The uncrowded window of object recognition

Denis G Pelli & Katharine A Tillman

It is now emerging that vision is usually limited by object spacing rather than size. The visual system recognizes an object by detecting and then combining its features. 'Crowding' occurs when objects are too close together and features from several objects are combined into a jumbled percept. Here, we review the explosion of studies on crowding—in grating discrimination, letter and face recognition, visual search, selective attention, and reading—and find a universal principle, the Bouma law. The critical spacing required to prevent crowding is equal for all objects, although the effect is weaker between dissimilar objects. Furthermore, critical spacing at the cortex is independent of object position, and critical spacing at the visual field is proportional to object distance from fixation. The region where object spacing exceeds critical spacing is the 'uncrowded window'. Observers cannot recognize objects outside of this window and its size limits the speed of reading and search.

Object recognition means calling a chair a chair, despite variations in style, viewpoint, rendering and surrounding clutter. Crowding is a breakdown of object recognition.

Let us begin by sketching a popular two-step model of object recognition: feature detection and combination. Features are components of images that are detected independently[1–4]. They are typically simple and nonoverlapping. The first step in object recognition is feature detection[4]. Each neuron in the primary visual cortex responds when a feature matches its receptive field. Only the features that drive neurons hard enough are detected[5]. In the second step, the brain combines some of the detected features to recognize the object. This combining step (including 'integration', 'binding', 'segmentation', 'pooling', 'grouping', 'contour integration' and 'selective attention') is still mysterious[3,4,6–11].

Some objects are recognized through a single combining of features over the whole object, whereas other objects require separate combining over each of several regions of the object[12–14]. These distinct regions define object parts. In an object with multiple parts, each part must be recognized before they are all joined together.

The best evidence that features are indivisible elements that we detect and combine is that, even with practice, people combine information across features much less well than within a feature. Searching for a conjunction of several features is usually much harder than searching for a single feature[3]. Despite reading a billion letters over a lifetime,

people still recognize letters inefficiently, by detecting and combining many simple features rather than by detecting each letter as a whole[4,15]. Crowding is inappropriate feature combination that spoils object recognition (reviewed in refs. 16,17).

This is an empirical review of crowding in object recognition. Science, in its many styles, creates theory to bind facts into an intelligible whole. This whole, as W.V.O. Quine noted, is a continuum from fact to theory. Broad empirical generalizations, such as those we present here, lie near one end of the continuum, with full explanatory models being present at the other end. Unlike a mature field such as physics, object recognition is an immature topic with only tentative theories, as scientists are still describing the empirical phenomena. A review of the "scattered and diverse" theoretical models of crowding in object recognition finds "a growing consensus" for the two-step account of feature detection and combination[16]. That account does not specify how the crucial combination happens and mostly serves to provide a vocabulary for describing results. This empirical review passes over the details of the diverse models to provide a broad survey of the underlying results, which we find notably consistent. We boil the results down as far as we can, achieving a short synthesis that we call the Bouma law. It binds together most of the facts on crowding and seems to be a useful step toward the computational model of recognition that we all yearn for.

This empirical review includes visual demonstrations that allow the reader to experience the phenomena. The bars in the 'A in chaff' demonstration (**Fig. 1**) represent elementary features. When you look at the demonstration, your brain detects the features and combines them to categorize the letter as A. We cannot yet explain how this process works, but we can easily break it. Fix your eyes on the red minus, far from the A, and the extra features (chaff) make it impossible to recognize the A. When you fixate this far from the A, your brain combines features over too large an area around the A, failing to isolate the relevant features of the A from the nearby junk, and comes up with a jumbled percept instead of a letter. This is crowding. Some well-known illusions are delicate, strongly affected by expectation and only work once. Unlike them, crowding is robust. No matter how many times you move your eyes back and forth from plus to minus, the A quickly comes and slowly goes away every time.

Crowding, unlike overlap masking (ordinary masking by nearby objects that overlap the target), never makes the target disappear[17]. Crowding impairs our ability to identify, count and locate objects, but does not affect detection (**Fig. 2**). As you can see, the jumbled percept produced by crowding looks like inappropriate combining rather than a failure to detect. The notion that crowding is a breakdown of the second step of object recognition, after feature detection, is consistent with experiments showing that crowding can knock out the observer's ability to judge target orientation while sparing (or largely sparing) the orientation-specific aftereffect of adapting

Department of Psychology and Center for Neural Science, New York University, 6 Washington Place, New York, New York 10003, USA. Correspondence should be addressed to D.G.P. (denis.pelli@nyu.edu).
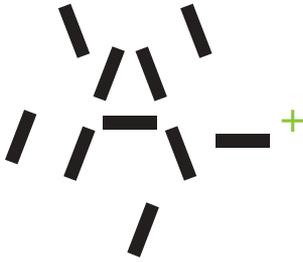
Figure 1 An A in chaff. The bars represent elementary visual features. Fixating close to the bars, at the green plus, makes it easy to recognize the letter A. If you fixate far away, on the red minus, you can still see the features, but you cannot identify the letter. Your visual system is combining over too large an area, including all the features from both the A and the surrounding chaff, which results in a jumbled percept. This is crowding. You can rule out acuity (letter size) as an explanation (for your inability to identify the A) by confirming that you can see the A while fixating the minus if your fingers hide the chaff (for a review, see ref. 17).

to that target[18,19]. Finding that we still adapt to stimuli that we cannot identify is evidence for two steps in object recognition, one (feature detection) that is susceptible to adaptation followed by another (feature combination) that is susceptible to crowding.

Crowding is usually specified by the observer's 'critical spacing'. Critical spacing is how far (measured center to center) the flanking objects ('flankers') must be from the target to allow unimpaired perception of the target. Critical spacing grows in proportion to eccentricity, the distance of the target object from fixation[20]. It has been suggested that critical spacing may reflect the spatial resolution (minimum area) of visual attention, but this is controversial[11,21,22].

Distinguishing crowding from overlap masking is easy, as the critical spacing of crowding is proportional to eccentricity, whereas that of overlap masking is independent of eccentricity[17]. Therefore, crowding dominates in the periphery and overlap masking dominates centrally[17].

Tilted flankers have a long-range effect on the perceived tilt of a foveal target. Unlike crowding and overlap masking, this 'stochastic recalibration' affects the orientation threshold, but not the contrast threshold, for orientation discrimination[23].

The Bouma law

Practically every paper on crowding reports critical spacing. Our story, here, is that (despite the great diversity of models) the results all boil down to a simple law, a generalization of an observation that Herman Bouma reported in 1970, that the critical spacing for identification of small letters is roughly half the eccentricity[20]. We take this observation to its most general form, which we call the Bouma law: for an object that can be identified in isolation, our ability to identify it among similar

Figure 3 Crowding in a word. While fixating the red minus, it is easy to identify the isolated letter on the left, but try to identify the middle letter on the right. It is hard. Fixate the green plus and try again. Now it is easy[24].

objects depends solely on the ratio of the object spacing to the observer's critical spacing at that location. The object is crowded whenever the ratio is less than one. For each observer, the critical spacing is independent of what the object is and depends only on where the object is in the visual field and the direction from target object to flanker object. The broad empirical support for this law is unexpected because object recognition is usually assumed to be limited by size, not spacing.
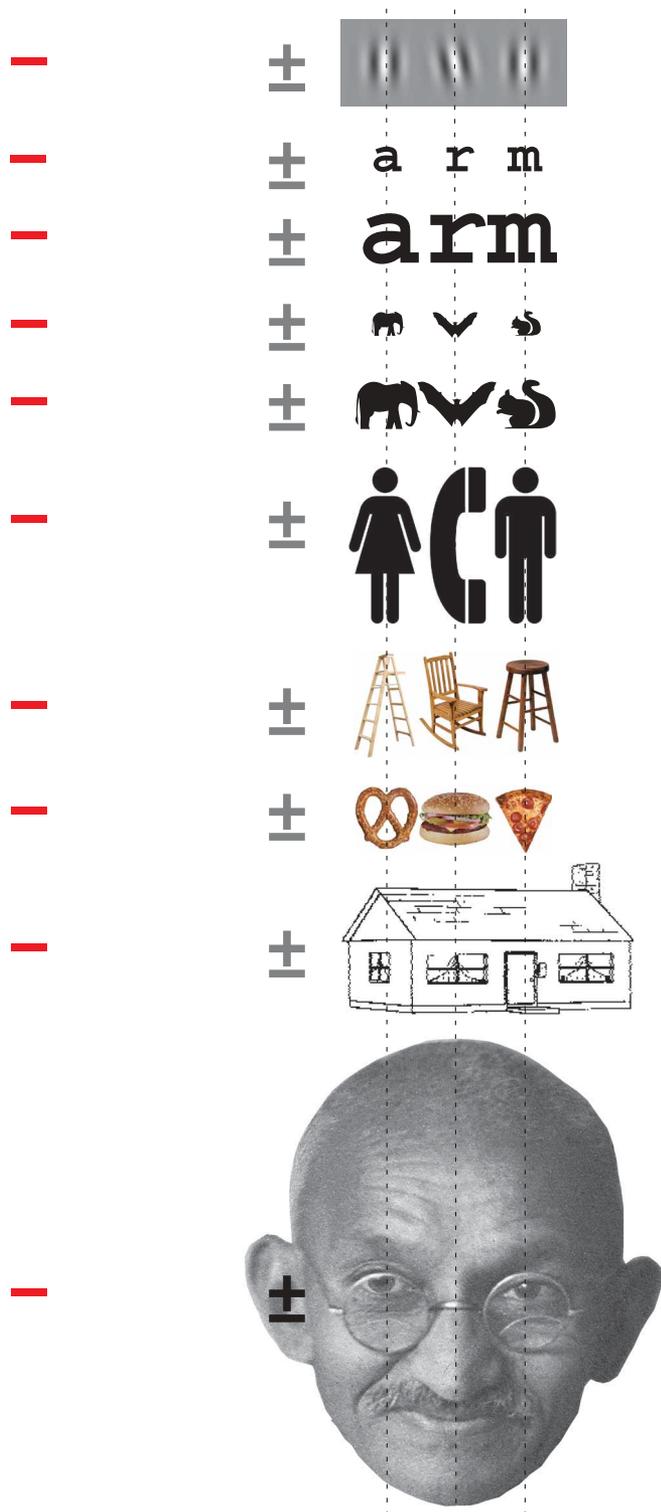
Most studies of crowding have used letters and words as stimuli. (However, a recent special issue of the *Journal of Vision* includes more than twenty articles on crowding, using a wide variety of stimuli. http://www.journalofvision.org/7/2/) **Figure 3** demonstrates the critical spacing of the letters in a word[24]. If you try to identify the middle letter in the word 'are', it is easy when you fixate near the word and becomes hard when you fixate far away. This is because, when fixation is too far away, the whole word falls within one critical spacing and features from all of the letters are jumbled together. Some objects, such as words, have parts. The parts of an object crowd each other when they are closer than the critical spacing. Faces, like words, are recognized only if the visual system can isolate their parts: eyes, nose and mouth[14]. Thus, we cannot recognize a face unless we look at or near it (**Fig. 4**).

