A letter in the peripheral visual field is much harder to identify in the presence of nearby letters. This is called “crowding”. In general, “masking” refers to any effect of a “mask” pattern on the discriminability of a signal. Crowding conforms to the masking paradigm, but the crowding effect is unlike ordinary masking. Here we characterize crowding, and present diagnostic tests that 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, 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, seven 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 feature integration gone awry — detected features are integrated over an inappropriately large area because smaller integration fields are absent — so the signal is ambiguous, jumbled with the mask. 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 phenomenon that one might call “temporal crowding,” which depends on time pressure, independent of spatial proximity.

Keywords: Crowding; Masking; Peripheral vision; Second order mechanisms; Feature integration; Illusory conjunction; Critical spacing; Letter identification; Object recognition; Integration field.

* Current address: Dept. of Psychological and Brain Sciences, Johns Hopkins University, Baltimore, MD 21228.
overlaps the signal. However, in the normal periphery or in amblyopic fovea, neighboring letters with no overlap severely impair the identification of a signal letter (Korte, 1923; Ehlers, 1936, 1953; Anstis, 1974; Bouma, 1970; Flom, 1991). This particular masking phenomenon is called “crowding” (Stuart and Burian, 1962), but it is still unclear how crowding happens.

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. The image that matches a detector’s receptive field is called its “feature”. All the 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 and Graham, 1981).

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

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. The image that matches a detector’s receptive field is called its “feature”. All the 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 and Graham, 1981).

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

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. The image that matches a detector’s receptive field is called its “feature”. All the 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 and Graham, 1981).

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

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. The image that matches a detector’s receptive field is called its “feature”. All the 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 and Graham, 1981).

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 in order to respond correctly (see Chubb, Olzak, and 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 won’t address that here. We will suggest that crowding is feature integration gone awry, integrating over an inappropriately large area that includes the flanking mask as well as the signal.

This Introduction presents a simple intuition (Sec. 1.1) that brings together ideas about feature detection (1.2) with facts of ordinary masking (1.3) and crowding (1.4, 1.5). Later, in Discussion, we will review the close connection between crowding and illusory conjunction (4.4).

1.1 Overview

Here we attempt to characterize crowding, and distinguish 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. A definitive test is proposed in Discussion (Sec. 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. It is the ambition of achieving a comprehensive account that forces upon us what is new in our model.

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 can’t 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, and 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. Tasks that require only a single feature-detection event are immune, or nearly so. This parallels the dichotomy found in searching for one feature vs. a conjunction of features, and is strong evidence that crowding interferes with feature integration, not feature detection. The multiple detections must be integrated, which 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, and 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 and Foley, 1980; Foley and Chen, 1999; Wilson and Kim, 1998; see footnote 2). A new finding, the measured effect of mask contrast as a function of spacing (Sec. 3.6), provides strong evidence that, in crowding, distinct feature detectors mediate the effects of mask and signal.

In Discussion, we’ll survey the literature on illusory conjunctions. Most of the papers’ results are consistent with crowding, as defined here, but a few papers, including Treisman and Schmidt (1982), describe a different phenomenon that we’ll 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 conjunction of features must be based on an integration (combination) of several detected features (e.g. Selfridge, 1959; Neisser, 1967; Campbell and Robson, 1968; Thomas et al., 1969; Treisman and Gelade, 1980; Sagi and Julesz, 1985; Olzak and Thomas, 1986). Despite its appealing simplicity, feature detection has been hard to establish convincingly. The grating detection literature is convincing (e.g. Campbell and Robson, 1968; Robson and Graham, 1981; Graham, 1989), but that leaves open the possibility that other tasks and targets — e.g.

---

4 Whether to call it a dichotomy has become controversial, as there is a continuum of intermediate cases (e.g. see Nakayama and Silverman, 1986).
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, and Moore, 2003). 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.

