

Visual crowding: a fundamental limit on conscious perception and object recognition

David Whitney^{1,3} and Dennis M. Levi^{2,3}

¹Department of Psychology, University of California, 3210 Tolman Hall, Berkeley, CA 94720-1650, USA

²School of Optometry, University of California, 351 Minor Addition, Berkeley, CA 94720-2020, USA

³Helen Wills Neuroscience Institute, University of California, 3210F Tolman Hall, Berkeley, CA 94720-3192, USA

Crowding, the inability to recognize objects in clutter, sets a fundamental limit on conscious visual perception and object recognition throughout most of the visual field. Despite how widespread and essential it is to object recognition, reading and visually guided action, a solid operational definition of what crowding is has only recently become clear. The goal of this review is to provide a broad-based synthesis of the most recent findings in this area, to define what crowding is and is not, and to set the stage for future work that will extend our understanding of crowding well beyond low-level vision. Here we define six diagnostic criteria for what counts as crowding, and further describe factors that both escape and break crowding. All of these lead to the conclusion that crowding occurs at multiple stages in the visual hierarchy.

Dispelling the illusion

With regular flicks of the eye, we establish and maintain the illusion of a continuous high-resolution representation of our visual environment. This compelling illusion is easy to dispel by trying to describe the details of objects in your peripheral visual field; scrutinizing or trying to count objects in the visual periphery is impossible. This partly reflects the well-known decline in visual acuity in peripheral vision. However, the most widespread impediment to reading and object recognition in the periphery is the mysterious process known as crowding, which is the deleterious effect of clutter on peripheral object recognition. Objects that can be easily identified in isolation seem indistinct and jumbled in clutter (Figure 1).

Crowding is an essential bottleneck, setting limits on object perception, eye and hand movements, visual search, reading and perhaps other functions in peripheral, amblyopic and developing vision. Crowding impairs not only discrimination of object features and contours, but also the ability to recognize and respond appropriately to objects in clutter. Thus, studying crowding might lead to a better understanding of the processes involved in object recognition. Crowding also has important clinical implications for patients with macular degeneration, amblyopia and dyslexia.

Interest in crowding has significantly increased in the past few years and has yielded a more sophisticated un-

derstanding of the phenomenon itself, as well as of the processes involved in object recognition and reading. Two reviews provide overviews of much of the relevant literature published at the time [1,2]. Levi concluded that 'Crowding is an enigma wrapped in a paradox and shrouded in a conundrum. Despite a great deal of new (and old) work, we do not yet have a full understanding of crowding' [1]. Since then, novel approaches, models and findings have provided new insights into the mysteries of crowding and suggest that crowding occurs at multiple stages in the visual hierarchy.

Operationally defining crowding

The significance of crowding is clear from phenomenological demonstrations of its power and ubiquity in natural scenes (Figure 1). Ultimately, however, characterizing and understanding the mechanism(s) of crowding require more than a phenomenological description. Recent work has established that there are several diagnostic criteria for crowding, and use of these as converging evidence can help

Glossary

Aftereffect: delayed or prolonged physiological or psychological response following exposure to a stimulus.

Amblyopia: developmental abnormality that results from physiological alterations in the visual cortex and impairs form vision (from the Greek, *amblyos* for blunt and *opia* for vision).

Anisotropy: difference in the extent of crowding when measured along different axes.

Eccentricity: where in the visual field, relative to the locus of fixation, a stimulus appears.

Flanker: object or feature that is close to the target and is irrelevant to the observer's task. Sometimes referred to as a distractor.

Isotropy: uniformity in all directions.

Lateral interaction: capacity of a feature, object or neuron to influence the perception (or excitation) of a neighboring feature, object or neuron.

Macular degeneration: condition affecting mostly older adults that results in a loss of central vision due to degenerative changes in the macula (the part of the retina with the highest cone density that is responsible for high visual acuity and reading). It is the leading cause of blindness in Americans over 65 years of age.

Metaccontrast masking: type of backward visual masking (see below) in which the visibility of a brief target stimulus is reduced when a second, non-overlapping but adjacent stimulus is presented in quick succession.

Surround suppression: reduction in visibility of a visual stimulus or of neuronal firing rate caused by the presentation of a surrounding stimulus.

Visual masking: reduction in visibility of a visual stimulus (target) caused by the presentation of a second stimulus, either at the same time as the target (simultaneous masking), before the target (forward masking) or following the target (backward masking).

Corresponding author: Levi, D.M. (dlevi@berkeley.edu).



Figure 1. Visual crowding – the deleterious effect of clutter on peripheral object recognition – is ubiquitous in natural scenes. (a) Visual crowding seriously impacts virtually all everyday tasks, including reading, driving and interacting with the environment. For example, when fixating the bull’s eye near the construction zone, note that it is difficult or impossible to recognize the child on the left-hand side of the road, simply because of the presence of the nearby signs. Conversely, it is relatively easy to recognize the child on the right-hand side. (b) While fixating the crosses, identification of the middle shape, letter, or line orientation – or even the number of tilted lines – is difficult or impossible on the bottom half of the panel. Crowding impairs the ability to recognize and scrutinize objects, but it does not make them disappear; one can see that something is present in panel a, but it is difficult to identify it as a child as opposed to another sign. Crowding defines the spatial resolution of conscious object recognition throughout most of the visual field.

studies to individuate and distinguish crowding from other effects, such as masking, lateral interaction and surround suppression. All of these share the characteristic of making a target more difficult to see, but each is distinct and most probably reflects different neural processes.

