressed relative to the (average positive) target eccentricity. We calculate average positive eccentricity by replacing the target by a concentric disk with the same height and computing the average distance from fixation of all its points. The average positive eccentricity of a disk with diameter d at fixation is $d/3$.

1980; McLean, Broadbent, & Broadbent, 1983; Gathercole & Broadbent, 1984; Intraub, 1985, 1989; Shimomura & Yokosawa, 1998; Botella, Suero, & Barriopedro, 2001). Intraub (1989, p. 98) says, “*Temporal migration* describes a situation in which subjects viewing rapidly presented stimuli (e.g., 9 – 20 items/s) confidently [but wrongly] report a target element as having been presented in [the] previous or following stimulus in the sequence. . . . Stimuli (letter, words, or pictures) were presented in rapid succession . . ., in the same spatial location on a screen. The subjects were required to immediately report the stimulus [having] a particular color [or] letter case, or [enclosed in] a black frame. . . . [S]ubjects often reported not the target stimulus (the one actually bearing the searched-for feature), but a temporally adjacent stimulus in the sequence.” Temporal migrations were frequent for frames and pictures at a presentation rate of 9/s, but when numerical digits were substituted for the pictures, there were no temporal migrations until the presentation rate was doubled, to 18/s. Intraub (1985) suggests that the increased presentation rate needed to produce migrations reflects the shorter time it takes to identify a digit than a picture. Similarly, Gathercole and Broadbent (1984) find that migrations require a higher presentation rate with a letter than with a word. Like the Wolford and Shum (1980) and Wolford and Chambers (1983) observations of whole-symbol movement among widely spaced objects, Intraub (1989, p. 99) says, “It is important to note that parts of objects themselves do not typically dissociate and merge with temporally adjacent objects . . ., but rather dissociation and merging frequently occur when an unrelated visual component (e.g., a black outline frame or a homogeneous colored background) is . . . in the same display as one of the objects.”

Treisman and Schmidt (1982) presented their five items all at once, in a single glimpse, rather than serially as in the temporal migration experiments, but their account of the phenomenon supposes serial processing of the items. Thus, internal to the observer, their paradigm may be very much like the paradigms used to produce temporal migration. The rates would be similar if we suppose that the observer takes 1/3 s all together to process the five items (letters and digits) in the glimpse. (Kanwisher, 1991, reports a similar near equivalence between serial and simultaneous presentation of letters in producing repetition blindness.) The experiments seem similar, too, in requiring pressure from both ends of the vice. The experiments press from above by increasing the rate of presentation or number of objects in a glimpse, and press from below by increasing the difficulty of identification of each object.⁶ A digit or letter is identified more quickly than a picture or word, and reducing the target’s duration reduces its effective contrast, making it harder and slower to identify.

Calling the Treisman and Schmidt paradigm “temporal crowding” is a different way to tell their tale, but is basically consistent with their interpretation. *Temporal crowding* produces illusory conjunction when there is insufficient time for the objects to be attended one at a time.

Span of apprehension

There is a long literature on the “visual span of apprehension,” the number of characters that an observer can take in from a single glimpse (e.g., Woodworth, 1938; Sperling, 1960, 1970; Legge, Mansfield, & Chung, 2001). Various kinds of observer response, especially partial report, allow the task to measure the limits of acquisition. The span of apprehension is a key parameter in understanding reading rate.

Treisman and Schmidt’s temporal crowding demonstrations must be very closely related to the span of apprehension because they are essentially the same paradigm: acquiring many symbols in a glimpse. Treisman and Schmidt used shapes, letters, and digits laid out in novel ways, but obtained similar results for the various kinds of object, independent of spacing. Surely, like us, they would expect the same result again if the symbols were just letters evenly spaced in a row, as in the standard span-of-apprehension paradigm, provided that the letters are spaced far enough apart to escape spatial crowding.

Attention

There seems to be little or no effect of attentional manipulations on spatial crowding. Nazir (1992) showed that precueing signal location did not increase identification rates of the signal (Landolt C) in the periphery among flankers (bars, Es, or Os). Similarly, Wilkinson et al. (1997) found no significant effect of precueing on crowded threshold for fine discrimination of grating contrast, frequency, or orientation.

That is not to say that cognitive factors play no role. Illusory conjunctions caused by spatial crowding do occur more often between items that are members of a perceptual group. Grouping may be induced by similarity in color or shape, proximity, good continuation, orthographic structure, or instructions (see Prinzmetal, 1995, for review; also see Chung et al., 2001). Even so, it seems that the critical spacing of crowding is a preattentive limitation.

In an influential paper, Intriligator and Cavanagh (2001) used the crowding paradigm to measure “the spatial resolution of attention.” That is to say, they measured the critical dot spacing for two attentive tasks – tracking and stepping – that emphasize dot location. They acknowledged that the critical spacing that they found is very similar to that obtained by many prior crowding studies, including the asymmetries with respect to upper/lower position in the visual field and radial/circumferential orientation with respect to the fixation point. Like us and prior investigators (e.g., Bouma, 1970), they interpreted the critical spacing as the extent of the selection region that isolates the object being individuated from those around it. However, unlike us, they assumed that “selection” is “the operation of attention”, and thus concluded that the measured resolution (critical spacing) is specific to attention. Stripped of that assumption, their data are compatible with Bouma’s (extended) rule that critical spacing is roughly half of the eccentricity and independent of everything else, including

attention (Nazir, 1992; Wilkinson et al., 1997). The most parsimonious interpretation is that the measured resolution is a preattentive limitation, the critical spacing of crowding, not a property of attention per se. This supposes that the control tasks that they used to assess “visual” resolution (acuity) require detecting only a single feature and are thus immune to crowding, and that the tracking and stepping tasks that they used to assess “attentional” resolution are fine spatial discriminations that require feature integration and thus are susceptible to crowding. In other words, their distinction between “visual” and “attentional” resolution was really the distinction between one-feature tasks, which are crowding-immune, and multi-feature tasks, which are crowding-susceptible.