The critical spacing is universal, independent of object and size (**Fig. 5**). The threshold eccentricity for recognition is the same for all objects with the same spacing, even when the objects are as diverse as gratings, letters, animals and furniture. Similarly, the critical spacing of crowding is unaffected by equal motion of the target and flankers[25]. Across different tasks, including discrimination of size, hue, saturation and orientation, the amplitude (maximum threshold elevation) of crowding varies, but the spatial extent of crowding is practically the same[26]. 'Second-order' letters (painted with texture) are more susceptible to crowding than 'first-order' letters (painted with homogeneous ink), but the spatial extent of crowding is the same[27].

The generality of the Bouma law suggests that the critical spacing of crowding is a fundamental parameter of human vision. It is

Figure 2 Effects of crowding. While fixating the red minus, can you tell that the clusters differ in letter identity, number and position? Crowding impairs your ability to judge these object properties[20,21]. Using your finger to cover all but the leftmost letter, you can confirm that even this most distant letter is well within your acuity (reprinted from ref. 21).

Figure 4 Faces are like words. Arnold Schwarzenegger and Elvis Presley are famous, and their faces may be familiar. Fixate on the red minus between them. Can you still recognize the governor and the King? How close to each face do you have to fixate to identify it? As you fixate closer and closer to the face, you will find that you remain unable to recognize it until you are near the cheek. As with words, the parts (eyes, nose and mouth) of faces must be isolated (separated by the observer's critical spacing) for the whole to be recognized[14].

Figure 5 Critical spacing is independent of object and size. Fixating on the red minus, you will be unable to identify the middle object in the first eight rows unless you isolate it by hiding the flanking objects with your fingers (or two pencils). In the last two rows, you will be unable to recognize the single object while fixating on the red minus. Grating patches, similar to those in the top row, are often taken to be one-feature objects. In the first row, is the middle grating vertical or tilted? The ± is our estimate of the fixation point where you can just barely identify the target. You can assess the accuracy of this threshold estimate by noting that the task is easy when you fixate to the right of the ± and hard when you fixate to the left. Critical spacing depends solely on position (and direction) in the visual field, which does not vary among rows in this demonstration. Note that halving object size has no effect on critical spacing. Critical spacing is independent of spatial frequency[46] (see **Supplementary Sources** online).

spacing at the cortex (6 mm at V1) that is independent of eccentricity (see **Supplementary Discussion** online).

Size or spacing?
The idea that spacing limits object recognition could not be simpler, but it has been very hard to accept because it displaces a firmly held belief that visibility is limited by size (acuity), not spacing (crowding). For example, an expert reviewer of a related article complained that "the presentation in terms of spacing [instead of size] … made it quite hard for me to understand".

When we view a scene from farther away, both size and spacing decrease. Viewing distance, *per se*, does not matter. What matters is the stimulus at the retina. Some visual tasks are limited by size. The Egyptians (5,000 years ago) and many since have assessed acuity of vision by the ability to distinguish the double star Alcor/Mizar in Ursa Major. Today, to measure a size threshold (acuity) that characterizes a person's vision, we ask the observer to identify a simple object, usually a letter. This measure is unaffected by crowding if done foveally, where critical spacing is only a few minutes of arc, or anywhere on a blank field. Measuring acuity is useful, especially in selecting the best optical correction. However, outside of the optometrist's office, most of us are well corrected (20/20) and, provided that there is enough contrast[28], our ability to see is more limited by object spacing than by size. We can see a bird in the sky without crowding, but most of our visual world is cluttered, and each object that we identify must be isolated from the clutter. When an object is not isolated, it is crowded, and we cannot recognize it. Isolation depends on spacing and not size. To escape crowding, the object spacing must exceed the observer's critical spacing at that location in the observer's visual field (that is, 6 mm at V1).

Critical spacing has profound effects on everyday life. Consider reading. It has long been known that reading consists of a series of eye fixations, 4 per second, rather than a continuous sweep of the eyes across the text[29]. Reading speed is independent of text size over a large 6:1 range, but drops precipitously for sufficiently small text. From ancient to modern times, this has been taken to be a size limit (acuity). Plato complained that he was asked "to read small letters from a distance". This statement shows that he both understood the concept of acuity and thought that it limited reading. In 1985, we said that, "the fairly rapid decline in reading rate for characters smaller than 0.3° is undoubtedly associated with acuity limitations"[30], but we were wrong. Reading speed depends on letter spacing and not size. Measuring with two texts, one widely and one normally spaced, at various viewing distances, it is found that reading speed drops at a particular letter spacing (in deg), independent of letter size[31]. Typographers routinely increase 'tracking' (spacing) to maintain the legibility of text when it is made smaller.

proportional to the distance from fixation (**Fig. 6**), and depends solely on position and direction in the visual field[17,20]. This proportionality matches the organization of the visual cortex. The known eccentricity dependence of the cortical magnification factor (mm on the cortex per deg of visual angle) produces a logarithmic map of the visual field on the primary visual cortex (V1). The logarithmic transformation of the proportional critical spacing at the visual field results in a fixed critical
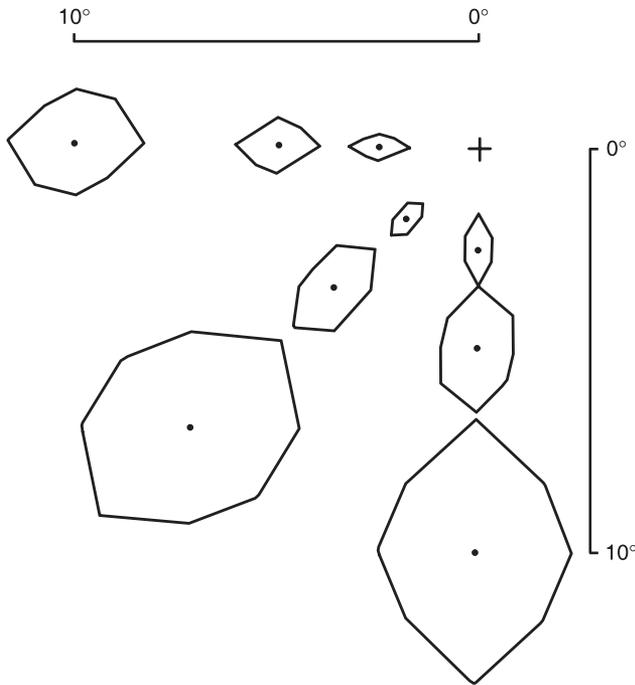
Figure 6 Critical spacing is proportional to eccentricity. The observer fixated on the point indicated by a plus in the upper right and identified the orientation of a target T (right-side up or upside down?) presented (in blocks) at one of the nine locations indicated by the dots. Two flanking Ts were shown symmetrically displaced from the target in opposite directions, –45°, 0°, 45° or 90° relative to horizontal. Each vertex in the roughly elliptical contours represents the measured critical spacing of the pair of flanking letters for 75% correct identification of target orientation. Note that the critical spacing contours are not circles; the direction from target to flanker matters. These were measured with one letter size at each eccentricity. Changing letter size has no effect on the results[28] (figure adapted from ref. 47).