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 and Movshon, 1982) and some MT neurons actually implement this combination rule (Movshon, Adelson, Gizzi, and Newsome, 1986). Speed discrimination is affected by whether the components are perceived to form an object (Verghese and 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, and Vaina, 1995; Graham and Sutter, 1998). Texture discrimination has been explained by supposing linear combination of nonlinearly transformed feature detection signals (for review, see Chubb, Olzak, and Derrington, 2001). Visual search and crowding experiments have also contributed hints, as we’ll 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 from the alignment approach championed by Poggio and Ullman (see Tarr and Buelthoff, 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 and orientation as the signal (Legge and Foley, 1980; Phillips and Wilson, 1984). Critical-band masking experiments have shown that the spatial frequency tuning of grating detection (Greis and Rohler, 1970; Stromeyer and Julesz, 1972; Solomon and Pelli, 1994) and letter identification (Solomon and Pelli, 1994; Majaj, Pelli, Kurshan, and Palomares, 2002; Chung, Levi, and Legge, 2001) is about two octaves wide. And it’s independent of eccentricity, having the same tuning in central and peripheral vision (Mullen and Losada, 1999).

Ordinary masking has very similar effects on detection and identification (Thomas, 1985a, b). 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’s 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., 2003) 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; Levi, Pelli, and Chung, in preparation). 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 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. We usually assume that visual object recognition takes in a section of the scene and accounts
for it by hypothesizing an “object” with appropriate properties. One supposes that all the object’s properties are estimated from the same section of the image. Surprisingly, this demo shows that a single object’s various properties can be estimates from regions of very different sizes. 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’ll 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 independently estimated properties. This demo, like the rest of the 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 and Burian (1962). Acuity is greatly impaired by crowding (Ehlers, 1936, 1953; Woodworth, 1938; Flom, Weymouth, and Kahneman, 1963; Bouma, 1970).
identification: their signals were still detectable when they could no longer be identified.


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 min arc, that one would expect from its proportionality to eccentricity.

Lateral masking studies with larger signals find no effect of non-overlapping flankers on foveal targets (Strasburger, Harvey, and Rentschler, 1991; Leat, Li, and Epp, 1999). Bondarko and Daniola (1997) showed that non-overlapping 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 five times the gap width of a Landolt C (Flom, 1991) and three times the wavelength of a grating (Polat and Sagi, 1993; Levi, 2000). This scaling with signal size is characteristic of ordinary masking and unlike crowding.

Since our experiments are done mostly with letters, we postpone until Discussion (Sec. 4.2) the rest of our review of crowding with other stimuli. What we’ve 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 (Sec. 4.4).

1.5 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). 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 and Levi, 1992). Most relevant to this conclusion are the effect of task and the combined effects of mask contrast and spacing.

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 executed 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 Committee on Vision (1980). Sloan letters were usually 0.32 deg high and wide. Sinusoid 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 and Zhang, 1991). The fixation point was a 0.15 deg black square. 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 (Fig. 1). 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 ten 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.

The position of the fixation point on the screen determined the eccentricity of the signal (always presented at the center of the display). Thus the signal was presented at various eccentricities along the horizontal meridian in the right visual field.

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

Figure 1. Typical condition for crowding. The black square is a 0.15 deg fixation mark. Signal is a 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 to best estimate threshold contrast.

| Table 1. The experiments. For gratings, “size” is the 1/e radius of the Gaussian envelope, and the observer “identified” the ±45° orientation. Re Fig. 8: observer MCP was tested at 4 instead of 6 deg eccentricity. |

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

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 (Fig. 2). Threshold elevation (a ratio) is measured from floor to ceiling. Critical spacing is the least spacing at which there is no threshold elevation in the fit (i.e. edge of the floor).
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).

3. Results

Figs. 3 to 15 present our results. The reader needn’t struggle to retain it all, as the Results section closes with a list of the nine empirical differences between crowding and ordinary masking (Table 2). We recommend focussing on the sheer strangeness of crowding. Our intuitions, based on familiarity with ordinary masking, were defied at every turn.

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 function 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. Fig. 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. Fig. 3b shows that critical spacing is proportional to eccentricity. Our data confirm the finding (Bouma, 1970; Strasburger et al., 1991; Toet and Levi, 1992) 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’ll 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 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 at various eccentricities. 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 3 observers. Like Bouma (1970), we find that critical spacing is half of viewing eccentricity. Observers MCP, SJR, and SSA.
Fig. 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 seven 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 out to greater eccentricities than to lesser eccentricities (Townsend, Taylor, and Brown, 1972; Wolford and Shum, 1980; Banks, Larson, and Prinzmetal, 1979; Toet and Levi, 1992).