Diagnostic criteria

- (i) Crowding impairs identification, not detection
When objects are crowded they do not simply disappear, as might be expected if crowding were a disruptive process that suppressed their signals. Figure 1b confirms that crowding has little or even no effect on detection of a feature or object [3–5].
- (ii) Crowded objects are perceived as having high contrast but are indistinct or jumbled together.
This can be easily confirmed from inspection of Figure 1b.
- (iii) Critical spacing: is Bouma’s rule a law?
Crowding depends on the eccentricity of a target object and how densely spaced the surrounding objects are (Figure 2). At a given eccentricity, identification of a crowded target improves as the distance between the target and flankers increases. In his highly influential report, Bouma stated ‘for complete visual isolation of a letter presented at an eccentricity of ϕ° , it follows that no other letters should be present (roughly) within $0.5 \phi^\circ$ distance.’ [6]. This gives rise to the notion of a critical spacing that is proportional to eccentricity. Bouma’s proportionality constant, b varies across studies, depending on how it is both measured and computed, but it is widely reported to be approximately 0.4–0.5. Thus, an object at an eccentricity of 10° might be crowded by other objects as much as 5° away. Bouma’s proportionality constant, or Bouma’s rule, is sometimes conferred the status of a law, but this is controversial (Box 1).
- (iv) Anisotropy
Crowding in peripheral vision is not isotropic. There is a very substantial radial–tangential anisotropy

(approx. 2:1) [7], such that radially positioned flankers are more effective than tangentially positioned ones. For example, in the vertical meridian, vertically arranged flankers are more potent than horizontally arranged flankers, whereas in the

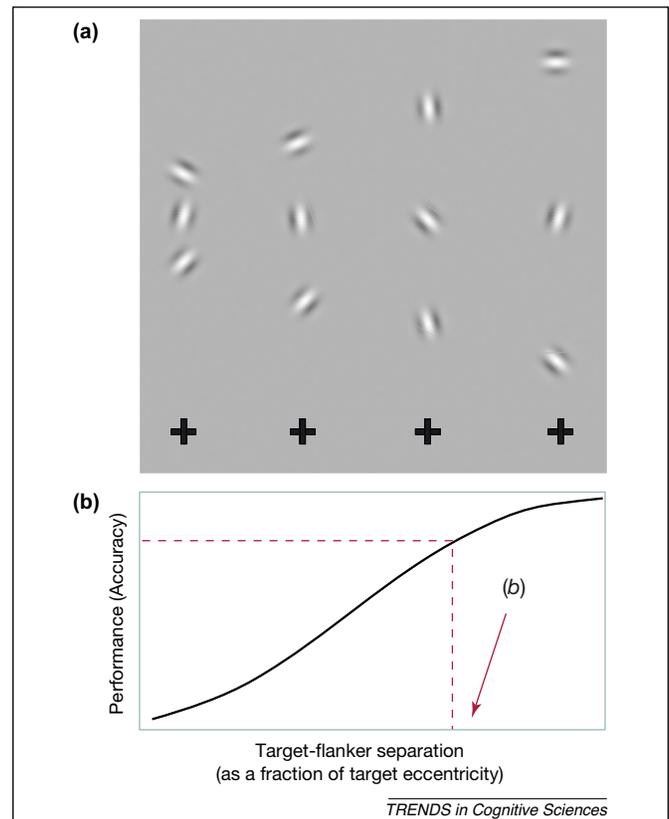


Figure 2. The critical spacing of crowding and Bouma’s proportionality constant b . (a) When fixating the crosses along the bottom, notice that the target orientation (central Gabor patch in each column) is easier to recognize on the right. (b) Performance accuracy increases as the target–flanker separation increases. Bouma’s constant, b , is defined as the target–flanker separation (as a ratio of target eccentricity) that results in criterion performance (shown by the dashed line). Although the analytic methods and criteria used to compute b vary from study to study, it generally corresponds to the point at which performance begins to decrease as flankers are advanced toward the target.

horizontal meridian, horizontally arranged flankers are more potent than vertically arranged ones. In diagonal locations of the four quadrants, crowding is significantly stronger (critical spacing is greater) when the target and distractors are horizontally rather than vertically arranged [8] (this does not occur on the vertical meridian because the radial–tangential anisotropy dominates). In addition, crowding is also stronger in the upper field than in the lower field [9]. Finally, a recent report suggests that crowding is also stronger when the distractors and target are within the same visual field than in separate visual fields, despite equal retinal distance [10].

(v) Asymmetry

Peripheral crowding is asymmetric. Bouma noted that two flankers (one on each side of the target letter) were much more potent than one, and that crowding was stronger with a single flanker at an eccentric locus greater than the target compared to a single flanker at an eccentric locus nearer to the fovea (at the same angular separation from the target) [6]. This inner–outer asymmetry occurs for recognition of letters [11], Gabor patch orientation [12] and face recognition [13]. Although this asymmetry might be readily explained on the basis of cortical geometry (the far flanker is actually closer to the target than the near flanker after mapping to cortical space [14]) we know of no similar explanation for the large radial–tangential anisotropy.

(vi) Temporal tuning

Although crowding is usually thought of as a spatial phenomenon, it also occurs over time. For example, moving objects are crowded more by flankers that slightly lead than trail the target [15] and the temporal relationship between the target and flankers modulates crowding [16–18]. This temporal tuning of crowding might be a diagnostic feature, or at least a distinguishing characteristic. That the crowding effect is strong for nearly simultaneous flankers and target [17,18] distinguishes it from object substitution and

metacrowding [19]. That the full-width at half-maximum of the temporal tuning function is approximately 150 ms and is not strongly skewed toward backward masking (i.e., targets presented before flankers are about as crowded as targets presented after flankers [17,18]) could further distinguish crowding from backward pattern masking. Whether there is an independent mechanism of temporal crowding remains unclear, but the effects of spatial crowding are correlated with those of temporal crowding across subjects [20], which supports the possibility that crowding involves spatiotemporal and perhaps attentional mechanisms [16,21].

Crowding and appearance

As discussed above, crowded objects do not simply disappear. On the contrary, crowding changes the appearance of the crowded zone, which is important because it can help to distinguish among the main models for crowding (masking, pooling and substitution, as discussed below). As examples, Tyler and Likova noted their impression of a crowded letter as a ‘gray, or inchoate, smudge between the two outer letters, including the inner parts of those letters’ [22] (see also Figures 1 and 2). Greenwood *et al.* elegantly showed that crowded objects seem to take on characteristics of the flankers, a finding that is consistent with the jumbled percept that accompanies crowding [23]. They conclude, in agreement with Levi and Carney [24], that crowding is a regularization process that simplifies the appearance of the peripheral array by promoting a consistent appearance among adjacent objects.