Spatial extent of crowding

The invariance of critical spacing demonstrated here (**Fig. 5**) is found when the target and flankers have similar features (for example, black letters flanking a black letter target). These typical cases produce maximum crowding. Flankers that have features that are different than those of the target (for example, white letters flanking a black letter target on a gray background) produce much less crowding or none at all. This weaker effect is usually reported as a reduction in critical spacing, but perhaps the spatial extent of crowding is unchanged and the effect is only reduced in amplitude. It seems that the reported reduction of critical spacing may be an artifact of defining critical spacing by a performance criterion. Compared with the effect of

xuncrowdedx

Figure 7 What is your uncrowded span? Fixate on the o in the center of the word. Your uncrowded span is 3 if you can read 'row', 4 for 'crow', 5 for 'crowd' and a whopping 9 for 'uncrowded', which many observers achieve. The variation in the uncrowded span reflects the substantial individual differences in critical spacing reported previously[47]. The Bouma law says that critical spacing is invariant across objects, not subjects (for reviews of uncrowded and visual spans, see refs. 28,32,39). Image reprinted from ref. 28 and adapted from ref. 33.

target-like flankers, dissimilar flankers may simply have a weaker effect over the same spatial extent (see **Supplementary Discussion** for more on similarity and effects of salience, grouping, and observer practice).

At present, the simplest account is that the spatial extent of crowding for any given location and direction is independent of the particular target and flanker. That conclusion is tentative because the majority of published studies have not disentangled the amplitude and extent of crowding, but it is supported by all the studies that have done a two-parameter analysis. For the rest of this review, we revert to using 'critical spacing', asking the reader to bear in mind that special cases demand a two-parameter (amplitude and extent) characterization of crowding.

The uncrowded window

Most of our visual field is crowded most of the time, sparing only a central uncrowded window. This window and the limitation it places on recognition are especially clear in the case of reading. To read text, we must identify letters. The rate at which we read depends on how many letters we take in on each fixation (**Fig. 7**), which is limited by crowding. The spacing of letters in text is uniform, but the observer's critical spacing increases with distance from fixation. Beyond some eccentricity, the reader's critical spacing exceeds the spacing of the text and the letters crowd each other, spoiling recognition. Peripheral vision, beyond that eccentricity, is crowded. Central vision, within that eccentricity, is uncrowded: the uncrowded window. Inside of the window, letters are uncrowded and we can read them. Outside of the window, letters are crowded and we cannot. To read the letters that now lie outside of the window, we must move our eyes to bring our window to those letters. The number of character positions in a line of text that fit inside the uncrowded window is the uncrowded span[28]. Incidentally, note that letters at the ends of words are much less crowded[24] and have a larger uncrowded window.

Figure 8 demonstrates the uncrowded window by simulating crowding in the periphery. The corruptions outside the uncrowded window are undetectable when you fixate on the center of the window.

It seems that each observer's critical spacing for crowding is the same for all objects. Together, the observer's critical spacing and the spacing of the viewed objects determine the size of the uncrowded window. Inside of the window, we can recognize objects, and outside of it, we cannot[32]. When the spacing is uniform, as in text, then the window will be central, where the critical spacing is smallest. When spacing is not uniform, the window need not be central, and there may be more than

Crowded periphery   Uncrowded center   Crowded periphery

Figure 8 The uncrowded window. This figure simulates crowding in reading by substituting letters in the peripheral field. Crowding spoils letter recognition, making reading impossible outside of the uncrowded window. Note that the substitutions are undetectable when you fixate on the center of the circle. As you read this caption, the words are clear and legible near your chosen point of fixation and illegibly crowded beyond that clear region. That central uncrowded field is a window through which we read (figure adapted from ref. 28).
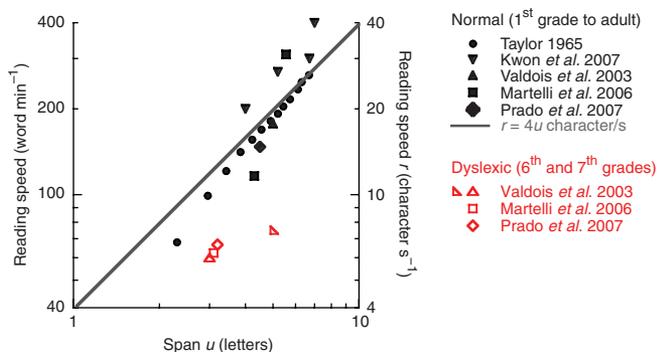
Figure 9 Reading speed versus span. Data are from five studies of normal (black filled symbols) and dyslexic (red empty symbols) readers[44,45,48–50]. The normal readers were of various ages, from 1st grade (age 6) through adult. Reading speed rose monotonically with age. The dyslexic readers were all in the 6th or 7th grades. The vertical scale is reading speed (1 word min[−1] = 0.1 character s[−1], assuming an average of five letters and a space for each word). The horizontal scale is letter span, estimated in various ways. Span is the width (in characters) of the uncrowded window. A reader making ρ eye movements per second, advancing an average of u characters per eye movement, reads at a rate $r = \rho u$ character s[−1]. The diagonal line plots this proportionality, assuming 4 eye movements per second ($\rho = 4$ Hz), showing that this simple 4 Hz rule gives a fairly good account of all the data from normal readers (see **Supplementary Methods** online).

one. Many have suggested that a central window (also known as the span of apprehension, visual span, visual attention span, area of focal attention, conspicuity area, association field or number of elements processed per fixation) limits reading or search[10,20,29,32–38] (reviewed in refs. 28,39), but they usually assumed that the window size is independent of object spacing. Often it has been supposed that the window size is limited by letter or object size (acuity), or sometimes by attention. Until recently, only Woodworth[33] and Bouma[34] claimed that the size of the window is set by spacing (crowding). They made good cases against acuity, but failed to convince their colleagues. Subsequent papers cited them, but persisted in assuming that the window is limited by acuity. However, recent detailed studies of search and reading validate the original claim, showing that the window is where the object spacing exceeds the critical spacing of crowding[28,32,40].

Following the success of the uncrowded window idea in explaining the reading speed of normal adults[28,39], one wonders whether it can help to explain why children and dyslexics read more slowly. Developmental dyslexia is now generally thought to be primarily a phonological deficit[41], but there is evidence that dyslexics have increased crowding[42].

We plotted data (**Fig. 9**) from all the studies for which we could estimate reading speed as a function of the number of characters in the uncrowded window. For all the normal readers, including both children and adults, reading speed was fairly well predicted (with no degrees of freedom) by the product of span and the standard 4-Hz rate of fixations. The large increase in uncrowded span during childhood contrasts with the small effect of practice on critical spacing (and thus uncrowded span) in adults. This warrants further investigation. Most of the dyslexics had smaller spans than age-matched controls, but they read much more slowly than is predicted by their span: They were all well below the normal line, reading at less than half of their span-predicted speed. This finding is contrary to the hypothesis that most cases of dyslexia are arrested development, with performance similar to that of younger normal individuals matched for reading level[41]. These data indicate that something else (for example, a phonological deficit

or longer fixations[43]) must account for the rest of the dyslexic impairment[35]. However, the most notable result is the accuracy of the reading speed prediction for normal readers. The normal development of reading speed seems to be mediated entirely by the uncrowded span[35,43,44].

Crowding also limits the speed of visual search. For searches in the real world (or in *Where's Waldo?*; **Supplementary Discussion**), where similarity and spacing are variable, it is helpful to trace out an uncrowded neighborhood relative to the target, the area in which you must fixate to see the target without crowding. This is the inverse of the uncrowded window, which is defined relative to the observer's fixation point. The size of the uncrowded neighborhood limits search rate (**Supplementary Fig. 1** online).

DISCUSSION
Peripheral vision and texture
Typically, only a small portion of the visual field falls in the uncrowded window. Most of our visual field is peripheral and crowded and cannot recognize objects. If we cannot recognize things in this part of our vision, what do we see? We see stuff (unnamed texture) and perceive space (the shape of the scene we are in). With an effort, observers can name and describe texture, but this rarely happens. Texture includes variations of color, depth and motion[8]. Many of the cues to depth (binocular disparity, motion parallax, scale gradients and shape from shading) seem to be immune to crowding. A sense of space is particularly important for mobility, which is greatly impaired by tunnel
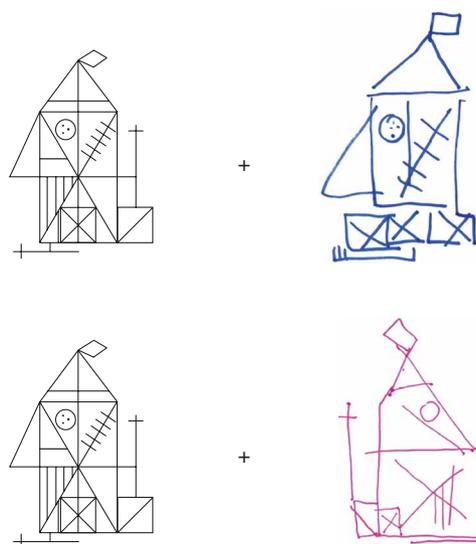

Figure 10 The Rey Complex Figure Test. The original diagram is on the left. The drawings on the right were made by normally sighted graduate students who were asked to copy, from left to right, while fixating on the central + (ignore the left-right reversal, which was the result of ambiguity of the copying instructions). A neurologist who examined these drawings found them to be typical of those produced with unrestricted viewing by patients with apperceptive agnosia. Despite the amateur drawing skill of the students, you can verify that these are reasonably good copies for your peripheral vision by fixating on the central +. Courtesy of M. Martelli (Università di Roma "La Sapienza").