Spatial vs. temporal crowding

This article is not about attention, but it does seem worth noting some implications of our interpretation of Table 4 for Treisman’s Feature Integration Theory. The experiments in the upper part of the table – consistent with Bouma’s bound – seem to be spatial crowding, and the experiments in the lower part of the table – violating Bouma’s bound (and closely related to temporal migration) – are temporal crowding. Feature Integration Theory supposes a serial “attention” process that binds features to make objects, which predicts object-rate-limited performance and provides a good account of temporal crowding. The critical spacing of spatial crowding seems to be a preattentive limitation, unaffected by attentional manipulations, and outside the scope of Feature Integration Theory. Spatial crowding of object identification occurs even when there is no attentional loading at all, as in the Section 1.5 demo (or Woodworth and Schlosberg, 1954, p. 104; Banks et al., 1979; Prinzmetal, Henderson, & Ivry, 1995).

Most of the papers on illusory conjunction seem to have the purpose of addressing Treisman and Schmidt’s evidence for Feature Integration Theory yet used conditions that produced spatial not temporal crowding. One could easily be misled by a passing comment of Treisman and Schmidt (1982, p. 113) about how to produce illusory conjunctions. Treisman and Schmidt explicitly reduced the duration of presentation in order to overload attention. It is clear that the task becomes harder as signal duration is reduced, but it is unclear why it’s harder. It seems that they thought that by reducing the signal duration they were reducing the observer’s time available for processing.⁷ In our interpretation, their trial-and-error reduction of duration to achieve a desired level of performance reduced object visibility to near threshold, which reduced the number of elementary features detected, making the identification computationally harder, which increased the required time to recognize each object enough to overload the observer with the five objects presented. Thus it seems that illusory conjunctions can be produced in two ways: temporally, by asking the observer to process many objects in a glimpse or at a great rate, or spatially, by putting two objects within Bouma’s bound, most easily in the periphery. Temporally,

processing is slower for pictures than for digits and slower for hard than easy-to-see letters (Intraub, 1989; Legge, Rubin, & Luebker, 1987). Not suspecting that making the given number of objects hard-to-see might be essential to replicating Treisman and Schmidt’s effect – temporal crowding – most authors simply took the occurrence of illusory conjunctions as their measure of success, and ended up studying spatial crowding.

Cohen and Ivry (1989, p. 656) report an experimental manipulation that spans the gulf between spatial and temporal crowding. They report pilot studies comparing a two-letter presentation in the peripheral visual field (2.71° eccentricity) accompanied either by two digits presented peripherally, left and right, outside the letters (following Treisman & Schmidt), or by one digit, centrally, at fixation. The observer is asked to report all the digits and letters. Each letter is displayed somewhere on an imaginary 2.71° radius circle centered on fixation. Unbeknownst to the observer, the outcome of interest is the frequency of illusory conjunction between the two letters as a function of letter spacing. When the digits are peripheral, one on each side of the display, Cohen and Ivry replicate Treisman and Schmidt’s result, finding illusory junctions at all letter spacings, including those well beyond Bouma’s bound, as one would expect for temporal crowding. When the digit is central, Cohen and Ivry get illusory conjunctions only at small spacings, within Bouma’s bound, as one would expect for spatial crowding. Cohen and Ivry attributed the difference in results to whether the letters are inside or outside the “attentional spotlight” spanning the digit or digits. Our interpretation of their finding is that the digit is easy to see centrally, so it and the two letters are processed quickly, the observer is not overloaded, there is no temporal crowding, and one gets the standard spatial crowding result. Going from one to two digits presumably increases the attentional load, and moving the digit(s) to the periphery reduces their visibility, which increases their processing time enough to overload the observer, producing temporal crowding. (We would expect closely spaced peripheral letters to always produce spatial crowding. In the peripheral-digits condition, this spatial crowding would be over and above the temporal crowding, but Cohen and Ivry’s summary of this pilot experiment doesn’t go into such detail.) Once again, all the violations of Bouma’s bound seem to be temporal, not spatial, crowding.

In sum, most of the papers on illusory conjunction report results that are consistent with Bouma’s bound and thus seem to be about spatial crowding, as defined here. Many have no time pressure at all. Spatial crowding depends on spatial separation, independent of time pressure. The 9 papers that exceed Bouma’s bound all overloaded attention by asking the observer to process many hard-to-see objects in a glimpse. This effect is produced by time pressure and is independent of spatial separation, so we call it “temporal crowding.”

In her Bartlett memorial lecture, Treisman (1988) singled out feature integration as a key bottleneck in percep-

tion. We agree, but distinguish two feature-integration bottlenecks, revealed by spatial and temporal crowding. When objects are too close in space or in time their features may be inappropriately combined, producing illusory conjunction. However, spatial crowding seems to combine elementary features, typically resulting in a hard-to-describe jumble, whereas temporal crowding seems to combine high-level nameable object properties (e.g., shape and color), typically resulting in a plausible nameable object. The isolation that is disturbed by spatial crowding seems to reflect local anatomy of the visual field, independent of attention. The isolation that is disturbed by temporal crowding seems to reflect both the maximum rate at which objects can be recognized and the visual span of apprehension, which may be closely related. Treisman and Schmidt (1982) give a good account of temporal crowding as a speed limit on object processing in a serial attentive scan. The critical spacing of spatial crowding seems to be a preattentive limitation, outside the scope of their theory. Spatial crowding determines the minimum spacing (at each eccentricity) at which we can recognize objects, and temporal crowding determines the maximum rate at which we can recognize them.

5. Conclusion

A diagnostic test for crowding is proposed in Section 4.1. The facts of crowding and our theoretical conclusions are summarized in Tables 2 and 3.

This test and characterization can help answer some longstanding questions about object identification, such as whether faces are recognized by parts and the roles of letter and word recognition in reading (Martelli et al., *in press*; Su et al., 2004).