The results reported by Greenwood *et al.* are consistent with a number of earlier studies suggesting that information about crowded objects is not lost [23]. In particular, Parkes *et al.* found that the orientation signals from the target and flankers in a cluttered peripheral display were pooled rather than being lost through masking [25]. They concluded that crowding reflects compulsory averaging of signals (but see

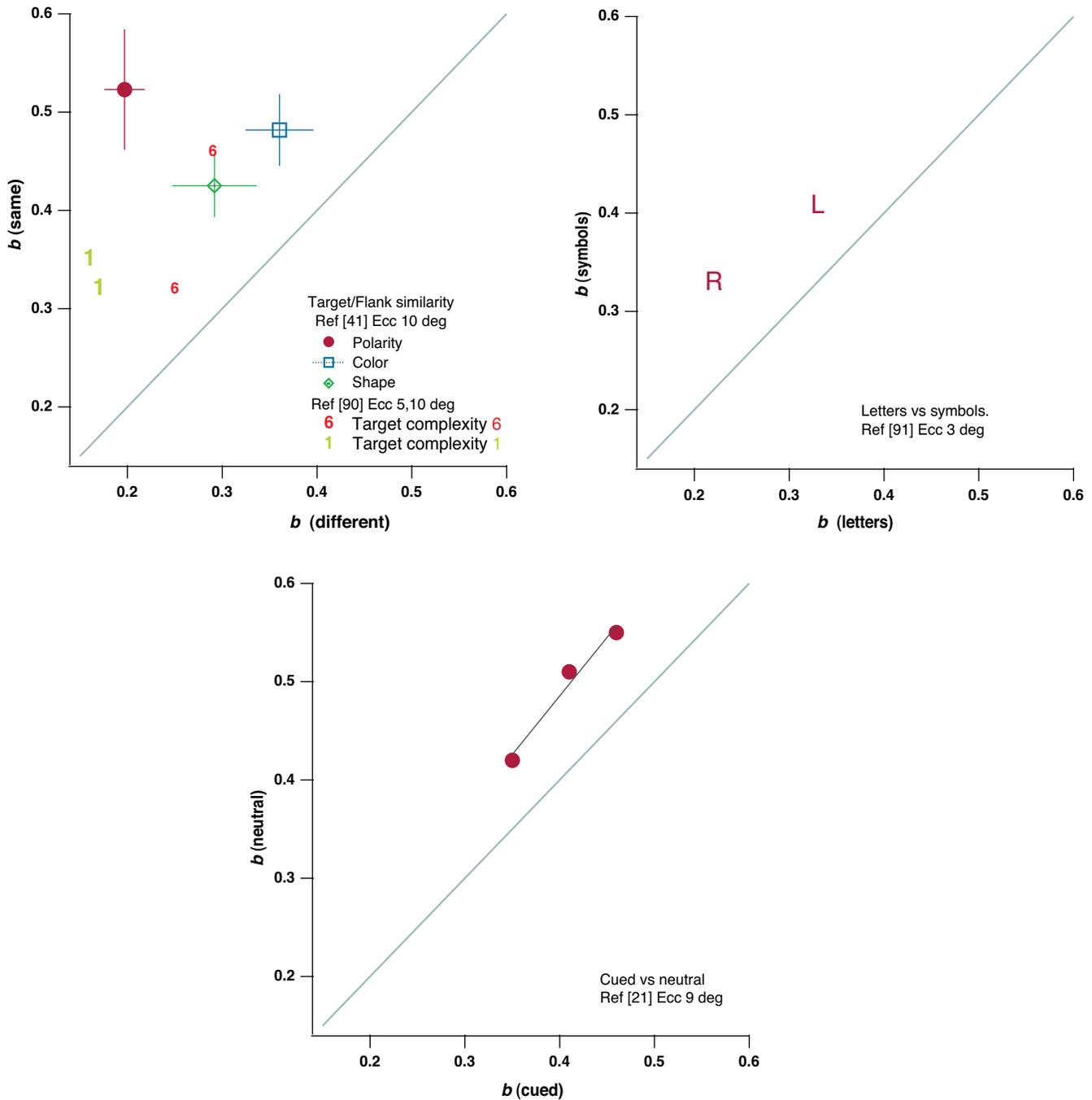
Box 1. Is Bouma’s law a law?

Many studies have confirmed that the critical spacing for crowding depends on target eccentricity, not target size [3–5,87]. It was recently suggested that Bouma’s rule of thumb should be elevated to the status of a law [2,67]. If it were a law, it would have important implications: because of the log conformal mapping of the visual world onto the retinotopic visual cortex [88], it suggests that regardless of the eccentricity, the critical distance represents a fixed distance on the cortex. Thus, objects can only be recognized when they are sufficiently separated on retinotopically organized cortex [67,89]. Pelli argued that the critical spacing on the cortex is ~6 mm in area V1 [67] (which is also the range of dichoptic interactions in the region of the cortex corresponding to the blind spot [66]).

The notion that crowding depends simply on cortical distance is simple and appealing. But is Bouma’s law truly a law? Is it physiologically plausible that the critical spacing is independent of stimulus, task, attention and other factors? Figure 1 suggests not. Specifically, the top left panel shows that Bouma’s so-called constant b is substantially greater when the polarity, color or shape of targets and flankers is the same compared to when they differ [41,46,47]. Figure 1 shows that b is also greater when the complexity of targets and flankers is the same compared to when they differ [90]. Recent work also suggests that b is greater when the target and flankers are letter-like symbols rather than actual letters (Figure 1, top right panel [91]) and that

b is smaller when the target location is cued compared to not cued (Figure 1, lower panel [21]). The study by Yeshurun and Rashal [21] differs from a number of previous studies that have shown either no effect or a small effect of cueing on crowding [42,71,92], because they took measures to avoid masking of the target by the cue. Even very weak target masking can greatly enhance the measured crowding effect [93], so it is crucial to control for interactions between crowding and masking. In addition to modulations by attention [21,39,60,61], b can be modulated by perceptual learning [94,95]. The implications of these and other related findings [13,57,96,97] are discussed in the main text (Figure 4), but this small sample of studies shows that Bouma’s constant strongly depends on a number of stimulus, task and attentional factors, as well as on the orientation of the target and flankers and their location in the visual field [7,9]. Differences in the methods used to calculate b cannot explain the asymmetries in Figure 1.

Collectively, the evidence shows that Bouma’s rule is not hard and fast. Critical spacing depends on stimulus characteristics, task requirements and attentional factors. Critical spacing is undoubtedly a hallmark of crowding, but Bouma’s rule should not be used in black-and-white terms or as a single value (e.g. 0.5). Rather, it should be understood as a continuum, and whether a particular stimulus is crowded should be evaluated based on where it falls on this continuum relative to other similar stimuli under similar conditions.