**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 flanker of the same size, as a function of spacing (i.e. flanker position). 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.
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. Does the range of crowding scale with the size of the letter? We measured threshold contrast for letters of various sizes at 4 deg viewing eccentricity. Fig. 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). Fig. 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 (Fig. 5c) since, as Fig. 5a shows, the ceiling threshold remains constant at about 0.7, while the floor threshold decreases with size. This is just the familiar fact that contrast sensitivity for letters depends on size (see Pelli and Farell, 1999).

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 3 observers, showing no effect. Average (horizontal line) is 1.2 deg. c. Threshold elevation increases somewhat with size: log-log slope of 0.6.
3.3 Effect of flanker size

We also measured threshold as a function of spacing for a 0.32 deg signal with masks of several sizes, from 0.32 to 3.2 deg. 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, Fig. 6 shows that the threshold curves nearly superimpose, hardly affected by mask size, retaining a critical spacing of about 1.3 deg. (Fig. 6a is for one observer; Fig. 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 (Fig. 5b) and mask size (Fig. 6), depending solely on eccentricity (Fig. 3b).

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.
3.4 Effect of font

We wondered whether perimetric complexity (perimeter, squared, over ink area, Pelli et al., 2003) or some other aspect of letter shape is important for crowding. Fig. 7a shows threshold as a function of spacing for several fonts, including a meaningless alphabet of twenty six 2x3 checkers, e.g. d, 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. (2003) showed that efficiency for letter identification is inversely proportional to perimetric complexity, but complexity seems to be irrelevant to crowding. Figs. 7b and c plot critical spacing and threshold elevation as a function of complexity, showing no systematic effect of complexity.

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, 2x3 Checkers d, 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.
3.5 Effect of number of flankers

Would adding more flankers increase the crowding effect? Fig. 8a plots threshold for letter identification in the periphery with 1, 2, and 4 flankers. Fig. 8b shows that critical spacing is independent of number of masks. It’s about 0.4 of the eccentricity. Fig. 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.

For a signal 6 deg to the right of fixation, we find a smaller critical spacing for flankers above and below, SKR, 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 in the orthogonal direction than than in the radial direction to the fovea.

Figure 8. Effect of number of flankers. Signal is 0.32 deg Sloan letter at 6 deg in the right visual field. a. Threshold contrast as a function of spacing, for 1, 2, or 4 flankers. The horizontal line at the bottom left of the graph represents the width of the signal. The single flanker was to the right of the signal. The two flankers were either to the left and right or above and below. The four flankers were above, to the right, below, and to the left. Note that, lacking data at zero spacing (mostly 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. b. Critical spacing (estimated separately for each condition) 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.

### 3.6 Effect of flanker contrast

The experiments presented above used flankers of a high contrast, 0.85. Fig. 9 demonstrates the effect of mask contrast, showing a fairly 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, Levi, and Legge (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.

Fig. 10a 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 three 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 (Fig. 10c, 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, Fig. 10b 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 9. 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.](image-url)
The series of functions plotted in Fig. 10a reveal something quite remarkable. It is hardly surprising that the threshold elevation (on the vertical scale) is reduced at greater spacings, as shown in Fig. 10c. But we were surprised to find that the critical mask contrasts (on the horizontal scale) are unaffected by the spacing. In Fig. 10a, 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. Fig. 10d shows explicitly that the critical mask contrasts are independent of spacing. We’ll come back to this in Discussion.

5 Comparing Figs. 8a and 10c, you’ll note that we’ve fit them differently, with a ceiling at small spacing in 8a and still rising in 10c. However, that difference is spurious. The data in the two figures are very similar, showing enough scatter to accept either fit.
Figure 10. Effect of flanker contrast, for 3 observers identifying a 0.32 deg Sloan letter at 4 deg in the right visual field.  

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., 2003). The lower asymptote is 1/10 because that is the chance of correctly guessing the identity of one of ten letters.  

c. The fits in panel a (and similar data for observers MCP and AG) have high threshold elevation (left) and log-log slope (right) 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 initially increases. 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 (panel a) or zero (panel b) mask contrast. Observers MCP, AG, and MLL.
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, Fig. 11 shows identification and detection thresholds for a letter among letters as a function of flanker spacing for two observers. (Fig. 15a shows similar results for a third observer.) For identification, averaging across the three observers, the threshold elevation is large (tenfold) and extends out to 1.3 deg (four signal widths). For detection, the threshold elevation is only 3-fold but extends about as far (average is 1.5 deg).