We have sketched a two-stage model – independent feature detection followed by feature integration – to account for the results. We have suggested that ordinary masking impairs feature detection and that crowding impairs feature integration. In ordinary masking, the mask interferes by stimulating the signal's feature detector. In crowding, the mask and signal features are detected separately and subsequently amalgamated by the integration field. Crowding exposes the inner workings of the feature integrator.

A single object's various perceived properties can be estimated based on regions of very different sizes: a small region for one-feature properties (e.g., presence or coarse location), which don't require integration, and a large region for multiple-feature properties (e.g., shape or any fine discrimination), which do require integration. Despite its unitary appearance, a perceptual "object" may be just a loose bundle of independently estimated properties.

Finally, a survey of the illusory conjunction literature finds that most of the illusory conjunction results are consistent with the spatial crowding described here, which depends on spatial proximity independent of time pressure, and that the rest seem to arise through a distinct phe-

nomenon that one might call "temporal crowding," which depends on time pressure ("overloading attention"), independent of spatial proximity.

6. Acknowledgments

This is the first in a series of papers about crowding and its cure, isolating to recognize (#2 Martelli et al., *in press*; #3 Su et al., 2004). Thanks to Gordon Legge, Dennis Levi, and Bosco Tjan, among others, for an intriguing discussion at ARVO that raised the question, to Jack Loomis for his advice at the beginning and thoughts at the end, to Marisa Carrasco for pointing out that Treisman and Schmidt's results are inconsistent with (spatial) crowding, to Susana Chung, Bart Farell, Elisabeth Fine, Marialuisa Martelli, Greg Murphy, Endel Pöder, Anne Treisman, Jeremy Wolfe, and an anonymous reviewer for their helpful critiques, to Emily Huang for discussion of "spatial resolution of visual attention," to Duncan Luce for noting the instability of the A-B percept in the Section 1.5 demo, to Mehrdad Jazayeri for asking about mask contrast, to Gina Cardazone for pilot foveal data, to Joshua Green for pilot grating-grating data, to Uta Wolfe for translation of Korte, and to Maria Carla LaPutt and Arusha Gupta for careful observations. Some results were presented in Gina Cardazone's 1998 Westinghouse National Science Talent Search essay and at the Annual Meeting of the Association for Research in Vision and Ophthalmology, Ft. Lauderdale, FL (Palomares, LaPutt, & Pelli, 1999; Pelli & Palomares, 2000). This work was supported by National Eye Institute grant EY04432 to DGP. Melanie Palomares is now a graduate student in the Department of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD, USA.

Commercial relationships: none.

Corresponding author: Denis G. Pelli.

Email: denis.pelli@nyu.edu.

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

7. Footnotes

¹This model is simple, omitting irrelevant details. Recent models of the "feature detector" to account for ordinary masking are more complicated than this (Foley, 1994; Watson & Solomon, 1997; Wilson & Kim, 1998; Foley & Chen, 1999; Itti, Koch, & Braun, 2000). They incorporate several similar receptive fields, some of which are used for normalization by divisive inhibition. The differences among these receptive fields are negligible for the purpose of explaining crowding, so our presentation will assume the simple old-fashioned model of the feature detector, with a single receptive field.

²Placing the nonlinearity and threshold in the feature detector is a fiction, a useful but unrealistic assumption

that is adequate for the purposes of this paper, but ultimately to be replaced by the nonlinear combination rule of the feature integrator. We cannot suppose that the feature detector and integrator are both linear because the cascade of two linear operations is equivalent to a single linear operation, so the two stages would collapse to one consisting of more-complex-shaped-feature detectors and no second stage. Selfridge's (1959) pandemonium model suffers from this defect. Minsky and Papert (1969) showed that the computational power of a one-layer network (i.e., simple feature detection), which they dubbed a "perceptron," is quite limited and cannot solve some important problems, such as closure. The second-order literature is full of tasks that people can do, but which an isolated feature detector cannot (see Chubb, Olzak, & Derrington, 2001; Landy & Graham, 2004). However, if the layers are separated by a nonlinearity then they don't collapse and the perceptron limits do not apply.

For the purposes of this paper we assume a high threshold within the feature detector. The high-threshold assumption is known to be wrong, because it cannot cope with the finding that when observers lower their criterion both the hit and false alarm rates rise together (Nachmias & Steinman, 1965). Nor can it cope with the finding that when noise is added to the display, the observer's threshold is at a constant effective signal-to-noise ratio (Pelli, 1985; Pelli & Farell, 1999). (Also see Palmer, Verghese, & Pavel, 2000.) However, for a given criterion and noise level, the high-threshold assumption remains a popular and useful simplification because it captures the steepness of real psychometric functions (much steeper than a cumulative normal), the relatively sudden disappearance of the signal as contrast is reduced below threshold, and allows surprisingly accurate calculations of the increase in visibility with added features (Pelli, 1985; Graham, 1989). In fact, the criterion and noise effects tell us that the nonlinearity cannot lie within the feature detector, and must, instead, be in the integrator, in the way the (linear) feature signals are combined. For example, for detection and near-threshold discrimination, a maximum rule (output is maximum of the many inputs) among 100 to 1000 linear detectors exhibits appropriate dependence on criterion, added noise, and number of features (Pelli, 1985). The tasks considered here (e.g., letter identification) clearly require different nonlinear combination rules, which are still a mystery.

³"A feature pops out and a conjunction does not." Whether to call this a dichotomy has become controversial, as there is a continuum of intermediate cases (Bergen & Julesz, 1983; Nakayama & Silverman, 1986; Duncan & Humphreys, 1989; Verghese & Pelli, 1992; Wolfe, Friedman-Hill, Stewart, & O'Connell, 1992; Verghese & Nakayama, 1994; Wolfe, 1998; Palmer et al., 2000). Search difficulty depends on how much difference there is between target and distractors relative to the difference between distractors (Duncan & Humphreys, 1989). This can be varied over a wide range, spoiling any claim that it's generally easy or hard to search for a feature or conjunction target. How-

ever, the critical spacing of crowding, independent of nearly everything else, seems likely to be independent of task difficulty, so we expect a robust feature vs. conjunction dichotomy in crowding: small critical spacing (no crowding) for a one-feature judgment and Bouma's large critical spacing for a multi-feature judgment. (Carrasco, Evert, Chang, & Katz, 1995, showed that detection of a conjunction target among distractors is harder at larger eccentricity, which may be an effect of crowding.)