TRENDS in Cognitive Sciences

Figure 1. Bouma's constant (reflecting the critical spacing between target and distractors) varies, depending on the similarity of the targets and flankers and their complexity (top left panel), the type of stimuli used (top right panel) and the attentional requirements of the task (bottom panel). The critical spacing can vary systematically, which suggests that Bouma's rule is a rule-of-thumb and not a hard-and-fast law. Differences in methods and analytic approaches to calculating b cannot explain these graphs. The graphs show a within-study modulation of crowding (and b), depending on similarity, complexity, stimulus type and attention.

[26–28]) and that crowding is the term we use to define texture perception 'when we do not wish it to occur.'

Under conditions of crowding, orientation perception is characterized by strong perceptual assimilation (e.g. the flanker orientation captures the target) near the target and perceptual repulsion (i.e. anti-crowding) further from the target [27,29–31]. Assimilation could regularize perception of the peripheral array, whereas repulsion could highlight salient differences among visual signals (making different stimuli pop out). Assimilation and repulsion reflect oppo-

nent influences on orientation perception, and a recent study suggests that the switch from assimilation (crowding) to repulsion depends on cortical distance [30]. Whether this is specific to orientation or is more general is not yet known.

What information survives crowding?

The type of information that gets through to conscious perception under conditions of crowding provides important clues about the nature of crowding.

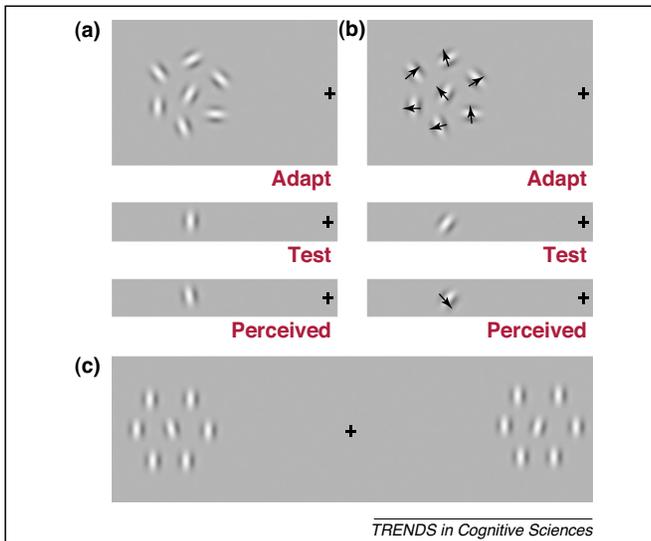


Figure 3. What escapes crowding? Although crowding limits conscious access to object identities in the periphery, there is much information that gets through or escapes crowding. (a,b) Adaptation to low-level features. Adapting to (a) the orientation or (b) the motion of a crowded pattern that is unidentifiable nevertheless causes strong local orientation and motion aftereffects, respectively. (c) While fixating the central cross, notice that the array of Gabor patches on the right seems to be tilted more clockwise, whereas the array on the left seems to be relatively more tilted counterclockwise. In fact, the only difference between the two arrays is the single central Gabor patch; all the flankers are identical. Crowded objects can be unidentifiable, but nevertheless contribute their features to the ensemble or texture.

Detection of a feature

Observers can easily detect the appearance of a feature under conditions that render identification or discrimination of a change in the feature impossible [3–5,24].

Aftereffects following adaptation to a crowded target

Whereas a target can be crowded and indistinguishable, adapting to it can produce a variety of aftereffects including orientation [9] and motion [32,33] (Figure 3a,b). Whether these aftereffects are modulated by crowding is debated [34], but their existence shows that crowding does not destroy feature-level information.

Statistical properties

As mentioned above, Parkes *et al.* elegantly demonstrated that although observers are unable to correctly report the orientation of an individual patch under conditions of crowding, they can reliably report the average ensemble orientation [25], which suggests that the local orientation signals are combined rather than lost (Figure 3c). This led to the widely held notion that crowded signals undergo a form of compulsory pooling or averaging of signals, a finding that has now been demonstrated under a variety of different conditions [24,35–37] and forms the basis of the faulty integration theory.

Some target identity information

A well-established effect of crowding is a substitution-like phenomenon: in crowded displays, observers frequently mistakenly report a flanker rather than the target. Whether this reflects positional uncertainty (the observer confuses the position of the flanker with that of the target) or simply the fact that the observer has to report something (if they are unable to see the crowded target, they simply

report what they could see, the flanker) is not clear. However, when required to report all the letters in a crowded display (give a full report of target and flanker letters), the proportion of correct target responses is much higher when the correct order (position) is not required [38,39]. For example, given a crowded display 'BTH', an observer might respond 'BHT'. Clearly, some information about the target is preserved (perhaps even semantic information [40]) but the location information is lost.

What information breaks crowding?

Under certain circumstances, crowding might be reduced or released completely.

Ungrouping of target and flankers

In peripheral vision, there is a predilection to perceptually group features into a *Gestalt*.

Target-flanker grouping When targets and flankers are similar, they are likely to be grouped, and when they are