![Graph showing identification and detection thresholds](image)

We measured the effect of eccentricity (2, 4, and 8 deg in right visual field) on detection thresholds for 0.75 deg Sloan letters. Fig. 12a plots threshold as a function of spacing for each eccentricity. Fig. 12b shows that the critical spacing for detection is independent of eccentricity, unlike the proportionality found for identification (Fig. 3b).

We also 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 Fig. 13b show that the critical spacing for letter detection is proportional to size, unlike for letter identification where it’s independent of size (Fig. 5b).

To us, this is the most telling difference: in ordinary masking (e.g. letter detection) the critical spacing is proportional to signal size (Fig. 13b), independent of eccentricity (12b), whereas in crowding (e.g. letter identification) the critical spacing is proportional to eccentricity (3b), independent of size (5b).
Figure 12. 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 Fig. 3b. Observer MLM.

Figure 13. 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 Fig. 5b. Observer MCP. Not shown: similar results for observer MLL.

We also changed the envelope size of 1 c/deg gratings in a ±45° orientation discrimination task at 20 deg viewing eccentricity (Fig. 14a). We saw earlier (Fig. 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 (Fig. 14b). There’s no
mystery here: The gratings mask each other only when they overlap; at their critical spacing they are abutting.

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 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 were 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 ±45° 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° or -45° orientation.

Figs. 15c & d 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 Figs. 15c & d we were led to think, wrongly as it 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 the feature detectors that allow us to see gratings have an orientation bandwidth of ±15° to ±30° (Phillips and Wilson, 1984). Thus orthogonal gratings are detected by distinct feature detectors, and we would expect the label of the feature detector to suffice for identifying the coarse orientation of the grating. Indeed, the thresholds for detection and identification in Figs. 15c & d 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 their 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 (see Verghese and Nakayama, 1994).
3.9 Summary

Table 2 summarizes the difference between crowding and ordinary masking.

Figure 15. 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 Figs. 11ab and 15a, 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.
### 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. 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 (Fig. 13b), not eccentricity (Fig. 12b). e. Figs. 11, 15a, b are examples of weak effects of nonoverlapping masks in ordinary masking. f. “Roughly half” can be as low as 0.3, as in Fig. 5b. 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). 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’s not certain whether their paradigm elicited crowding or ordinary masking (see Sec. 4.6).

<table>
<thead>
<tr>
<th>Fact</th>
<th>Ordinary masking</th>
<th>Crowding</th>
<th>Figs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>Similar in fovea and periphery.</td>
<td>Normally evident only in the periphery (Korte, 1923; Stuart and Burian, 1962; Flom et al., 1963b; Bouma, 1970).</td>
<td>3, 4</td>
</tr>
<tr>
<td>b</td>
<td>Signal disappears, suppressed by mask.</td>
<td>Signal is visible but ambiguous, incorporating features from mask (Korte, 1923; Flom et al., 1963b; Andriessen and Bouma, 1976; Wolford and Shum, 1980; Wilkinson et al., 1997; Parkes et al., 2001).</td>
<td>1, 9</td>
</tr>
<tr>
<td>c</td>
<td>Occurs for any task and signal.</td>
<td>So small, specific to identification of letters (Flom, 1991) and fine discrimination of contrast, spatial frequency, and orientation (Andriessen and Bouma, 1976; Wilkinson et al., 1997; Parkes et al., 2001).</td>
<td>11 - 15</td>
</tr>
<tr>
<td>d</td>
<td>Similar effect on identification and detection.</td>
<td>Little or no effect on detection (Wilkinson et al., 1997) and coarse discrimination.</td>
<td>11 - 13, 15</td>
</tr>
<tr>
<td>e</td>
<td>Narrow critical spacing, little or no effect of nonoverlapping mask.</td>
<td>Wide critical spacing can be more than ten times bigger than a small signal (Korte, 1923; Stuart and Burian, 1962; Bouma, 1970; Toet and Levi, 1992).</td>
<td>3, 4</td>
</tr>
<tr>
<td>f</td>
<td>Critical spacing scales with signal, independent of eccentricity.</td>
<td>Critical spacing is roughly half of viewing eccentricity (Bouma, 1970; Toet and Levi, 1992), independent of signal size (Strasburger et al., 1991), mask size, and number of masks.</td>
<td>3 - 8, 12 - 14</td>
</tr>
<tr>
<td>g</td>
<td>Spatiotemporal selectivity more or less consistent with a receptive field.</td>
<td>Remarkably unselective, showing equal effect over a wide range of flanker type (letter, black disk or square; Eriksen and Hoffman, 1973; Loomis, 1978), flanker size (10:1), and flanker number (≥2).</td>
<td>6, 8</td>
</tr>
<tr>
<td>h</td>
<td>Shallow power-law contrast response (log-log slope of 0.5 to 1).</td>
<td>Steep sigmoidal contrast response. Log-log slope of 2 for close spacing. Log ceiling and log slope fall exponentially with spacing.</td>
<td>9, 10</td>
</tr>
<tr>
<td>i</td>
<td>Threshold mask contrast depends on spacing. No saturation.</td>
<td>Threshold and saturation mask contrasts are independent of spacing.</td>
<td>10</td>
</tr>
</tbody>
</table>