1133

vision of 20 deg or less[45]. Location of fixation affects perception of texture much less than it affects perception of objects (**Supplementary Fig. 2** online).

The Rey Complex Figure Test is widely used to assess the ability of neurological patients to copy a line drawing. Surprisingly, normal observers copying the figure with just their peripheral vision produce drawings that are similar to those produced with free viewing by patients with apperceptive agnosia, a type of object blindness (**Fig. 10**). These drawings suggest that the crowded peripheral vision of normal observers may be a good model for the central vision of these object-blind patients. Clinically, the excessive feature combination of crowding may account for apperceptive agnosia and strabismic amblyopia[31], whereas insufficient feature combination may correspond to simultanagnosia (see **Supplementary Discussion**).

Visual dichotomy

The uncrowded window and crowded surrounding field follow a long tradition of visual dichotomies: direct versus indirect, foveal versus peripheral, focal versus ambient, with versus without scrutiny, attentive versus pre-attentive, sustained versus transient, ventral versus dorsal, what versus where and perception versus action. This history of dichotomies distinguishes two kinds of vision. The first is typically central, acute, serial and 'conscious', and it recognizes and names objects. The second is typically peripheral, indistinct (blurry, vague, fuzzy, uncertain, confused and jumbled), parallel and 'unconscious', and it does not recognize or name objects, but helps to guide movement. Technically, these dichotomies are distinct, but in practice they have been used more or less interchangeably, following the fashions of vision science.

Crowding may be responsible for some of these dichotomies (a very close correspondence between pre-attentive and crowded vision can be seen in **Supplementary Fig. 3** online). Similarly, there is a strong association between crowded and unconscious vision. One sign of conscious awareness is reporting what we see, which is much harder when object recognition fails, leaving only unnamed texture. The failure of crowded viewing to produce object names may be why peripheral vision is so rarely described in science and literature. Acuity and other measures have been graphed as a function of eccentricity, but there are very few published descriptions of the everyday experience of crowded viewing (see **Supplementary Discussion**).

In everyday life, most of the things that we recognize are susceptible to crowding (by surrounding clutter) or self-crowding (among the parts). We see these things through a keyhole, the uncrowded window. Reading and searching speeds are proportional to the size of this window. We talk about and remember the things that we identify. The rest of our visual field is crowded, does not recognize or name things, and is hardly ever mentioned, but it lets us perceive space.

Attention

Attention is one of the most-studied topics in psychology (PsycInfo lists nearly 4,000 peer-reviewed articles on visual attention). If we take attention to be awareness of the target, it is clearly necessary for most object recognition tasks. Our purpose here is not to review attention as a general factor in object recognition, but rather to focus on a narrower question: the possible connection between attention and crowding (see **Supplementary Discussion**).

Selective attention is the filtering of a scene by the observer to emphasize a target. It is natural to interpret the critical spacing of crowding as the spatial resolution of selective attention[11]. Although there is evidence supporting this view (see **Supplementary Discussion** and **Supplementary Figs. 3** and **4** online), an alternative interpretation

sees crowding and selective attention as independent phenomena that affect object recognition separately. This alternative view is possible because selective attention enhances object recognition without affecting the critical spacing. These two interpretations differ in taking crowding to be either the resolution of attention or independent of attention, yet they agree in supposing that the critical spacing defines the area over which features are combined. Many investigators are trying to establish a link between crowding and attention.

Our ultimate goal is to achieve a computational model of the object recognition process. So far we have said only that features beyond the critical spacing for crowding are ignored. What happens inside of the critical spacing? How are features combined? Psychophysics, physiology and engineering all suggest that the first step is a reduction in the spatial precision of the internal representation of the stimulus through feature pooling (see **Supplementary Discussion**; a demonstration allows the reader to witness this imprecision, **Supplementary Fig. 5** online).

In this empirical review, the various studies of crowding all merge to tell a single story. Although the roles of learning, development, similarity and selective attention in crowding are still being worked out, there is a growing consensus that crowding is the combining of features over an inappropriately large area. Object recognition is usually limited by spacing and not by size. To be identified, simple objects must be separated by at least the observer's critical spacing, which corresponds to 6 mm at the primary visual cortex. Compound objects, such as words and faces, can crowd themselves. Their parts must be separated by at least the critical spacing. Thus, in our cluttered world, observers can identify objects only in an uncrowded window, determined by the object spacing. When the spacing is uniform, as in text, then the window will be central, where the critical spacing is smallest. These conclusions all spring from the consistent observations that, for each observer, the critical spacing of crowding depends solely on location and direction, which we call the Bouma law.

Note: Supplementary information is available on the Nature Neuroscience website.

1. Barlow, H.B. Summation and inhibition in the frog's retina. *J. Physiol. (Lond.)* **119**, 69–88 (1953).
2. Robson, J.G. & Graham, N. Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Res.* **21**, 409–418 (1981).
3. Treisman, A.M. & Gelade, G. A feature-integration theory of attention. *Cognit. Psychol.* **12**, 97–136 (1980).
4. Pelli, D.G., Burns, C.W., Farell, B. & Moore-Page, D.C. Feature detection and letter identification. *Vision Res.* **46**, 4646–4674 (2006).
5. Hubel, D.H. & Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (Lond.)* **160**, 106–154 (1962).
6. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
7. Ullman, S. *High-Level Vision: Object Recognition and Visual Cognition* (MIT Press, Cambridge, Massachusetts, 2000).
8. Parkes, L., Lund, J., Angelucci, A., Solomon, J.A. & Morgan, M. Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* **4**, 739–744 (2001).
9. Motter, B.C. Crowding and object integration within the receptive field of V4 neurons. *J. Vis.* **2:274**, 247a (2002).
10. Ledgeway, T., Hess, R.F. & Geisler, W.S. Grouping local orientation and direction signals to extract spatial contours: empirical tests of "association field" models of contour integration. *Vision Res.* **45**, 2511–2522 (2005).
11. Intriligator, J. & Cavanagh, P. The spatial resolution of visual attention. *Cognit. Psychol.* **43**, 171–216 (2001).
12. Biederman, I. Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* **94**, 115–147 (1987).
13. Tanaka, J.W. & Farah, M.J. Parts and wholes in face recognition. *Q. J. Exp. Psychol. A* **46**, 225–245 (1993).

14. Martelli, M., Majaj, N.J. & Pelli, D.G. Are faces processed like words? A diagnostic test for recognition by parts. *J. Vis.* **5**, 58–70 (2005).
15. Geisler, W. & Murray, R. Cognitive neuroscience: practice doesn't make perfect. *Nature* **423**, 696–697 (2003).
16. Levi, D.M. Crowding - an essential bottleneck for object recognition: a mini-review. *Vision Res.* **48**, 635–354 (2008).
17. Pelli, D.G., Palomares, M. & Majaj, N.J. Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J. Vis.* **4**, 1136–1169 (2004).
18. He, S., Cavanagh, P. & Intriligator, J. Attentional resolution and the locus of visual awareness. *Nature* **383**, 334–337 (1996).
19. Blake, R., Tadin, D., Sobel, K.V., Raissian, T.A. & Chong, S.C. Strength of early visual adaptation depends on visual awareness. *Proc. Natl. Acad. Sci. USA* **103**, 4783–4788 (2006).
20. Bouma, H. Interaction effects in parafoveal letter recognition. *Nature* **226**, 177–178 (1970).
21. Freeman, J. & Pelli, D.G. An escape from crowding. *J. Vis.* **7**, 1–14 (2007).
22. Reddy, L. & VanRullen, R. Spacing affects some, but not all, visual searches: implications for theories of attention and crowding. *J. Vis.* **7**, 1–17 (2007).
23. Solomon, J.A. & Morgan, M.J. Stochastic re-calibration: contextual effects on perceived tilt. *Proc Biol Sci* **273**, 2681–2686 (2006).
24. Bouma, H. Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Res.* **13**, 767–782 (1973).
25. Bex, P.J., Dakin, S.C. & Simmers, A.J. The shape and size of crowding for moving targets. *Vision Res.* **43**, 2895–2904 (2003).
26. van den Berg, R., Roerdink, J.B.T.M. & Cornelissen, F.W. On the generality of crowding: visual crowding in size, saturation and hue compared to orientation. *J. Vis.* **7**, 1–11 (2007).
27. Chung, S.T.L., Li, R.W. & Levi, D.M. Crowding between first- and second-order letter stimuli in normal foveal and peripheral vision. *J. Vis.* **7**, 1–13 (2007).
28. Pelli, D.G. *et al.* Crowding and eccentricity determine reading rate. *J. Vis.* **7**, 1–36 (2007).
29. Huey, E.B. *The Psychology and Pedagogy of Reading* (Macmillan, New York, 1908).
30. Legge, G.E., Pelli, D.G., Rubin, G.S. & Schleske, M.M. Psychophysics of reading. I. Normal vision. *Vision Res.* **25**, 239–252 (1985).
31. Levi, D.M., Song, S. & Pelli, D.G. Amblyopic reading is crowded. *J. Vis.* **7**, 1–17 (2007).
32. Vlaskamp, B.N., Over, E.A. & Hooge, I.T. Saccadic search performance: the effect of element spacing. *Exp. Brain Res.* **167**, 246–259 (2005).
33. Woodworth, R.S. *Experimental Psychology* (Holt, New York, 1938).
34. Bouma, H. Visual search and reading: eye movements and functional visual field: a tutorial review. in *Attention and Performance VII* (Requin, J.) (Erlbaum, Hillsdale, New Jersey, 1978).
35. Bosse, M.L., Tainturier, M.J. & Valdois, S. Developmental dyslexia: the visual attention span deficit hypothesis. *Cognition* **104**, 198–230 (2007).
36. McConkie, G.W. & Rayner, K. The span of the effective stimulus during a fixation in reading. *Percept. Psychophys.* **17**, 578–586 (1975).
37. Engbert, R., Nuthmann, A., Richter, E.M. & Kliegl, R. SWIFT: a dynamical model of saccade generation during reading. *Psychol. Rev.* **112**, 777–813 (2005).
38. O'Regan, J.K. Eye movements and reading. *Rev. Oculomot. Res.* **4**, 395–453 (1990).
39. Legge, G.E. *Psychophysics of Reading in Normal and Low Vision* (Erlbaum, Mahwah, New Jersey, 2007).
40. Motter, B.C. & Simoni, D.A. The roles of cortical image separation and size in active visual search performance. *J. Vis.* **7**, 1–15 (2007).
41. Stanovich, K.E., Siegel, L.S. & Gottardo, A. Converging evidence for phonological and surface subtypes of reading disability. *J. Educ. Psychol.* **89**, 114–127 (1997).
42. Bouma, H. & Legein, C.P. Foveal and parafoveal recognition of letters and words by dyslexics and by average readers. *Neuropsychologia* **15**, 69–80 (1977).
43. Prado, C., Dubois, M. & Valdois, S. The eye movements of dyslexic children during reading and visual search: impact of the visual attention span. *Vision Res.* **47**, 2521–2530 (2007).
44. Kwon, M., Legge, G.E. & Dubbels, B.R. Developmental changes in the visual span for reading. *Vision Res.* **47**, 2889–2900 (2007).
45. Lévy-Schoen, A. Exploration et connaissance de l'espace visuel sans vision périphérique. *Trav. Hum.* **39**, 63–72 (1976).
46. Chung, S.T.L., Levi, D.M. & Legge, G.E. Spatial-frequency and contrast properties of crowding. *Vision Res.* **41**, 1833–1850 (2001).
47. Toet, A. & Levi, D.M. The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Res.* **32**, 1349–1357 (1992).
48. Taylor, S.E. Eye movements in reading: facts and fallacies. *Am. Ed. Res. J.* **2**, 187–202 (1965).
49. Valdois, S. *et al.* Phonological and visual processing deficits can dissociate in developmental dyslexia: evidence from two case studies. *Read. Writing* **16**, 541–572 (2003).
50. Martelli, M., Di Filippo, G., Spinelli, D. & Zoccolotti, P. Crowding, reading and developmental dyslexia. *Perception* **35**, supp 174 (2006).