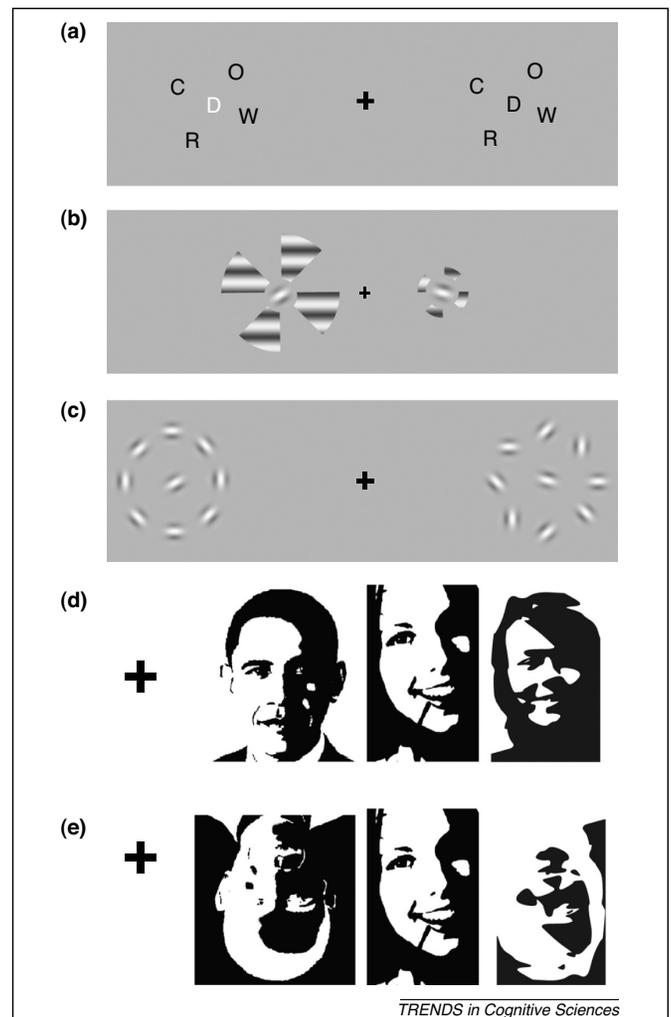


Figure 4. Crowding can be modulated or released depending on the nature of the flankers and the target-flanker relationship. (a) Reversing the contrast polarity of the target (left panel) reduces crowding compared to the right panel. (b) Flankers of the same size as the target crowd more effectively (more crowding in right panel). (c) Perceptual grouping of the flankers on the left reduces crowding (more crowding in the right panel). (d,e) Object-centered holistic crowding. Faces crowd each other and upright faces are more effective flankers than inverted ones. Crowding therefore occurs not just between features (Gabor patches, letters), but also between holistic representations of faces.

dissimilar they are ungrouped and the target pops out (Figure 4). Thus, crowding is reduced when targets and flankers are dissimilar in shape and size [41,42], orientation [4,43,44], polarity [16,41], spatial frequency [45], depth [41] color [41,46–48], synesthetic color to some degree [49], motion [50] and order (first- vs second-order [51]). Temporal grouping also modulates crowding [24]. Crowding is maximal when targets and flankers are presented nearly simultaneously; presentation of targets before or after the flankers (by ~150 ms) is sufficient to break crowding [17,18].

Flanker-flanker grouping In multi-element flankers, when the flankers are grouped separately from the target, crowding can be reduced [24,28,52–56]. Thus, when a target and flanker seem to have a regular texture, it is difficult to make judgments about the target and crowding is strong, whereas when the target seems to be distinct from the flankers, crowding is weak or absent [56].

Object-centered or holistic crowding

Crowding can occur between configural, high-level representations of objects. Inversion of face flankers can release face crowding. Specifically, it is harder to recognize an upright target face when it is surrounded by a crowd of nearby upright faces than by a crowd of inverted faces (Figure 4) [13,57]. The inversion effect in crowding also occurs between Mooney (two-tone) faces [13], stimuli that require holistic processing [58]. These object-centered crowding effects adhere to all of the diagnostic criteria for crowding and are not due to masking, similarity effects or grouping of low-level features. Likewise, holistic crowding – crowding between upright face representations – can be distinguished from within-face or facial feature crowding [13,59]. In fact, these object-centered crowding effects demonstrate compound crowding within the same stimulus – crowding between the whole upright faces, and between the low level features that comprise each face – which suggests that crowding operates at multiple stages. Recent work demonstrates that object-centered crowding effects also occur for letter-like stimuli [35], which raises the possibility that object-centered holistic crowding might occur independently at different levels of visual processing.

Attention

Several recent studies have demonstrated that cueing of a crowded target location reduces the effects of crowding [21,39,60,61]. Dakin and colleagues also found that deviation of attention to the target region ameliorated the effect of crowding [62]. The only explanation for these findings, as well as the demonstration by Cavanagh and Holcombe of attentionally gated crowding effects (*Vision Sciences Society Annual Meeting*, 2007), is that attention can modulate the critical spacing in crowded arrays (Box 1).

Masked flankers

Crowding can be released when flankers are masked. However, Chakravarthi and Cavanagh showed that this release only occurred with noise and metacontrast masks, and not with object substitution masks [63]. They argue that noise

and metacontrast masks act early in the visual processing cascade, degrading the features, whereas object substitution masks do not interfere with feature encoding, but act much later by replacing the representation of the stimulus.

Suppression of flankers from visual awareness

Wallis and Bex used adaptation-induced blindness to render flankers perceptually invisible and used a dual report paradigm to obtain a trial-by-trial assessment of awareness and crowding [64]. Target identification was dependent on the number of flanking letters perceived in a given trial, independent of the number that were physically present, and they concluded that crowding is released when flankers are suppressed from visual awareness.

Where in the brain does crowding take place?

There has been a great deal more psychophysical than neurophysiological work on crowding. Above all, neurophysiological studies of crowding have proven that it is surprisingly difficult to isolate the neural mechanism(s) of crowding *per se* (Box 2). Nevertheless, psychophysical studies have helped to narrow down the level(s) at which crowding occurs and will help to guide the design of more stringent future neurophysiological experiments.

Although there is great diversity in the stimuli used (ranging from oriented bars, Gabor patches and shapes to letters, words and faces, among others [1,2]), most studies on crowding implicitly (if not explicitly) argue that crowding is a unitary phenomenon, occurring at a single circumscribed level of visual processing, or perhaps in a particular visual area. Early work demonstrated that crowding works dichoptically (target to one eye, distractors to the other eye) [65,66], which suggests that crowding arises in the cortex. Since then, various researchers have suggested that the site of crowding might be V1 [67], V2 (Freeman and Simoncelli, *Vision Sciences Society Annual Meeting*, 2010), V3 [22,68], V4 [10,69] (but compare [70]) or even later in visual processing [32,57]. The evidence that crowding occurs in each of these and other visual areas is mixed, and difficulties comparing across studies is compounded by the fact that crowding can occur selectively between different kinds of stimuli (Box 1). Even within a stimulus type (e.g. orientation, motion or faces), crowding is modulated by stimulus similarity, context and attention (Figure 4, Box 1), as discussed above. Concluding that any single

Box 2. Neurophysiological studies of crowding

Very few neurophysiological studies have been conducted to measure crowding *per se*. Motter observed crowding-like effects in V4 neurons using letter-like stimuli (*Vision Sciences Society Annual Meeting*, 2002), although Merigan [70] found no effect of V4 lesions on crowding. Many other physiological studies have investigated responses to pairs or groups of stimuli, natural scenes and visual search arrays [98–100]. All of these would probably involve crowding to some extent. However, the stimuli in the majority of the neurophysiological studies were not presented to ensure or tightly control crowding, and thus are related in either an uncertain or indirect way. As discussed in the main text, there are very specific diagnostic criteria for what constitutes crowding, and future neurophysiological and imaging studies will need to specifically control and manipulate these factors to isolate the mechanism(s) of crowding.

visual area could explain this range of effects is therefore tenuous at best. Neurophysiological studies have provided relatively little additional evidence about the neural mechanism(s) of visual crowding (Box 2).