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 and Schmidt (1982) make a similar suggestion for illusory conjunction. c. Treisman (1991) makes a similar suggestion for illusory conjunctions. e. 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. f. 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.
4. Discussion

Table 2 summarized the empirical findings. Table 3 summarizes the theoretical conclusions. We begin the discussion by proposing a definition.

4.1 A definitive test for crowding

Using published and new results, we have established that the original crowding task — identifying a letter among letters in the periphery — is unlike ordinary masking. We suggest that the term “crowding” should also be applied to other tasks that exhibit the same symptoms. We suggest refraining from indiscriminately calling all lateral masking “crowding”, as letter-letter crowding has several striking properties that are not shared by, for example, the lateral interactions in grating detection (Sagi, 1990; Polat and Sagi, 1993), and these special properties make the crowding phenomenon seem a very promising probe into the feature integration process.

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 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?

Stuart and Burian (1962) coined the term “crowding” for the impairment of identification of a peripheral letter by neighboring letters. Some subsequent investigators have tentatively applied the term “crowding” more broadly to describe the effect of laterally displaced masks on other kinds of target, i.e. the orientation discrimination of a grating among gratings (He, Cavanagh, and Intrilligator, 1996; Parkes, Lund, Angelucci, Solomon, and Morgan, 2001) and vernier acuity (Levi and Klein, 1983, 1985; Levi, Klein, and Aitsebaomo, 1985) and Landolt C acuity (Hess, Dakin, Kapoor, and 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 haven’t 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 (Bowma, 1970; Toet and Levi, 1992). The flanking bars in Landolt C and vernier tasks mask the signal only at intermediate spacing (five times the gap width of the C), losing effectiveness at close and far spacing (Flom et al., 1963b; Flom, 1991; Jacobs, 1979; Levi, Klein, and Aitsebaomo, 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, and Carney, 1990; Levi, Klein, and Aitsebaomo, 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 and Morgan, 2000; Sagi, 1990), but it seems unlike crowding and we suspect it will fail our test (Sec. 4.1).

Besides identification of a letter among letters, what other conditions yield crowding? In Figs. 14 and 15 the task was identification (i.e. discrimination) of a grating’s ±45° 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. Parkes et al. (2001) also tested a grating among gratings, and they reproduce their stimulus as a figure, allowing the reader to assess the pure effect of eccentricity by changing fixation. With eccentric viewing, the grating loses its tilt (amalgamating the mask...
orientations), but remains visible. This satisfies the initial part of our proposed test for crowding.

In an elegant adaptation paradigm, He et al. (1996) showed that a grating whose orientation could not be identified because of crowding by other gratings still produced orientation-specific adaptation. But don’t let their suggestion that crowding “restrict[ed] … conscious awareness” mislead you to think that the grating disappeared. Disappearance would be unlike crowding and inconsistent with Wilkinson et al. (1997). In fact, a glance at their stimulus, in their paper, shows that the target grating remains visible, even though one cannot accurately report its orientation.

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.

4.3 Integration field (Table 3a)

Why does the visual system do something so silly as to integrate the features of a remote flanker into the signal letter? We don’t usually 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.6 Why extend it perversely for signals in the periphery? Our guess is that the visual system has many integration fields of various sizes, 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 radius of half the eccentricity (or more), allowing a response based on feature conjunctions anywhere in that integration field. This may seem at once familiar and fanciful: Familiar because it overlaps with popular existing theories, specifically feature detection (Graham, 1980) and feature integration (Treisman and Gelade, 1980). It differs from Treisman (and He et al., 1996) in attributing the peripheral (and amblyopic) 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 and Bouma, 1976; Wilkinson et al., 1997; Hazeltine, Prinzmetal, and Elliott, 1997).