Erratum: The uncrowded window of object recognition

Denis G Pelli & Katharine A Tillman

In the version of this article initially published, the legend to Figure 5 contained several errors. The second sentence should read "Fixating on the red minus, you will be unable to identify the middle object in the first eight rows unless you isolate it by hiding the flanking objects with your fingers (or two pencils)." The fourth sentence should read "Grating patches, similar to those in the top row, are often taken to be one-feature objects." These errors have been corrected in the HTML and PDF versions of the article.

nature
neuroscience

The uncrowded window of object recognition

Denis G Pelli & Katharine A Tillman
Psychology & Neural Science, NYU
http://psych.nyu.edu/pelli/

This is the online supplement to our Perspective review.

SUPPLEMENTARY DISCUSSION
The Bouma law at the cortex
The known eccentricity-dependence of the *cortical magnification factor* (mm on the cortex per deg of visual angle) produces a logarithmic map of the visual field on the primary visual cortex (V1). The logarithmic transformation of the proportional critical spacing at the visual field results in a fixed critical spacing at the cortex (6 mm at V1), independent of eccentricity.

Let us work this out. Bouma showed that critical spacing $\Delta\varphi$ is $b\varphi$, where b is Bouma's proportionality constant between critical spacing and eccentricity φ (ref. [1]). In V1 and many other areas in the visual cortex, eccentricity φ in the visual field is an exponential function of position d (in mm) on the cortex, $\varphi=\exp[\beta(d+\alpha)]$, where α and β are empirical constants, unique to each cortical area[2]. So position d is $(\log\varphi)/\beta - \alpha$, a logarithmic map. If the target is at eccentricity φ, then a flanker one critical spacing farther from fixation will be at eccentricity $\varphi+b\varphi$. b and β are fixed constants, so the cortical separation $\Delta d = d_{\varphi+b\varphi} - d_{\varphi} = \log(1+b)/\beta$ is a fixed number of mm, independent of target location, in every cortical area that is logarithmically mapped[3, 4]. We take $b = 0.4$, as in **Fig. 5**, so in V1, where $\beta = 0.0577$/mm, Δd is 6 mm.

Spatial extent of crowding
The invariance of critical spacing demonstrated in **Fig. 5** is found when the target and flankers have similar features (e.g., black letters flanking a black letter target). These typical cases produce maximum crowding. Flankers that have features unlike the target (e.g., white letters flanking a black letter target, on a gray background) produce much less crowding or none at all[5-7]. This weaker effect is usually reported as a reduction in critical spacing, but perhaps the spatial extent of crowding is unchanged and the effect is only reduced in amplitude (strength). The reported reduction of critical spacing may be an artifact of defining critical spacing by a performance criterion, as discussed below. Compared to the effect of target-like flankers, dissimilar flankers may simply have a weaker effect over the same spatial extent. Incidentally, the field still lacks an objective definition of similarity to predict crowding. We were surprised to learn that first- and second-order letters crowd each other, with the usual critical spacing, despite having very different features[8]. It might seem

that similarity, for the purpose of crowding, would be just the opposite of salience, but in fact salience has little or no effect on crowding[9].

Crowding has usually been characterized by just one number, "critical spacing", i.e., spacing threshold, the spacing required to achieve a criterion level of performance. That single number seems to be enough to characterize crowding when the flanker is similar to the target, but may not adequately describe the weaker crowding produced by dissimilar flankers. Disentangling the amplitude and extent of crowding demands a two-number description. The complete 'psychometric function', plotting proportion correct as a function of spacing, tells us little more than the critical spacing. Proportion correct has a small dynamic range bounded by the floor at chance, when spacing is below critical, and by the ceiling at 100%, when spacing is above critical. To get the whole story, we must replace proportion correct by a better dependent measure: threshold. To measure threshold, one varies a physical parameter of the stimulus to achieve a particular level of performance[10]. Thus, threshold is measured on a physical scale with a wide dynamic range. For example, several studies have measured orientation discrimination thresholds as a function of spacing. These plots show that the weaker crowding produced by less-similar flankers has much less amplitude (maximum threshold elevation) but practically the same spatial extent[6, 7, 11].

Changing the orientation of the flankers from parallel to orthogonal (to the target) halves the amplitude without obvious reduction of extent[6]. Arranging the flankers in a ring to form a closed contour reduces the threshold elevation by a factor of 6 without reducing its spatial extent (defined as spacing for half maximum log threshold elevation)[12].

Crowding diminishes somewhat with practice, but the improvement is specific to the trained strings[13] and does not transfer from 3-letter strings to reading[14]. The benefit of practice has been reported as a reduction in critical spacing, but a two-parameter analysis might reveal, as above, that only amplitude (not extent) is affected.

At present, the simplest account is that the spatial extent of crowding for a particular location and direction is independent of the particular target and flanker. However, that conclusion is tentative because most published studies have not disentangled the amplitude and extent of crowding. Thus, this review focuses on "critical spacing", but the reader should bear in mind that the special cases we just discussed demand a two-parameter (amplitude and extent) characterization of crowding.

Uncrowded neighborhood and search

An object must be in the observer's uncrowded window to be recognized. Inverting the idea, **Supplementary Fig. 1** shows the object's *uncrowded neighborhood* — that is, the area around the object within which you must fixate to see it uncrowded. The uncrowded neighborhood is much like Engel's "conspicuity area", though he did not mention crowding[15]. Only those objects whose uncrowded neighborhoods include the observer's point of fixation are recognizable. If the observer fixates randomly, then the probability that the fixation will land in a particular object's uncrowded neighborhood (and thus that the object can be recognized) is equal to the fraction of the image area occupied by the uncrowded neighborhood. In the popular children's book *Where's Waldo?*[16], your chance of finding Waldo in your first glimpse is proportional to the area of his uncrowded neighborhood[15, 17].

Descriptions of crowded viewing

In 1936, the Gestalt psychologist Wolfgang Metzger described crowded viewing: "Farther out [in the periphery], the structure becomes ever weaker and cruder … The unifying effect of proximity becomes overwhelming. … [D]ifferences … cause an imbalance and restlessness in each intrafigural organization that is difficult to describe and can best be compared with what, in clearly seen objects, is called … texture … You see in that region ... no clearly segregated, countable, or, above all, individually identifiable component parts"[18]. In 1976, Jerry Lettvin added, "Things are less distinct as they lie farther from my gaze. It is not as if things there go out of focus … it's as if somehow they lost the quality of 'form'"[19].

Peripheral vision and texture

What do we see when vision is crowded? We see stuff (unnamed texture) and perceive space (the shape of the scene we are in). Location affects perception of texture much less than perception of objects (**Supplementary Fig. 2**).