Models of crowding

There is no shortage of ideas about crowding, but few are computational or make specific quantitative predictions. The large number of different models can be distilled down to three basic classes: (i) masking, (ii) pooling (either of low-level features or by attention) and (iii) substitution. Within each class many different architectures and algorithms have been proposed. These extant models are largely descriptive and have been reviewed in detail elsewhere [1]. There are few quantitative models of crowding and most are quite recent. Wilkinson *et al.* proposed a model in which complex cells and simple cells interact by mutual inhibition [71]. In this model, isolated visual contours are processed by simple cells, which suppress weak complex-cell responses. However, in the presence of nearby similarly oriented flanking contours in a small area, complex cells respond vigorously because of spatial pooling, and they then suppress simple cell activity within their receptive field area. This texture model nicely predicts several aspects of their data; however, the pooling parameter was based on simulations with the best fit to the data rather than on physiology or some other principled approach.

There are three more recent approaches to modeling crowding. van den Berg and colleagues proposed a quantitative model for spatial integration of orientation signals that is based on the principles of population coding [72]. Their model nicely predicts several properties of crowding, including critical spacing, compulsory averaging and inner-outer asymmetry. However, in its current form it fails to predict the effect of target flank similarity [41,42], configuration effects [24,28,52] and object-centered or holistic crowding [13,35,57]. Dayan and Solomon take a very different approach, in which spatial selection of a target among flankers emerges through a process of Bayesian inference in a computational form [73]. Interference (also known as crowding) in this model results from the spatial uncertainty inherent in large receptive fields, and receptive field size is assumed to increase with eccentricity according to the cortical magnification factor. The model was developed to explain the Eriksen flanker task. It remains to be seen how well it accounts for many of the key features of crowding reviewed here.

None of the models naturally accounts for the radial/tangential anisotropy in a principled way. van den Berg *et al.* simply used different parameters to define the radial and tangential integration fields [72]. By contrast, Nandy and Tjan (*Vision Sciences Society Annual Meeting*, 2010) began with a model of cortical area V1 and its geometry and lateral connections, quite similar to the model of Neri and Levi [74], combined with the important role of natural image statistics [75]. However, the novel insight and advance in their model is the idea that image statistics are acquired primarily at attended spatial locations via a gating mechanism, and that spatial attention and any subsequent eye movement that it elicits overlap in time. Nandy and Tjan further argue that learning of image

statistics during development leads to the formation of lateral connections that distort the true image statistics in the peripheral field, which in turn leads to the radial-tangential anisotropy of crowding. Whether this model will be able to account for grouping, similarity and feature- and object-specific crowding remains to be seen.

Any successful computational model of crowding needs to account for each of the characteristics above, including the diagnostic criteria and the factors that modulate crowding. It is not sufficient for a model of crowding to simply mimic or reproduce the phenomenological jumble that is representative of crowding [75]; the model must conform to the diagnostic criteria for crowding and must account for what escapes crowding, what breaks crowding and what modulates crowding.

Multiple levels of crowding

Although there is great heterogeneity in results on crowding, there is sufficient evidence to cast doubt on the idea that crowding is a unitary effect due to a single stage of processing, although this is implicitly assumed in most studies on crowding [1,2]; rather, the collective work suggests that crowding happens independently at several stages of visual processing. In support of this view are observations that crowding is specific to the similarity between and the configuration of target and flanks (discussed above and in Box 1), and the fact that there is compound crowding: in a given scene, crowding occurs selectively between features [1], object parts [59] and whole objects [13,57]. These make an all-convergent crowding stage unlikely. If crowding occurs at multiple levels of visual analysis, or if different channels (chromatic, spatial frequency, object, etc.) each has its own unique crowding bottleneck, then we would expect that the crowding gradient as a function of eccentricity might be channel- or stimulus-specific. Intriguing evidence in favor of this view is increasing (Box 1), but much work remains (Box 3).

Box 3. Questions for future research

- Does crowding happen in different modalities (e.g. in audition, touch)? Does it happen crossmodally (visual flankers with an auditory target)? Does optimal cue combination (e.g. between vision and audition) reduce or counteract crowding?
- Is there a benefit of crowding? Ensemble statistics can be perceived in crowds, but are they always? Do crowding and ensemble perception share a common mechanism?
- How and when does crowding limit action? Does the motor system have a compensatory mechanism? Is there dissociation between the crowding of perception and action?
- What is the developmental time course of crowding and how does it impact the development of visually guided action? How does crowding change over the lifespan?
- Crowding reveals the spatial resolution of conscious vision; many neurodevelopmental disorders are anecdotally reported as having coarse-grained attention or perception. Is there a significant difference in crowding thresholds in clinical populations with autism, fragile X disorder, 22q deletion syndrome, Williams syndrome or other neurodevelopmental disorders?
- How can crowding be mitigated to improve reading, particularly in those with macular degeneration or amblyopia?
- Does the gradient of crowding as a function of eccentricity differ systematically for different channels or stimuli (e.g. chromatic, first vs second order, spatial frequency, shapes, letters, object parts, object wholes, etc.)?

In natural scenes, crowding can be ubiquitous, but it can occur in layers, with dependence on location, content and attention. One of the major challenges in future work on crowding will be to develop a parsimonious model that can account for the diversity of findings. Individual models that ignore content, category boundaries, similarity and attention are not likely to succeed.