⁴Ehlers (1953) recounts, "When one is teaching a child his letters, it is the custom to present at the beginning only a few letters on each page. This is because of another characteristic of the eyesight. The sense of form is most easily utilized in an otherwise empty visual field. When one is testing amblyopic children with isolated letters or Es, the visual acuity recorded is often much better than with the ordinary test chart. If the visual field is crowded with letters, the area of the visual field in which the letters can be recognized narrows. This is very easy to demonstrate, as I showed at the Congress of Scandinavian Ophthalmologists in 1936." [The published 1936 paper has no mention of this.]

⁵One may well ask here, what is an object, and how does the visual system segment the scene to isolate an object? In this context an object is a perceptual hypothesis, accounting for part of the image by a solid with a specific velocity, etc. Does segmentation precede identification? Or does vision attempt to recognize using all available integration fields and infer the segmentation from the best identification? We don't know.

⁶Treisman and Schmidt's experiments used various kinds of report (e.g., "were any two targets identical?") to show that the observer's limitation was in the scrutiny of the objects, not in the remembering and reporting. They "controlled exposure duration separately for each subject in order to produce a feature error rate of 10%" (Treisman & Schmidt, 1982, p. 113). In their Experiment 1, the mean target duration was 120 ms, and was followed by a visual noise mask. On average, the objects reported by the observer included only half of the features actually presented. Thus, visibility was restricted, making the targets hard to identify. Pelli, Burns, Farell, and Moore (*in press*) show that the identification of a letter seems to require the independent detection of a modest number of elementary features of the letter. As visibility is reduced, approaching threshold, fewer features are detected. Presumably the observer is slower when struggling to identify ambiguous percepts based on fewer detected elementary features. In the same way, reading rate is largely independent of contrast, but slows near threshold contrast (Legge et al., 1987). Treisman and Schmidt (1982) presented their brief signal at threshold duration, which is probably similar in effect to presenting it at threshold contrast. Legge et al. (1987) found that the fixation duration during reading is prolonged at low near-threshold contrasts.

⁷Treisman and Schmidt (1982, p. 113) say that their experiment has "conflicting requirements." To overload

attention they “need to present several items and to use brief exposures. ... [Yet,] illusory conjunctions can be formed only from correctly identified features, ... [and] the briefer the exposure, the poorer the quality of the sensory information,” which introduces errors in identifying features. “Thus we were forced to trade off the need to load resources [create time pressure] against the risk of introducing data limits [reducing visibility].” Contrary to their interpretation, we suspect that reducing the duration affected the probability of illusory conjunction primarily by reducing visibility (i.e., “limiting data”). The briefer signals are weaker, so not all elementary features are detected, making the recognition (feature integration) harder and slower.⁶

8. References

- Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, 300(5892), 523-525. [PubMed]
- Andriessen, J. J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research*, 16(1), 71-78. [PubMed]
- Anstis, S. M. (1974). Letter: A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14(7), 589-592. [PubMed]
- Banks, W. P., Larson, D. W., & Prinzmetal, W. (1979). Asymmetry of visual interference. *Perception & Psychophysics*, 25(6), 447-456. [PubMed]
- Banks, W. P., & White, H. (1984). Lateral interference and perceptual grouping in visual detection. *Perception & Psychophysics*, 36(3), 285-95. [PubMed]
- Baylis, G. C., & Driver, J. (1992). Visual parsing and response competition: The effect of grouping factors. *Perception & Psychophysics*, 51(2), 145-162. [PubMed]
- Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature*, 303(5919), 696-698. [PubMed]
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115-147. [PubMed]
- Bischof, W. F., & Di Lollo, V. (1995). Motion and metacontrast with simultaneous onset of stimuli. *Journal of the Optical Society of America A*, 12(8), 1623-1636. [PubMed]
- Bjork, E. L., & Murray, J. T. (1977). On the nature of input channels in visual processing. *Psychological Review*, 84(5), 472-484. [PubMed]
- Bondarko, V. M., & Danilova, M. V. (1996). Crowding effects with different surroundings and interstimulus distances. *Sensory Systems*, 10(3), 170-177.
- Bondarko, V. M., & Danilova, M. V. (1997). What spatial frequency do we use to detect the orientation of a Landolt C? *Vision Research*, 37(15), 2153-2156. [PubMed]
- Botella, J., Suero, M., & Barriopedro, M. I. (2001). A model of the formation of illusory conjunctions in the time domain. *Journal of Experimental Psychology: Human Perception and Performance*, 27(6), 1452-1467. [PubMed]
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(241), 177-178. [PubMed]
- Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research*, 13(4), 767-782. [PubMed]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433-436. [PubMed]
- Breitmeyer, B. G. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.
- Brindley, G. S. (1960). Two more visual theorems. *Quarterly Journal of Experimental Psychology*, 12, 110-112.
- Burr, D. C. (1984). Summation of target and mask metacontrast stimuli. *Perception*, 13(2), 183-92. [PubMed]
- Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology*, 197(3), 551-566. [PubMed]
- Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, 57(8), 1241-1261. [PubMed]
- Cavanagh, P. (2001). Seeing the forest but not the trees. *Nature Neuroscience*, 4(7), 673-674. [PubMed]
- Chambers, L., & Wolford, G. (1983). Lateral masking vertically and horizontally. *Bulletin of the Psychonomic Society*, 21(6), 459-461.
- Chastain, G. (1981). Inhibition of feature extraction with multiple instances of the target feature in different orientations. *Psychological Research*, 43(1), 45-56.
- Chastain, G. (1983). Task and contrast effects on performance with parafoveal stimulus pairs. *Psychological Research*, 45(2), 147-156. [PubMed]
- Chastain, G., & Lawson, L. (1979). Identification asymmetry of parafoveal stimulus pairs. *Perception & Psychophysics*, 26(5), 363-368.
- Chubb, C., Olzak, L., & Derrington, A. (2001). Second-order processes in vision: Introduction. *Journal of the Optical Society of America A*, 18(9), 2175-2178. [PubMed]
- Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research*, 41(14), 1833-1850. [PubMed]

- Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance*, 15(4), 650-663. [PubMed]
- Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: Sensitivity to stimulus form. *Journal of Neurophysiology*, 57(3), 835-868. [PubMed]
- Desimone, R., Schein, S. J., Moran, J., & Ungerleider, L. G. (1985). Contour, color and shape analysis beyond the striate cortex. *Vision Research*, 25(3), 441-452. [PubMed]
- Didner, R., & Sperling, G. (1980). Perceptual delay: A consequence of metacontrast and apparent motion. *Journal of Experimental Psychology: Human Perception and Performance*, 6(2), 235-243. [PubMed]
- Donk, M. (1999). Illusory conjunctions are an illusion: The effects of target-nontarget similarity on conjunction and feature errors. *Journal of Experimental Psychology: Human Perception and Performance*, 25(5), 1207-1233.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433-458. [PubMed]
- Ehlers, H. (1936). The movements of the eyes during reading. *Acta Ophthalmologica*, 14, 56-63.
- Ehlers, H. (1953). Clinical testing of visual acuity. *A.M.A. Archives of Ophthalmology*, 49(4), 431-434. [PubMed]
- Enns, J. T. (2004). Object substitution and its relation to other forms of visual masking. *Vision Research*, 44(12), 1321-1331. [PubMed]
- Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science*, 8(2), 135-139.
- Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Science*, 4(9), 345-352. [PubMed]
- Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, 12(2-B), 201-204.
- Eriksen, C. W., & Hoffman, J. E. (1973). The extent of processing of noise elements during selective encoding from visual displays. *Perception & Psychophysics*, 14(1), 155-160.
- Eriksen, C. W., & Rohrbaugh, J. W. (1970). Some factors determining efficiency of selective attention. *American Journal of Psychology*, 83(3), 330-342.
- Estes, W. K. (1982). Similarity-related channel interactions in visual processing. *Journal of Experimental Psychology: Human Perception and Performance*, 8(3), 353-382. [PubMed]
- Estes, W. K., Allmeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics*, 19(1), 1-15.
- Estes, W. K., & Wolford, G. L. (1971). Effects of spaces on report from tachistoscopically presented letter strings. *Psychonomic Science*, 25(2), 77-80.
- Fine, E. M. (2003). Reduced contrast does not reduce visual crowding [Abstract]. *Journal of Vision*, 3(9), 814a, <http://journalofvision.org/3/9/814/>, doi:10.1167/3.9.814.
- Flom, M. C. (1991). Contour interaction and the crowding effect. *Problems in Optometry*, 3, 237-257.
- Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effect. *Science*, 142, 979-980. [PubMed]
- Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America*, 53(9), 1026-1032. [PubMed]
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, 11(6), 1710-1719. [PubMed]
- Foley, J. M., & Chen, C. C. (1999). Pattern detection in the presence of maskers that differ in spatial phase and temporal offset: Threshold measurements and a model. *Vision Research*, 39(23), 3855-3872. [PubMed]
- Gathercole, S. E., & Broadbent, D. E. (1984). Combining attributes in specified and categorized target search: Further evidence for strategic differences. *Memory & Cognition*, 12(4), 329-337. [PubMed]
- Gilbert, C., Ito, M., Kapadia, M., & Westheimer, G. (2000). Interactions between attention, context and learning in primary visual cortex. *Vision Research*, 40(10-12), 1217-1226. [PubMed]
- Graham, N. (1980). Spatial-frequency channels in human vision: Detecting edges without edge detectors. In C. Harris (Ed.), *Visual coding and adaptability* (pp. 215-252). Hillsdale, NJ: Erlbaum.
- Graham, N. (1985). Detection and identification of near-threshold visual patterns. *Journal of the Optical Society of America A*, 2(9), 1468-1482. [PubMed]
- Graham, N. (1989). *Visual pattern analyzers*. Oxford: Oxford University Press.
- Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision Research*, 38(2), 231-257. [PubMed]
- Greis, U., & Rohler, R. (1970). Untersuchung der subjektiven Detailerkennbarkeit mit Hilfe der Ortsfrequenzfilterung. *Optica Acta*, 17, 515-526.