Clinical conditions

Crowding is the fruitless combining of features over too large an area. Crowding with abnormally large critical spacing may account for several clinical conditions, including apperceptive agnosia and strabismic amblyopia[20, 21].

In principle it would be similarly fruitless to combine over too small an area, getting only a fraction of the object. This matches some clinical descriptions of simultanagnosia: "It often appeared as if he were looking through a peephole which was too narrow to include the entire stimulus"[22].

Supp. Fig. 1. The uncrowded neighborhoods (white polygons) of two objects: water bottle and magazine. You must fixate within its neighborhood to recognize the object. Fixating outside the uncrowded neighborhoods (e.g., on the little girl's face), you cannot recognize either of these objects. The neighborhood size (white polygon) depends on the local density of the clutter around the object and the similarity of the clutter to the object. The polygon is the measured threshold eccentricity for recognition in eight directions. These thresholds are subjective; the observer knew all along what the object was. Train station, Moscow, Russia, 2006. Photo by K.A. Tillman.

Supp. Fig. 2. A forest. This is mostly texture, with very few recognizable objects. Unlike perception of objects, the perception of texture is little affected by the location of fixation[23]. We suggest that one might define "texture" as what one can see without object recognition. Copyright © Ray K. Metzker, Courtesy Laurence Miller Gallery, New York.

Attention

Is attention involved in crowding and object recognition? It depends on what you mean by "attention". In William James's broadest view, when looking at the world, attention means awareness. Most object recognition tasks ask the observer to report a target (or its absence). This report (over many trials) typically communicates information that could only have come from the target, which strongly suggests that the observer was aware of it. (The exceptions are the much-discussed special cases of stimuli that may affect reports without entering consciousness, e.g., blindsight and subliminal priming.) If awareness of the target is attention, then attention is a near-essential part of most object recognition tasks, but this general fact is not related to crowding in particular.

More specifically, there has been great interest in selective attention in visual search and texture segregation. *Selective attention* refers to the observer's ability to filter the visual scene, emphasizing some areas (or things) and ignoring others. Helmholtz said, "A human being cannot attend to more than one object at a time, … [but,] in spite of the vagueness of the broad field of view, the eye is capable of taking in at a rapid glance the main features of the whole surroundings, and of noting immediately the sudden appearances of new objects … to divert our attention to any new or extraordinary phenomenon that may arise out toward the periphery of the field"[24]. A century later, Julesz and Bergen[25] express the same idea: "Thus preattentive vision serves as an 'early warning system' by pointing out those loci of texton [i.e. feature] differences that should be attended to"[25, p. 1619]. Treisman's Feature Integration Theory (FIT) goes on to suppose that observers have a *focused attention* process (i.e., selective attention) that can be directed to any area or object in the visual field, and that only this process can correctly integrate (i.e., combine) features for object recognition[26]. Most of the two thousand or so papers on visual search interpret their results as characterizing the limits of selective attention[27]. In this framework, it is natural to interpret the critical spacing for crowding as being the spatial resolution of attention[28]. That interpretation is parsimonious in attributing the crowding phenomenon to FIT's mechanism of focused attention.

Wolfe and Bennett[29] describe shapeless "preattentive object files" of "unbound" [not combined] features. This is similar to the "bag of features" idea in machine learning[30], and seems to be a good description of crowding. Indeed, the "attentive" vs. "preattentive" dichotomy seems to correspond to uncrowded vs. crowded vision. In particular, Treisman and Gelade[26] show that finding an R among P's and Q's is serial, "requiring focused attention". Similarly, Julesz & Bergen[25, p. 1621] say that "element-by-element scrutiny, called 'focal attention', is required to find the T's embedded in the L's." And Wolfe and Bennett[29] report serial search for a target plus sign, +, consisting of a green vertical and a red horizontal bar among distractor pluses that each consist of a red vertical and a green horizontal bar. **Supplementary Fig. 3** shows minimal versions of these stimuli. In each of the top three rows,

Supp. Fig. 3. Is crowded vision "preattentive"? These are minimal versions of stimuli that are known to require serial search[29, 31, 32]. In the top three rows, fixating the central minus, you will easily identify the isolated target at the left, but you will fail to identify the same now-crowded (middle) target on the right. The bottom row shows self-crowding: Fixating on the minus you will find that the left and right targets are indistinguishable. In every case, you can readily identify the target if you fixate the nearby small grey plus. Thus, reducing eccentricity changes your vision from "preattentive" to "attentive" even though you are concentrating on the target throughout.

fixating on the minus, you can easily identify the isolated target to the left, but you cannot identify the same target on the right, where it is flanked by distractors. This is crowding. While fixating the minus, no amount of willpower will rescue the crowded target. However, moving fixation closer, to the small grey plus, does restore your ability to identify the target. Thus, for this fixed spacing, reducing eccentricity changes vision from "preattentive" to "attentive", even though the observer can concentrate on the target throughout. The bottom row shows self-crowding. While fixating the minus, you will find that the isolated target '10' on the right is indistinguishable from the isolated distractor 'S' on the left. Julesz says, "The 'S' and '10' shaped elements … in isolation appear quite different, [but are] preattentively indistinguishable"[25, p. 1626]. These demonstrations are consistent with the finding that target eccentricity strongly affects performance of searches for multi-feature targets[3, 17, 33].

One could say that these crowding effects are independent of attention. Or, as above, one could take the critical spacing of crowding to be the resolution (minimum area) of attention, treating the two as one.

While it is clear that directing the observer's attention to the object can enhance recognition, it is not clear that this focused attention is essential (despite FIT's claim that it is). Motter and Simoni recorded eye position and analyzed the trial-by-trial results as the probability of target recognition as a function of spacing (to nearest distractor) as a fraction of target eccentricity[3, 17]. They concluded that each target recognition is limited by the critical spacing of crowding, without invoking selective attention. Similarly, using a target embedded at a known or unknown location in a ring of distractors, all at the same eccentricity, several investigators have looked for and failed to find an effect of selective attention on the critical spacing for crowding[7, 34, 35]. This suggests another interpretation, equally parsimonious: Perhaps crowding may be understood on its own, independent of selective attention. Selective attention does enhance object

recognition, but since it does not affect critical spacing, perhaps it is a separate factor in object recognition, independent of crowding. These interpretations differ in taking crowding to be either the resolution of attention or independent of attention, yet they agree in describing crowding as the combining of features over an inappropriately large area.

Non-retinotopic crowding. A popular crowding demo, http://visionlab.harvard.edu/Members/Patrick/CrowdingMovies/, by Cavanagh and Holcombe, is taken by many to be evidence for the attentional-resolution account of crowding. Our view is that the demo shows just what one would expect given the Bex et al.[36] finding that the critical spacing of crowding is unchanged when measured in a moving reference frame. The Cavanagh and Holcombe demo alternately displays a target and flanking distractors[37]. Even though the target and distractors are never displayed simultaneously, the target is crowded unless the alternation rate is very slow (few Hz). They called that flickering linear array of letters an "arm", and arranged 16 such arms radially, like the spokes in a wheel, around the central fixation point. The arms are numbered 1 to 16. All the odd-numbered arms (i.e., every other arm) are in sync, all showing the target at the same time. The even-numbered arms are in sync with each other but out-of-sync with the odd-numbered arms, displaying the distractors when the odd arms display the target. Again, unless the alternation rate is very low, the target is crowded. However, the authors then add an attentional guide, brightening one sector (containing arm *t*) at time *t*, and ask the observer to maintain central fixation while mentally tracking the bright sector as it goes around. The brightened arm is just a target letter, in every frame, and observers are much better able to identify it with the guide than without it. Cavanagh and Holcombe call this "non-retinotopic crowding". This phenomenon needs more than a strictly spatial account, since the result is spatiotemporal, "suggesting [that] the crowding is specific to the flankers, if any, that move with the target and not to the letters that surround each target locally in retinotopic coordinates."