The crowded future

Our discussion of crowding has been confined exclusively to visual perception, but the implications of crowding are far and wide and are likely to increase rapidly in the near future (Box 3). For example, crowding might exert a fundamental limit on visually guided actions in naturally cluttered scenes. Although the impact of crowding *per se* on visually guided action has rarely been investigated [14,76,77], several studies suggest that clutter impairs action [78,79]. More intriguingly, there is evidence that crowded visual information is differentially used by perceptual and visuomotor systems [76]. Such dissociations between the perceptual and motor responses to crowded scenes might help to address the paradoxical and unanswered question of why eye (and hand) movements are not more random than they are in natural scenes; if peripheral object identities are crowded and unrecognizable, how do we make accurate, non-random eye and hand movements towards those objects? The practical consequences of visuomotor crowding, as distinct from perceptual crowding, are extensive, ranging from clinical settings to ergonomics and human factors; therefore, the search should be for dissociations between perception and action in crowded scenes (Box 3).

Crowding limits visually guided action and defines the resolution of conscious visual perception, so another important question is how crowding develops from infancy to old age. Although little work has been done specifically on crowding in older populations [80], studies with infants as young as 6 months [81], toddlers [82] and adolescents from 8 to 11 years of age [83] have reported that children are much more impaired by crowding than adults, even when their acuity is fully developed [83]. Given this protracted development of crowding, an important question is what sorts of knock-on effects might occur later in life if crowding develops abnormally in infancy. For example, crowding might be uniquely and differentially impacted in neurodevelopmental, visual and cognitive disorders, including autism [84], dyslexia [85], amblyopia [86] and macular degeneration [1], among others. Therefore, the clinical and practical implications of crowding, as well as the value of its training and possible rehabilitation, are widespread.

Acknowledgements

This work was supported by grants from the National Eye Institute: R01EY01728 to D.L. and R01EY018216 and NSF CAREER 0748689 to D.W. We are grateful to Jason Fischer for helpful comments on an earlier draft of this manuscript. Both authors contributed equally to this work.

References

- Levi, D.M. (2008) Crowding – an essential bottleneck for object recognition: a mini-review. *Vision Res.* 48, 635–654
- Pelli, D.G. and Tillman, K.A. (2008) The uncrowded window of object recognition. *Nat. Neurosci.* 11, 1129–1135
- Levi, D.M. *et al.* (2002) Suppressive and facilitatory spatial interactions in peripheral vision: peripheral crowding is neither size invariant nor simple contrast masking. *J. Vis.* 2, 167–177
- Levi, D.M. *et al.* (2002) Suppressive and facilitatory spatial interactions in amblyopic vision. *Vision Res.* 42, 1379–1394
- Pelli, D.G. *et al.* (2004) Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J. Vis.* 4, 1136–1169
- Bouma, H. (1970) Interaction effects in parafoveal letter recognition. *Nature* 226, 177–178
- Toet, A. and Levi, D.M. (1992) The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Res.* 32, 1349–1357
- Feng, C. *et al.* (2007) Horizontal and vertical asymmetry in visual spatial crowding effects. *J. Vis.* 7, 13 1–10
- He, S. *et al.* (1996) Attentional resolution and the locus of visual awareness. *Nature* 383, 334–337
- Liu, T. *et al.* (2009) Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli. *Curr. Biol.* 19, 127–132
- Banks, W.P. *et al.* (1977) Asymmetry of lateral interference in visual letter identification. *Percept. Psychophys.* 22, 232–240
- Petrov, Y. *et al.* (2007) Crowding and surround suppression: not to be confused. *J. Vis.* 7, 12 1–9
- Farzin, F. *et al.* (2009) Holistic crowding of Mooney faces. *J. Vis.* 9(6), 18 11–15
- Motter, B.C. and Simoni, D.A. (2007) The roles of cortical image separation and size in active visual search performance. *J. Vis.* 7, 1–15
- Bex, P.J. *et al.* (2003) The shape and size of crowding for moving targets. *Vision Res.* 43, 2895–2904
- Chakravarthi, R. and Cavanagh, P. (2007) Temporal properties of the polarity advantage effect in crowding. *J. Vis.* 7(2), 11 1–13
- Huckauf, A. and Heller, D. (2004) On the relations between crowding and visual masking. *Percept. Psychophys.* 66, 584–595
- Ng, J. and Westheimer, G. (2002) Time course of masking in spatial resolution tasks. *Optom. Vis. Sci.* 79, 98–102
- Enns, J.T. (2004) Object substitution and its relation to other forms of visual masking. *Vision Res.* 44, 1321–1331
- Bonneh, Y.S. *et al.* (2007) Spatial and temporal crowding in amblyopia. *Vision Res.* 47, 1950–1962
- Yeshurun, Y. and Rashal, E. (2010) Precueing attention to the target location diminishes crowding and reduces the critical distance. *J. Vis.* 10, 1–12
- Tyler, C.W. and Likova, L.T. (2007) Crowding: a neuroanalytic approach. *J. Vis.* 7(2), 16 1–9
- Greenwood, J.A. *et al.* (2010) Crowding changes appearance. *Curr. Biol.* 20, 496–501
- Levi, D.M. and Carney, T. (2009) Crowding in peripheral vision: why bigger is better. *Curr. Biol.* 19, 1988–1993
- Parkes, L. *et al.* (2001) Compulsory averaging of crowded orientation signals in human vision. *Nat. Neurosci.* 4, 739–744
- Baldassi, S. *et al.* (2006) Visual clutter causes high-magnitude errors. *PLoS Biol.* 4, e56
- Bulakowski, P.F. *et al.* (2011) Reexamining the possible benefits of visual crowding: dissociating crowding from ensemble percepts. *Atten. Percept. Psychophys.* DOI: 10.3758/s13414-010-0086-2
- Livne, T. and Sagi, D. (2007) Configuration influence on crowding. *J. Vis.* 7(2), 4 1–12
- Felisbert, F.M. *et al.* (2005) The role of target salience in crowding. *Perception* 34, 823–833
- Mareschal, I. *et al.* (2010) Cortical distance determines whether flankers cause crowding or the tilt illusion. *J. Vis.* 10, 13
- Song, S. and Levi, D.M. (2010) Spatiotemporal mechanisms for simple image feature perception in normal and amblyopic vision. *J. Vis.* 10(13), 21 1–22
- Aghdaee, S.M. (2005) Adaptation to spiral motion in crowding condition. *Perception* 34, 155–162
- Whitney, D. (2005) Motion distorts perceived position without awareness of motion. *Curr. Biol.* 15, R324–326
- Blake, R. *et al.* (2006) Strength of early visual adaptation depends on visual awareness. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4783–4788
- Dakin, S.C. *et al.* (2010) Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *J. Vis.* 10, 1–16
- Greenwood, J.A. *et al.* (2009) Positional averaging explains crowding with letter-like stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13130–13135