- Hazeltine, R. E., Prinzmetal, W., & Elliott, W. (1997). If it's not there, where is it? Locating illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, 23(1), 263-277. [PubMed]
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, 383(6598), 334-337. [PubMed]
- Hess, R. F., Dakin, S. C., & Kapoor, N. (2000). The foveal 'crowding' effect: Physics or physiology? *Vision Research*, 40(4), 365-370. [PubMed]
- Hess, R. F., Dakin, S. C., Kapoor, N., & Tewfik, M. (2000). Contour interaction in fovea and periphery. *Journal of the Optical Society of America A*, 17(9), 1516-1524. [PubMed]
- Huckauf, A., & Heller, D. (2004). On the relations between crowding and visual masking. *Perception & Psychophysics*, 66(4), 584-595. [PubMed]
- Intraub, H. (1985). Visual dissociation: An illusory conjunction of pictures and forms. *Journal of Experimental Psychology: Human Perception and Performance*, 11(4), 431-442. [PubMed]
- Intraub, H. (1989). Illusory conjunctions of forms, objects, and scenes during rapid serial visual search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(1), 98-109. [PubMed]
- Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology*, 43(3), 171-216. [PubMed]
- Itti, L., Koch, C., & Braun, J. (2000). Revisiting spatial vision: Toward a unifying model. *Journal of the Optical Society of America A*, 17(11), 1899-1917. [PubMed]
- Ivry, R. B., & Prinzmetal, W. (1991). Effect of feature similarity on illusory conjunctions. *Perception & Psychophysics*, 49(2), 105-116. [PubMed]
- Jacobs, R. J. (1979). Visual resolution and contour interaction in the fovea and periphery. *Vision Research*, 19(11), 1187-1195. [PubMed]
- James, W. (1890). *Principles of psychology* (Vol. 1, Chap. II). New York: Holt.
- Kahneman, D., & Henik, A. (1977). Effects of visual grouping on immediate recall and selective attention. In S. Dornic (Ed.), *Attention and performance VI*. Hillsdale, NJ: Erlbaum.
- Kanwisher, N. (1991). Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance*, 17(2), 404-421. [PubMed]
- Keele, S. W., Cohen, A., Ivry, R., Liotti, M., & Yee, P. (1988). Tests of a temporal theory of attentional binding. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 444-452. [PubMed]
- King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, 34(7), 885-912. [PubMed]
- Klein, S. A., Casson, E., & Carney, T. (1990). Vernier acuity as line and dipole detection. *Vision Research*, 30(11), 1703-1719. [PubMed]
- Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, 8(2), 255-279. [PubMed]
- Korte, W. (1923). Über die Gestaltauffassung im indirekten Sehen. *Zeitschrift für Psychologie*, 93, 17-82.
- Krumhansl, C. L., & Thomas, E. A. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception & Psychophysics*, 21(3), 269-279.
- Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1106-1118). Cambridge, MA: MIT Press.
- Latham, K., & Whitaker, D. (1996). Relative roles of resolution and spatial interference in foveal and peripheral vision. *Ophthalmic and Physiological Optics*, 16(1), 49-57. [PubMed]
- Lawrence, D. H. (1971). Two studies of visual search for word targets with controlled rates of presentation. *Perception & Psychophysics*, 10, 85-59.
- Leat, S. J., Li, W., & Epp, K. (1999). Crowding in central and eccentric vision: The effects of contour interaction and attention. *Investigative Ophthalmology and Visual Science*, 40(2), 504-512. [PubMed]
- Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *Journal of the Optical Society of America*, 70(12), 1458-1471. [PubMed]
- Legge, G. E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading. XX. Linking letter recognition to reading speed in central and peripheral vision. *Vision Research*, 41(6), 725-743. [PubMed]
- Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading. V. The role of contrast in normal vision. *Vision Research*, 27(7), 1165-1177. [PubMed]
- Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*, 2(2), 167-177, <http://journalofvision.org/2/2/3/>, doi:10.1167/2.2.3. [PubMed][Article]

- Levi, D. M., & Klein, S. A. (1983). Spatial localization in normal and amblyopic vision. *Vision Research*, 23(10), 1005-1017. [PubMed]
- Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Research*, 25(7), 979-991. [PubMed]
- Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, 25(7), 963-977. [PubMed]
- Levi, D. M., Klein, S. A., & Hariharan, S. (2002). Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking. *Journal of Vision*, 2(2), 140-166, <http://journalofvision.org/2/2/2/>, doi:10.1167/2.2.2. [PubMed][Article]
- Liu, L., & Arditi, A. (2000). Apparent string shortening concomitant with letter crowding. *Vision Research*, 40(9), 1059-1067. [PubMed]
- Loomis, J. M. (1978). Lateral masking in foveal and eccentric vision. *Vision Research*, 18(3), 335-338. [PubMed]
- Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, 42(9), 1165-1184. [PubMed]
- Martelli, M., Pelli, D. G., & Majaj, N. J. (in press). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*.
- McLean, J. P., Broadbent, D. E., & Broadbent, M. H. (1983). Combining attributes in rapid serial visual presentation tasks. *A Quarterly Journal of Human Psychology*, 35(Pt 1), 171-186. [PubMed]
- Minsky, M. L., & Papert, S. (1969). *Perceptrons: An essay in computational geometry*. Cambridge, MA: MIT Press.
- Morrone, M. C., Burr, D. C., & Vaina, L. M. (1995). Two stages of visual processing for radial and circular motion. *Nature*, 376(6540), 507-509. [PubMed]
- Motter, B. C. (1994a). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *Journal of Neuroscience*, 14(4), 2178-2189. [PubMed]
- Motter, B. C. (1994b). Neural correlates of feature selective memory and pop-out in extrastriate area V4. *Journal of Neuroscience*, 14(4), 2190-2199. [PubMed]
- Motter, B. C. (2002). Crowding and object integration within the receptive field of V4 neurons [Abstract]. *Journal of Vision*, 2(7), 274a, <http://journalofvision.org/2/7/274/>, doi:10.1167/2.7.274.
- Movshon, J. A., Adelson, E. H., Gizzi, M. S., & Newsome, W. T. (1986). The analysis of moving visual patterns. In C. Chagas, R. Gattass, & C. Gross (Eds.), *Pattern recognition mechanisms* (pp. 117-151). New York: Springer-Verlag.
- Mullen, K. T., & Losada, M. A. (1999). The spatial tuning of color and luminance peripheral vision measured with notch filtered noise masking. *Vision Research*, 39(4), 721-731. [PubMed]
- Nachmias, J., & Steinman, R. M. (1965). Brightness and discriminability of light flashes. *Vision Research*, 5(9), 545-557. [PubMed]
- Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320(6059), 264-265. [PubMed]
- NAS-NRC (1980). Recommended standard procedures for the clinical measurement and specification of visual acuity: Report of working group 39. Committee on vision. Assembly of Behavioral and Social Sciences, National Research Council, National Academy of Sciences, Washington, D.C. *Advances in Ophthalmology*, 41, 103-148. [PubMed]
- Navon, D., & Ehrlich, B. (1995). Illusory conjunctions: Does inattention really matter? *Cognitive Psychology*, 29(1), 59-83.
- Nazir, T. A. (1992). Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Research*, 32(4), 771-777. [PubMed]
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- Olzak, L., & Thomas, J. P. (1986). Seeing spatial patterns. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Sensory processes and perception* (Vol. 1). New York: Wiley.
- Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. *Vision Research*, 40(10-12), 1227-1268. [PubMed]
- Palomares, M., LaPutt, M. C., & Pelli, D. G. (1999). Crowding is unlike ordinary masking. *Investigative Ophthalmology and Visual Science* [Abstract], 40, 1864.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739-744. [PubMed]
- Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America A*, 2(9), 1508-1532. [PubMed]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437-442. [PubMed]
- Pelli, D. G., Burns, C. W., Farell, B., & Moore, D. C. (in press). Identifying letters. *Vision Research*.
- Pelli, D. G., & Farell, B. (1999). Why use noise? *Journal of the Optical Society of America A*, 16(3), 647-653. [PubMed]