It seems that, without a guide, the observer is using a stationary integrator (continuous oval in **Supplementary Fig. 4**). With a guide, the observer seems to be using a moving integrator (dashed oval). It matters little how many frames the integrator integrates provided it's more than one. It is well known that "translation *per se* across the retina has little effect on temporal summation" of an object[38]. Apparently the moving attentional cue helps the observer select a moving, instead of a stationary, integrator. Since motion *per se* does not seem to affect crowding[36], the cueing here seems to allow the observer to select an uncrowded (moving) representation of the stimulus. A rotating bright bar is a powerful motion stimulus. We would expect it to push the observer towards seeing motion, even if attention were not involved.
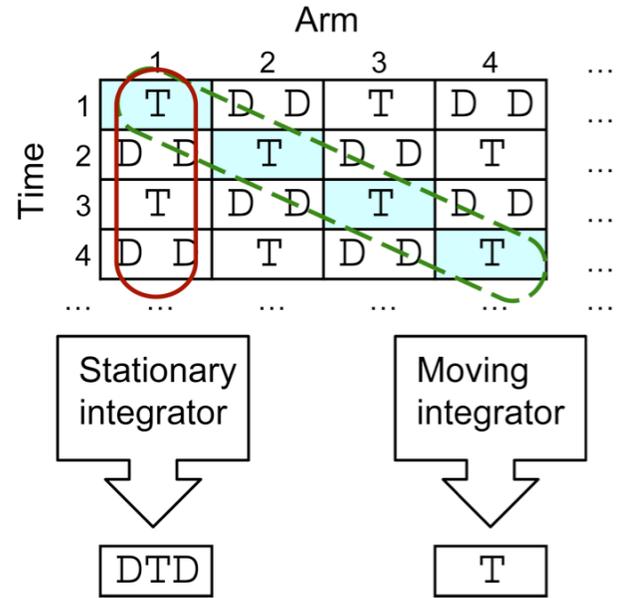
Supp. Fig. 4. Non-retinotopic crowding. Analysis of the Cavanagh and Holcombe demo.[37] http://visionlab.harvard.edu/Members/Patrick/CrowdingMovies/ ***Their demo (not shown):*** Sixteen radial arms are presented, each with a target letter in the middle, flanked by distractors on each side. The display counterphases targets and distractors. In one frame (presented at odd times 1, 3, …), every odd arm (1, 3, …) shows only the target without distractors, while the even arms (2, 4, …) show only the distractors without the targets. In the other frame (presented at even times), even arms show targets while odd arms show the distractors. Thus, when a guide (a brightened radial sector) moves from arm to arm in phase with the alternation (arm 1 at time 1, arm 2 at time 2, …), it contains only the target letter and no distractors. Subjects fixate the center of the circular array and report the orientation of the target letter. When attention is directed to one fixed location, there is substantial interference from the distractors. However, when following the guide, crowding is much reduced, suggesting that distractors only crowd the target if they remain with it over time. ***Our space-time diagram:*** The table shows the content of each arm at each time, which is either the target letter (T) or the distractors (D D). The shading represents the attentional cue: the brightening of arm *t* at time *t*.

Feature pooling

It seems that combining features to recognize objects carries the risk of crowding. Above, we reviewed evidence that the critical spacing is the spatial extent over which features are combined, but we have said little about what goes on within this combining area. Some of the complaints about crowding — especially the impaired judgment of position and shape — seem to stem from uncertainty (confusion) about feature position. One may complain about this uncertainty, but we all benefit from the positional invariance of recognition. Although the relative positions of features vary among fonts and handwritings, for a letter to be read, it must be assigned to the same category, e.g., "a", regardless of its rendering.

How are features combined? Three lines of investigation (psychophysics, physiology, and engineering) converge on *maximum pooling* as a key step. In maximum pooling, many feature detectors with similar receptive fields, differing only in position, all respond to the stimulus independently, but only

the maximum detector response, regardless of detector position, is passed on. This immediate loss of precision of feature position is an important aspect of psychophysical and physiological models and engineering solutions for object recognition[39-41].

Psychophysically, when attempting to identify something, human observers act as though they are always considering many possible positions, like the ideal observer for an uncertain signal, which does maximum pooling[42-45]. **Supplementary Fig. 5** allows you to witness this vagueness of feature position. Physiologically, in the primary visual cortex, complex cell responses are position invariant and do not summate, consistent with maximum pooling[46, 47]. In practical engineering, some of the most successful machine classifiers of handwritten digits (and other objects) use maximum pooling to tolerate "deformations and shifts in position"[48-50]. In all these cases, invariance of object recognition is achieved by maximum pooling, which results in uncertainty of feature location, which, in turn, limits the precision of judgments of object position and shape.

SUPPLEMENTARY SOURCES
Figure 4 images. The Elvis image is used by permission, Elvis Presley Enterprises, Inc. The Arnold image is publicly available from the State of California website. **Figure 5 images.** The gratings were created in MATLAB. The letters are in the Courier font. The animal silhouettes are in our Animals font, which is available for research purposes. The men, women, and telephone signs are from aiga.com[51]. The ladder is licensed from and copyright Stockbyte. The rocking chair is copyright 2008 Jupiter Images Corporation. In the following credits, we use the convention (Photographer/Name of collection/Source). The stool (C Squared Studios /Photodisc/GettyImages), pretzel (Steve Wisbauer/Photodisc/Getty Images), hamburger (Ryan McVay/Photodisc/Getty Images), and pizza (Raimund Koch/Riser/Getty Images) are from Getty Images. The house is courtesy of Snodgrass and Vanderwart[52]. The image of Gandhi is copyright Vithalbhai Jhaveri/GandhiServe.

SUPPLEMENTARY METHODS
Figure 9 methods. Taylor[53] tested thousands of normal students in 1st grade through college. We plot one point per grade, 1-12, plus college. Reading speed (vertical scale):

Subjects read age-appropriate paragraphs. Carver corrected the speeds for text difficulty [54, Table 2.1]. Span (horizontal scale): Taylor measured eye movements. We plot Taylor's "span of recognition", the average length of forward saccades (the product of words per saccade, from Taylor, and characters per word, for text at each difficulty level, from Carver's Eq. 2.2). **Kwon et al.**[55] tested normal 3rd, 5th, and 7th graders, and adults (4 points). Reading speed: Subjects read sentences displayed one at a time. Span: Kwon et al. measured 'visual span profiles', which trace out the subject's accuracy for identifying a triplet of random letters as a function of position in the visual field. We plot the number of letter positions in the visual span profile (for 1 deg letters) for which the triplet accuracy is at least 80%. **Valdois et al.**[56] tested two dyslexic subjects (in the 6th and 7th grades), classified as a surface dyslexic (△) and a phonological dyslexic (◿), and age-matched controls (▲). Reading speed: Valdois provided reading speeds for ordinary text. Span: They measured accuracy versus letter position for reporting a string of 5 briefly-presented random letters (their Fig. 1). We plot the number of positions (out of 5) at which the subject got at least 80% correct. **Martelli et al.**[57] tested normal and dyslexic 6th graders. Normal adult data were provided separately. Reading speed: The normal and dyslexic children read ordinary text printed on paper. The adults read 8-letter nouns in rapid serial visual presentation. Span: They measured the critical spacing required to identify the central letter in a triplet of three random letters with 90% accuracy as a function of eccentricity. They then calculated Bouma's factor b (proportionality constant between critical spacing and eccentricity). We plot the uncrowded span $u = 1+2/b$ [1]. **Prado et al.**[58] tested dyslexics and age-matched controls. (They did not report the students' grade level, but average age was 11 years, which is typical for the 6th grade.) Reading speed: They measured eye movements as subjects read short passages. We plot rate as the number of words in the passage divided by the product of the total number of fixations and the mean fixation duration (their Table 2). Span: We plot the average number of letters reported correctly from a string of 5 briefly presented letters (their Table 1).

For reviews of reading, see refs [59-63]. *For further reading, see refs* [64-67].

+ K

Supp. Fig. 5. Experience the vagueness of feature position predicted by maximum pooling. Viewing the page from 17 cm away (though distance hardly matters), fixate the plus. The letter (3.3° at an eccentricity of 46°) is too small to recognize, and looks like "a jumble of lines or an unorganized heap of marks"[68]. Optical blur is noticeable, but does not prevent you from seeing the lines. The feature position errors are so large that you see only a jumble of floating features. One observer said, "I see something that appears to be composed of straight lines about half an inch high. Could be a drawing. Could be a letter or letters. I cannot see clearly what it is. At the moment it looks like a capital Y, but it's indefinable. The lines are not precise. They appear to be shimmering, fading in and out. Very unstable figure". Such confusion of feature position is predicted by maximum pooling. We think that maximum pooling not only contributes to crowding, but also limits acuity, as shown here. There is no crowding here because there is only one simple letter, which consists of only one part, though it has many features. The uncertainty of feature position seems to be a fixed fraction of the combining area. For example, the just noticeable difference in position of a grating patch (not shown) is independent of the spatial frequency of the grating and is about 4% of whichever is larger: the extent of the grating or 50% of the eccentricity[69]. We think that this uncertainty contributes to crowding, but its spatial extent is much too small to be the main cause of crowding.

1. Pelli, D.G., Tillman, K.A., Freeman, J., Su, M., Berger, T.D. & Majaj, N.J. Crowding and eccentricity determine reading rate. *Journal of Vision* **7(2)20**, 1-36 (2007). [Pubmed]
2. Larsson, J. & Heeger, D.J. Two retinotopic visual areas in human lateral occipital cortex. *J Neurosci* **26**, 13128-13142 (2006).
3. Motter, B.C. & Simoni, D.A. The roles of cortical image separation and size in active visual search performance. *Journal of Vision* **7(2)6**, 1-15 (2007).
4. Levi, D.M., Klein, S.A. & Aitsebaomo, A.P. Vernier acuity, crowding and cortical magnification. *Vision Res* **25**, 963-977 (1985).
5. Kooi, F.L., Toet, A., Tripathy, S.P. & Levi, D.M. The effect of similarity and duration on spatial interaction in peripheral vision. *Spat Vis* **8**, 255-279 (1994).
6. Andriessen, J.J. & Bouma, H. Eccentric vision: adverse interactions between line segments. *Vision Res* **16**, 71-78 (1976).
7. Wilkinson, F., Wilson, H.R. & Ellemberg, D. Lateral interactions in peripherally viewed texture arrays. *J Opt Soc Am A* **14**, 2057-2068 (1997).
8. Chung, S.T.L., Li, R.W. & Levi, D.M. Crowding between first- and second-order letter stimuli in normal foveal and peripheral vision. *Journal of Vision* **7(2)10**, 1-13 (2007).
9. Felisbert, F.M., Solomon, J.A. & Morgan, M.J. The role of target salience in crowding. *Perception* **34**, 823-833 (2005).
10. Farell, B. & Pelli, D.G. Psychophysical methods, or how to measure threshold, and why. In *A Practical Guide to Vision Research* (ed. J.G. Robson & R.H.S. Carpenter) 129-136 (Oxford University Press, New York, 1999).
11. Levi, D.M., Hariharan, S. & Klein, S.A. Suppressive and facilitatory spatial interactions in peripheral vision: peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision* **2(2)3**, 167-177 (2002).
12. Livne, T. & Sagi, D. Configuration influence on crowding. *Journal of Vision* **7(2)4**, 1-12 (2007).
13. Huckauf, A. & Nazir, T.A. How odgcrnwi becomes crowding: Stimulus-specific learning reduces crowding. *Journal of Vision* **7(2)18**, 1-12 (2007).
14. Chung, S.T.L. Learning to identify crowded letters: Does it improve reading speed? *Vision Res* **47**, 3150-3159 (2007).
15. Engel, F.L. Visual conspicuity, visual search and fixation tendencies of the eye. *Vision Res* **17**, 95-108 (1977).
16. Handford, M. *Where's Waldo?* (Little Brown, Boston, 1987).