The new theory may seem fanciful in making 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 and Levi, 1992). 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. Receptive fields detect features, which are subsequently combined by integration fields.

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 Sec. 1.4 we showed that different properties of the same object may be based on regions of very different size. It seems that the region an object property is based on 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, and Takahashi, 1963a; Taylor and Brown, 1972; Tripathy and 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 mm at the cortex, which roughly matches the length of horizontal connections in V1 (Gilbert, Ito, Kapadia, and Westheimer, 2000). However, the half-the-eccentricity critical spacing is just as good a match to the receptive field radius of V4 neurons (Desimone and Schein, 1987;
Desimone, Schein, Moran, and Ungerleider, 1985; Piñon, Gattass, and Sousa, 1998). Thus, crowding occurs somewhere in the visual cortex, but it's hard to be more precise than that.

### 4.4 Illusory conjunctions (Table 3b)

Treisman and Schmidt (1982) coined the term “illusory conjunctions” to describe the impression of seeing nonexistent objects that each combine features from several real objects in the stimulus (for antecedents see Snyder, 1972; Wolford, 1975; Estes, Allmeyer, and Reder, 1976; Bjork and Murray, 1977; Krumhansl and 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.

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 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 separations less than roughly half the eccentricity. This seems to be crowding. However, some experimental results violate 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 (21, 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 8 papers, below the line, including the one that introduced the term “illusory conjunction” (Treisman and Schmidt, 1982), which are obviously inconsistent with Bouma’s bound and presumably describe some other phenomenon, not the spatial crowding we’ve been discussing here. There is a hint of a difference in the introspective reports. Recall that the descriptions of crowding (including the Sec. 1.4 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.”

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

Table 4. A representative list of papers on illusory conjunction among letters, plus antecedent work, before 1982. (One paper, Donk 1999, appears twice, above and below the line. Thus there are 21 papers above the line, 8 below the line, and 28 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
Treisman and Schmidt 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 primitive 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 migrations of a tick mark (plausibly a primitive feature) was affected differently by stimulus and task than were migrations 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 5 objects (e.g., digits and colored letters) presented in a single brief display. The various experiments used different 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 and Schmidt, 1982, p. 113). In their Exp. 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 (2003) show that the identification of a letter seems to require the independent detection of a modest number of primitive 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 primitive features. In the same way, reading rate is largely independent of contrast, but slows near threshold contrast (Legge, Rubin, and Luebker, 1987). All the experiments in Table 4 that violate Bouma’s bound loaded attention by demanding that the observer examine and report on many hard-to-see objects in a single glimpse.

The conditions in Table 4 that violate Bouma’s bound seem closely related to the phenomenon of “temporal migration” (James, 1890; Lawrence, 1971; Sperling and Reeves, 1980; McLean, Broadbent, and Broadbent, 1983; Gathercole and Broadbent, 1984; Intraub, 1985, 1989; Shimomura and Yokosawa, 1998; Botella, Barriopedro, and Suero, 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 report a target element as having been presented in the same display as a 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 presented in a particular color, in a particular letter case, or simultaneous with a black frame. ... subjects 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) suggest that the increased presentation rate needed to produce migrations reflects the fact that it takes much less time 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. 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.”
Temporal crowding but is basically consistent with their interpretation. “temporal crowding” is a different way to tell their tale, contrast, making it harder and slower to identify. 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. 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. 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.

There seems to be little or no effect of attention on spatial crowding. Nazir (1992) showed that precuing signal location did not increase identification rates of the signal (Landolt C) in the periphery among flankers (bars, E’s, or O’s). 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).

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.

Temporal crowding does provide evidence for Feature Integration Theory, as Treisman and Schmidt say, but spatial crowding does not. Spatial crowding occurs even when there is no attentional loading at all, as in the Sec. 1.4 demo (or Woodworth and Schlosberg, 1954, p. 104; Banks, Larson, and Prinzmetal, 1979). In Table 4, the paradigms that are consistent with Bouma’s bound seem likely to all be spatial crowding. It is interesting and worthwhile to determine how attention-related factors modulate spatial crowding, as many of these studies do, but it’s not clear to us that such studies have any bearing on Treisman’s Feature Integration Theory.