- Pelli, D. G., Farell, B., & Moore, D. C. (2003). The remarkable inefficiency of word recognition. *Nature*, 423(6941), 752-756. [PubMed]
- Pelli, D. G., Levi, D. M., & Chung, S. T. L. (2004). Using noise to characterize amblyopic letter identification. *Journal of Vision*, 4(10), <http://journalofvision.org/4/10/6/>, doi:10.1167/4.10.6 [PubMed][Article]
- Pelli, D. G., & Palomares, M. (2000). The role of feature detection in crowding. *Investigative Ophthalmology and Visual Science* [Abstract], 41, S37.
- Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research*, 31(7-8), 1337-1350. [PubMed]
- Phillips, G. C., & Wilson, H. R. (1984). Orientation bandwidths of spatial mechanisms measured by masking. *Journal of the Optical Society of America A*, 1(2), 226-232. [PubMed]
- Piñon, M. C., Gattass, R., & Sousa, A. P. (1998). Area V4 in *Cebus* monkey: Extent and visuotopic organization. *Cerebral Cortex*, 8(8), 685-701. [PubMed]
- Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research*, 33(7), 993-999. [PubMed]
- Prinzmetal, W. (1995). Visual feature integration in a world of objects. *Current Directions in Psychological Science*, 4(3), 90-94.
- Prinzmetal, W., Henderson, D., & Ivry, R. (1995). Loosening the constraints on illusory conjunctions: Assessing the roles of exposure duration and attention. *Journal of Experimental Psychology: Human Perception and Performance*, 21(6), 1362-1375. [PubMed]
- Prinzmetal, W., & Millis Wright, M. (1984). Cognitive and linguistic factors affect visual feature integration. *Cognitive Psychology*, 16(3), 305-340. [PubMed]
- Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? *Journal of Experimental Psychology: Human Perception and Performance*, 12(3), 361-369. [PubMed]
- Quick, R. F., Jr. (1974). A vector-magnitude model of contrast detection. *Kybernetik*, 16(2), 65-67. [PubMed]
- Raghavan, M. (1995). *Sources of visual noise*. Doctoral dissertation, Syracuse University, Syracuse, NY.
- Reeves, A. (1982). Metacontrast U-shaped functions derive from two monotonic processes. *Perception*, 11(4), 415-426. [PubMed]
- Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, 21(3), 409-418. [PubMed]
- Rosch, E., & Lloyd, B.B. (1978). *Cognition and categorization*. Hillsdale, NJ: Erlbaum.
- Sagi, D. (1990). Detection of an orientation singularity in Gabor textures: Effect of signal density and spatial-frequency. *Vision Research*, 30(9), 1377-1388. [PubMed]
- Sagi, D., & Julesz, B. (1984). Detection versus discrimination of visual orientation. *Perception*, 13(5), 619-628. [PubMed]
- Santee, J. L., & Egeth, H. E. (1982a). Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception and Performance*, 8(4), 489-501. [PubMed]
- Santee, J. L., & Egeth, H. E. (1982b). Independence versus interference in the perceptual processing of letters. *Perception and Psychophysics*, 31(2), 101-116. [PubMed]
- Selfridge, O. G. (1959). *Pandemonium: A paradigm for learning*. London: H. M. Stationery Office.
- Shaw, P. (1969). Processing of tachistoscopic displays with controlled order of characters and spaces. *Perception & Psychophysics*, 6(5), 257-266.
- Sherrington, C. S. (1906). *The integrative action of the nervous system*. New Haven: Yale University Press.
- Shimomura, M., & Yokosawa, K. (1998). Temporal illusory conjunction in RSVP tasks: Effects of target complexity. *Shinrigaku Kenkyu [Japanese Journal of Psychology]*, 68(6), 449-456. [PubMed]
- Snyder, C. R. (1972). Selection, inspection, and naming in visual search. *Journal of Experimental Psychology*, 92(3), 428-431. [PubMed]
- Solomon, J. A., & Morgan, M. J. (2000). Facilitation from collinear flanks is cancelled by non-collinear flanks. *Vision Research*, 40(3), 279-286. [PubMed]
- Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, 369(6479), 395-397. [PubMed]
- Sperling, G. (1960). The information available in brief visual presentation. *Psychological Monographs*, 74, 1-29.
- Sperling, G. (1970). Short-term memory, long-term memory, and scanning in the processing of visual information. In G. A. Young & D. B. Lindsley (Eds.), *Early experience and visual information processing in perceptual and reading disorders* (pp. 198-218). Washington, DC: National Academy of Sciences.
- Sperling, G., & Reeves, A. (1980). Measuring the reaction time of an unobservable response: A shift of visual attention. In R. Nickerson (Ed.), *Attention & Performance VIII* (pp. 347-360). Hillsdale, NJ: Erlbaum.
- Strasburger, H. (2001). Invariance of the psychometric function for character recognition across the visual field. *Perception & Psychophysics*, 63(8), 1356-1376. [PubMed]
- Strasburger, H. (2002). *Indirektes Sehen: Formerkennung im zentralen und peripheren Gesichtsfeld*. Göttingen: Hogrefe-Verlag. [English translation expected in 2005.]