17. Vlaskamp, B.N., Over, E.A. & Hooge, I.T. Saccadic search performance: the effect of element spacing. *Exp Brain Res* **167**, 246-259 (2005).
18. Metzger, W. *Laws of seeing* (MIT Press, Cambridge, MA, 2006).
19. Lettvin, J.Y. On seeing sidelong. *N. Y. Acad. Sci.* **16**, 10-20 (1976).
20. Levi, D.M., Song, S. & Pelli, D.G. Amblyopic reading is crowded. *Journal of Vision* **7(2)21**, 1-17 (2007). [Pubmed]
21. Song, S., Levi, D.M. & Pelli, D.G. Can "equivalent eccentricity" account for two measures of performance in the amblyopic visual system? *ARVO 2008 Annual Meeting* April 27-May 1 in Fort Lauderdale, Florida (2008).
22. Trobe, J.R. & Bauer, R.M. Seeing but not recognizing. *Surv Ophthalmol* **30**, 328-336 (1986).
23. Braun, J. & Sagi, D. Texture-based tasks are little affected by second tasks requiring peripheral or central attentive fixation. *Perception* **20**, 483-500 (1991).
24. Helmholtz, H.v. *Treatise on Physiological Optics (Dover ed.), vol. 1, sec. 10, pg. 93 and vol. 3, sec. 28, pg. 205* (Dover Publications, Mineola, NY, 1867/2005).
25. Julesz, B. & Bergen, J.R. Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal* **62**, 1619-1645 (1983).
26. Treisman, A.M. & Gelade, G. A feature-integration theory of attention. *Cognit Psychol* **12**, 97-136 (1980).
27. Wolfe, J.M. Visual search. In *Attention* (ed. H. Pashler) 13-73 (Psychology Press, New York, 1998).
28. Intriligator, J. & Cavanagh, P. The spatial resolution of visual attention. *Cognit Psychol* **43**, 171-216 (2001).
29. Wolfe, J.M. & Bennett, S.C. Preattentive object files: shapeless bundles of basic features. *Vision Res* **37**, 25-43 (1997).
30. Breiman, L. Bagging predictors. *Machine Learning* **24**, 123-140 (1996).
31. Treisman, A. Features and objects: the fourteenth Bartlett memorial lecture. *Q J Exp Psychol A* **40**, 201-237 (1988).
32. Julesz, B. A brief outline of the texton theory of human vision. *Trends in Neurosciences* **7**, 41-45 (1984).
33. Carrasco, M., Evert, D.L., Chang, I. & Katz, S.M. The eccentricity effect: target eccentricity affects performance on conjunction searches. *Percept Psychophys* **57**, 1241-1261 (1995).
34. Scolari, M., Kohnen, A., Barton, B. & Awh, E. Spatial attention, preview, and popout: which factors influence critical spacing in crowded displays? *Journal of Vision* **7(2)7**, 7 1-23 (2007).
35. Nazir, T.A. Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Res* **32**, 771-777 (1992).
36. Bex, P.J., Dakin, S.C. & Simmers, A.J. The shape and size of crowding for moving targets. *Vision Res* **43**, 2895-2904 (2003).
37. Cavanagh, P. & Holcombe, A.O. Non-retinotopic crowding. *Journal of Vision* **7(9)**, 338a (2007).
38. Burr, D.C. Temporal summation of moving images by the human visual system. *Proc R Soc Lond [Biol]* **211**, 321-339 (1981).
39. Baldassi, S., Megna, N. & Burr, D.C. Visual clutter causes high-magnitude errors. *PLoS Biol* **4**, e56 (2006).

40. Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. Receptive field organization of complex cells in the cat's striate cortex. *J Physiol (Lond)* **283**, 79-99 (1978).

41. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**, 2278-2324 (1998).

42. Pelli, D.G. Uncertainty explains many aspects of visual contrast detection and discrimination. *J Opt Soc Am A* **2**, 1508-1532 (1985). [Pubmed]

43. Tjan, B.S. & Nandy, A.S. Classification images with uncertainty. *Journal of Vision* **6(4)8**, 387-413 (2006).

44. Shaw, M.L. Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology* **14**, 353-409 (1982).

45. Hess, R.F., McIlhagga, W. & Field, D.J. Contour integration in strabismic amblyopia: the sufficiency of an explanation based on positional uncertainty. *Vision Res* **37**, 3145-3161 (1997).

46. Hubel, D.H. & Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* **160**, 106-154 (1962).

47. Lampl, I., Ferster, D., Poggio, T. & Riesenhuber, M. Intracellular measurements of spatial integration and the MAX operation in complex cells of the cat primary visual cortex. *J Neurophysiol* **92**, 2704-2713 (2004).

48. Fukushima, K. & Miyake, S. Neocognitron - A new algorithm for pattern-recognition tolerant of deformations and shifts in position. *Pattern Recognition* **15**, 455-469 (1982).

49. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M. & Poggio, T. Robust object recognition with cortex-like mechanisms. *IEEE Trans Pattern Anal Mach Intell* **29**, 411-426 (2007).

50. Ranzato, M.A., Huang, F.-J., Boureau, Y.-L. & LeCun, Y. Unsupervised learning of invariant feature hierarchies with applications to object recognition. *Proc. Computer Vision and Pattern Recognition Conference (CVPR'07)*, 1-8 (2007).

51. AIGA. *Symbol signs* (Hastings House, New York, 1981).

52. Snodgrass, J.G. & Vanderwart, M. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *J Exp Psychol [Hum Learn]* **6**, 174-215 (1980).

53. Taylor, S.E. Eye movements in reading: Facts and fallacies. *American Educational Research Journal* **2(4)**, 187-202 (1965).

54. Carver, R.P. *Reading rate: a review of research and theory* (Academic Press, San Diego, 1990).

55. Kwon, M., Legge, G.E. & Dubbels, B.R. Developmental changes in the visual span for reading. *Vision Res* **47**, 2889-2900 (2007).

56. Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D. & Pellat, J. Phonological and visual processing deficits can dissociate in developmental dyslexia: Evidence from two case studies. *Reading and Writing* **16**, 541-572 (2003).

57. Martelli, M., Di Filippo, G., Spinelli, D. & Zoccolotti, P. Crowding, reading, and developmental dyslexia. *Perception* **35, supp**, 174 (2006).

58. Prado, C., Dubois, M. & Valdois, S. The eye movements of dyslexic children during reading and visual search: Impact of the visual attention span. *Vision Res* **47**, 2521-2530 (2007).

59. Huey, E.B. *The psychology and pedagogy of reading* (Macmillan, New York, 1908).

60. Snowling, M.J. From language to reading and dyslexia. *Dyslexia* **7**, 37-46 (2001).

61. Legge, G.E. *Psychophysics of reading in normal and low vision* (Erlbaum, Mahwah, NJ, 2007).

62. Pelli, D.G. & Tillman, K.A. Parts, wholes, and context in reading: a triple dissociation. *PLoS ONE* **2**, e680 (2007). [Pubmed]

63. Hutzler, F. & Wimmer, H. Eye movements of dyslexic children when reading in a regular orthography. *Brain Lang* **89**, 235-242 (2004).

64. Plato. *Republic, 2, line 368^d* (Hackett Publishing Co., Indianapolis, 1992).

65. Quine, W.V.O. Two dogmas of empiricism. *Philosophical Review* **60**, 20-43 (1951). [Reprinted in Quine, W. V. O. *From a logical point of view*. (Harvard University Press, Cambridge, MA, 1980)]

66. Rey, A. L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives de Psychologie* **28**, 286-340 (1941).

67. Faye, E. *Clinical low vision* (Little Brown, Boston, 1984).

68. Zigler, M.J., Cook, B., Miller, D. & Wemple, L. The perception of form in peripheral vision. *American Journal of Psychology* **42**, 246-259 (1930).

69. Levi, D.M. & Tripathy, S.P. Localization of a peripheral patch: the role of blur and spatial frequency. *Vision Res* **36**, 3785-3803 (1996).

70. Wade, N.J. Image, eye, and retina. *J Opt Soc Am A Opt Image Sci Vis* **24**, 1229-1249 (2007).