It is dismaying that 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. We suspect that everyone was misled by a passing comment of Treisman and Schmidt (1982, p. 113) about how to produce illusory conjunctions. They explicitly minimized the duration of presentation to overload attention. They don’t spell it out, but presumably they thought that this effect was mediated by a correspondence between the physical presentation time and the observer’s internal time available for processing. They do 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 against the risk of introducing data limits.” Contrary to their interpretation, we suspect that minimizing the duration affected the resource limit (and probability of illusory conjunction) primarily by limiting data. In our interpretation, their experiments depend on reducing the duration or contrast of the objects to near threshold in order to increase the required processing time per 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 hard-to-see objects in a glimpse, or spatially, by putting two objects within Bouma’s bound, most easily in the periphery. Not suspecting that making the 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 accompanied by two digits presented peripherally (following Treisman and Schmidt) or centrally. The digits are the primary task; the letters are the secondary task; presumably the observer attempts to process all four objects. When the digits are central, Cohen and Ivry get temporal crowding, i.e. they replicate Treisman and Schmidt’s finding that there is little or no effect of separation among objects, violating Bouma’s bound. When the digits are central, Cohen and Ivry get spatial crowding, i.e. their results are consistent with Bouma’s bound. Cohen and Ivry attributed the difference in results to whether the letters are inside or outside the “attentional spotlight” spanning the digits. Our
interpretation of their finding is that the digits are easy to see centrally, so they are processed quickly, the observer is not overloaded, and there is no temporal crowding; moving the digits to the periphery reduces their visibility and thus increases their processing time enough to overload the observer, producing temporal crowding. Spatial crowding between the letters is unaffected by the location of the digits.

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 8 papers that violate 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.”

4.5 Task-specific: No crowding of detection (Table 3c)

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 (Figs. 12 & 13).

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. As noted in Section 3.9, one-feature-detection-event tasks (like detecting a gratings 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 and Robson, 1968; Graham, 1980, 1989; Watson and Robson, 1981), and the evidence that letter identification is mediated by feature detection (Pelli et al., 2003; Solomon and 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 and 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.

4.6 Selectivity (Table 3d,e)

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 and 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; Fig. 6), and flanker number (≥2; Fig. 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. Banks, Larson, and Prinzmetal (1979) and Baylis and Driver (1992) show lessened interference of letter identification by flanking letters of different grouping (e.g. color). Kooi et al. (1994) measured crowding of a T among T’s, 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. Further measurements of the selectivity of crowding could help characterize the computation performed by the feature integrator.

Nazir (1992) varied the number of features in the flankers by making the flankers more complex (e.g. E is more complex than C and I). She expected, as in visual search, that increasing the number of distracter features would further impair subject performance. However, like Leat, Li, and Epps (1999), she found that the impairment of signal acuity is independent of the number of flanker features.

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., 2003). 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, Levi, and Legge (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 gratings 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 Fig. 6. The experiments differ in important ways, and it’s hard to know which differences account for the qualitative difference in selectivity of masking. Firstly, 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. Secondly, 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. 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 Sec. 3.7. However, we also found ordinary masking for letter detection at approximately that size and spacing (Fig. 13). (This was at a larger eccentricity, but eccentricity has no effect on ordinary masking, Fig. 12.) 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).

4.7 Mask and signal detected separately (Table 3f)

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 (Fig. 10d, Table 2h). 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 (Fig. 10a; 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, Fig. 10b shows the proportion correct identification of an isolated letter. The curves in Figs. 10a & b 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; Fig. 7c) and no effect of number of masks, beyond two (Fig. 8c).

5. Conclusion

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

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 provides a window into the inner workings of the feature integrator.

A single object’s various properties can be estimates based on regions of very different sizes: a small region for one-feature properties (e.g. presence or 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 phenomenon that one might call “temporal crowding,” which depends on time pressure, independent of spatial proximity.

7. Acknowledgements

Thanks to Gordon Legge and Dennis Levi 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 crowding, to Susana Chung, Greg Murphy, Anne Treisman, and an anonymous reviewer for their helpful critiques, to Bart Farell and Marialuisa Martelli for discussion, to Gina Cardazone for pilot foveal data, to Joshua Green for pilot gratings-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, et al., 1999; Pelli and Palomares, 2000). Supported by National Eye Institute grant EY04432 to Denis Pelli.

8. References


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)).