- Strasburger, H., Harvey, L. O., Jr., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, 49(6), 495-508. [PubMed]
- Stromeyer, C. F., 3rd, & Julesz, B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America*, 62(10), 1221-1232. [PubMed]
- Stuart, J. A., & Burian, H. M. (1962). A study of separation difficulty: Its relationship to visual acuity in normal and amblyopic eyes. *American Journal of Ophthalmology*, 53, 471-477. [PubMed]
- Su, M., Berger, T. D., Majaj, N., & Pelli, D. G. (2004). *Crowding and reading*. Manuscript in preparation.
- Swift, D. J., & Smith, R. A. (1983). Spatial frequency masking and Weber's Law. *Vision Research*, 23(5), 495-505. [PubMed]
- Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition*, 67(1-2), 1-20. [PubMed]
- Tata, M. S. (2002). Attend to it now or lose it forever: Selective attention, metacontrast masking, and object substitution. *Perception & Psychophysics*, 64(7), 1028-1038. [PubMed]
- Taylor, S. G., & Brown, D. R. (1972). Lateral visual masking: Supraretinal effects when viewing linear arrays with unlimited viewing time. *Perception & Psychophysics*, 12(1-B), 97-99.
- Thomas, J. P. (1985a). Detection and identification: How are they related? *Journal of the Optical Society of America A*, 2(9), 1457-1467. [PubMed]
- Thomas, J. P. (1985b). Effect of static-noise and grating masks on detection and identification of grating targets. *Journal of the Optical Society of America A*, 2(9), 1586-1592. [PubMed]
- Thomas, J. P. (1987). Effect of eccentricity on the relationship between detection and identification. *Journal of the Optical Society of America A*, 4(8), 1599-1605. [PubMed]
- Thomas, J. P., & Gille, J. (1979). Bandwidths of orientation channels in human vision. *Journal of the Optical Society of America A*, 69(5), 652-660. [PubMed]
- Thomas, J. P., Padilla, G. J., & Rourke, D. L. (1969). Spatial interactions in identification and detection of compound visual stimuli. *Vision Research*, 9(2), 283-292. [PubMed]
- Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, 32(7), 1349-1357. [PubMed]
- Townsend, J. T., Taylor, S. G., & Brown, D. R. (1971). Lateral masking for letters with unlimited viewing time. *Perception & Psychophysics*, 10(5), 375-378.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology A*, 40(2), 201-237. [PubMed]
- Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 17(3), 652-676. [PubMed]
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14(1), 107-141. [PubMed]
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97-136. [PubMed]
- Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research*, 42(20), 2357-2369. [PubMed]
- Tripathy, S. P., & Levi, D. M. (1994). Long-range dichoptic interactions in the human visual cortex in the region corresponding to the blind spot. *Vision Research*, 34(9), 1127-1138. [PubMed]
- Vergheze, P., & Nakayama, K. (1994). Stimulus discriminability in visual search. *Vision Research*, 34(18), 2453-2467. [PubMed]
- Vergheze, P., & Pelli, D. G. (1992). The information capacity of visual attention. *Vision Research*, 32(5), 983-995. [PubMed]
- Vergheze, P., & Stone, L. S. (1995). Combining speed information across space. *Vision Research*, 35(20), 2811-2823. [PubMed]
- Vergheze, P., & Stone, L. S. (1996). Perceived visual speed constrained by image segmentation. *Nature*, 381(6578), 161-163. [PubMed]
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33(2), 113-120. [PubMed]
- Watson, A. B., & Robson, J. G. (1981). Discrimination at threshold: Labelled detectors in human vision. *Vision Research*, 21(7), 1115-1122. [PubMed]
- Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A*, 14(9), 2379-2391. [PubMed]
- Weber, E. H. (1834). *De pulsus, resorptione, auditu et tactu. Annotationes anatomicae et physiologicae*. Leipzig: C. F. Koehler. [Translated in Weber, 1978]
- Weber, E. H. (1846). Der Tastsinn und das Gemeingefühl. In R. Wagner (Ed.), *Handwörterbuch der Physiologie III, ii* (pp. 546-549) [or 596-100 in the 1905 reprint]. [Translated in Weber, 1978]
- Weber, E. H. (1978). *The Sense of Touch* (H. E. Ross & D. J. Murray, Trans.). London: Academic Press.

- Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A*, 14(9), 2057-2068. [PubMed]
- Wilson, H. R., & Kim, J. (1998). Dynamics of a divisive gain control in human vision. *Vision Research*, 38(18), 2735-2741. [PubMed]
- Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, 9(1), 33-39.
- Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research*, 37(1), 25-43. [PubMed]
- Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 34-49. [PubMed]
- Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review*, 82(3), 184-199. [PubMed]
- Wolford, G., & Chambers, L. (1983). Lateral masking as a function of spacing. *Perception & Psychophysics*, 33(2), 129-138. [PubMed]
- Wolford, G., & Hollingsworth, S. (1974a). Lateral masking in visual information processing. *Perception & Psychophysics*, 16(2), 315-320.
- Wolford, G., & Hollingsworth, S. (1974b). Retinal location and string position as important variables in visual information processing. *Perception & Psychophysics*, 16(3), 437-442.
- Wolford, G., & Shum, K. H. (1980). Evidence for feature perturbations. *Perception & Psychophysics*, 27(5), 409-420. [PubMed]
- Woodrow, H. (1938). The effect of pattern upon simultaneous letter-span. *American Journal of Psychology*, 51, 83-96.
- Woodworth, R. S. (1938). *Experimental psychology*. New York: Holt.
- Woodworth, R. S., & Schlosberg, H. (1954). *Experimental psychology* (Rev. ed.). New York: Holt.