- 37 Haberman, J. and Whitney, D. (2007) Rapid extraction of mean emotion and gender from sets of faces. *Curr. Biol.* 17, R751–753
- 38 Popple, A.V. and Levi, D.M. (2005) The perception of spatial order at a glance. *Vision Res.* 45, 1085–1090
- 39 Strasburger, H. (2005) Unfocused spatial attention underlies the crowding effect in indirect form vision. *J. Vis.* 5, 1024–1037
- 40 Huckauf, A. et al. (2008) Semantic processing of crowded stimuli? *Psychol. Res.* 72, 648–656
- 41 Kooi, F.L. et al. (1994) The effect of similarity and duration on spatial interaction in peripheral vision. *Spat. Vis.* 8, 255–279
- 42 Nazir, T.A. (1992) Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Res.* 32, 771–777
- 43 Andriessen, J.J. and Bouma, H. (1976) Eccentric vision – adverse interactions between line segments. *Vision Res.* 16, 71–78
- 44 Hariharan, S. et al. (2005) ‘Crowding’ in normal and amblyopic vision assessed with Gaussian and Gabor C’s. *Vision Res.* 45, 617–633
- 45 Chung, S.T. et al. (2001) Spatial-frequency and contrast properties of crowding. *Vision Res.* 41, 1833–1850
- 46 Gheri, C. et al. (2007) The relationship between search efficiency and crowding. *Perception* 36, 1779–1787
- 47 Kennedy, G.J. and Whitaker, D. (2010) The chromatic selectivity of visual crowding. *J. Vis.* 10, 15
- 48 Pöder, E. and Wagemans, J. (2007) Crowding with conjunctions of simple features. *J. Vis.* 7(2), 23 1–12
- 49 Hubbard, E.M. et al. (2005) Individual differences among grapheme-color synesthetes: brain–behavior correlations. *Neuron* 45, 975–985
- 50 Banton, T. and Levi, D.M. (1993) Spatial localization of motion-defined and luminance-defined contours. *Vision Res.* 33, 2225–2237
- 51 Chung, S.T. et al. (2007) Crowding between first- and second-order letter stimuli in normal foveal and peripheral vision. *J. Vis.* 7(2), 10 1–13
- 52 Livne, T. and Sagi, D. (2010) How do flankers’ relations affect crowding? *J. Vis.* 10(3), 1 1–14
- 53 Malania, M. et al. (2007) Grouping of contextual elements that affect Vernier thresholds. *J. Vis.* 7(2), 1 1–7
- 54 Pöder, E. (2006) Crowding, feature integration, and two kinds of ‘attention’. *J. Vis.* 6, 163–169
- 55 Saarela, T.P. et al. (2009) Global stimulus configuration modulates crowding. *J. Vis.* 9(2), 5 1–11
- 56 Saarela, T.P. et al. (2010) The effect of spacing regularity on visual crowding. *J. Vis.* 10, 1–7
- 57 Louie, E.G. et al. (2007) Holistic crowding: selective interference between configural representations of faces in crowded scenes. *J. Vis.* 7(2), 24 1–11
- 58 Cavanagh, P. (1991) What’s up in top-down processing? In *Representation of Vision: Trends and Tacit Assumptions in Vision Research* (Gorea, A., ed.), pp. 295–304, Cambridge Press
- 59 Martelli, M. et al. (2005) Are faces processed like words? A diagnostic test for recognition by parts. *J. Vis.* 5, 58–70
- 60 Chakravarthi, R. and Cavanagh, P. (2009) Bilateral field advantage in visual crowding. *Vision Res.* 49, 1638–1646
- 61 Freeman, J. and Pelli, D.G. (2007) An escape from crowding. *J. Vis.* 7(2), 22 1–14
- 62 Dakin, S.C. et al. (2009) Dissociable effects of attention and crowding on orientation averaging. *J. Vis.* 9(11), 28 1–16
- 63 Chakravarthi, R. and Cavanagh, P. (2009) Recovery of a crowded object by masking the flankers: determining the locus of feature integration. *J. Vis.* 9, 1–9
- 64 Wallis, T.S. and Bex, P.J. (2011) Visual crowding is correlated with awareness. *Curr. Biol.* 21, 254–258
- 65 Flom, M.C. et al. (1963) Contour interaction and visual resolution: contralateral effects. *Science* 142, 979–980
- 66 Tripathy, S.P. and Levi, D.M. (1994) Long-range dichoptic interactions in the human visual cortex in the region corresponding to the blind spot. *Vision Res.* 34, 1127–1138
- 67 Pelli, D.G. (2008) Crowding: a cortical constraint on object recognition. *Curr. Opin. Neurobiol.* 18, 445–451
- 68 Bi, T. et al. (2009) The effect of crowding on orientation-selective adaptation in human early visual cortex. *J. Vis.* 9(11), 13 1–10
- 69 Motter, B.C. (2006) Modulation of transient and sustained response components of V4 neurons by temporal crowding in flashed stimulus sequences. *J. Neurosci.* 26, 9683–9694
- 70 Merigan, W.H. (2000) Cortical area V4 is critical for certain texture discriminations, but this effect is not dependent on attention. *Vis. Neurosci.* 17, 949–958
- 71 Wilkinson, F. et al. (1997) Lateral interactions in peripherally viewed texture arrays. *J. Opt. Soc. Am. A. Opt. Image Sci. Vis.* 14, 2057–2068
- 72 van den Berg, R. et al. (2010) A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Comput. Biol.* 6, e1000646
- 73 Dayan, P. and Solomon, J.A. (2010) Selective Bayes: attentional load and crowding. *Vision Res.* 50, 2248–2260
- 74 Neri, P. and Levi, D.M. (2006) Spatial resolution for feature binding is impaired in peripheral and amblyopic vision. *J. Neurophysiol.* 96, 142–153
- 75 Balas, B. et al. (2009) A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9(12), 13 1–18
- 76 Bulakowski, P.F. et al. (2009) Visuomotor crowding: the resolution of grasping in cluttered scenes. *Front Behav. Neurosci.* 3, 49
- 77 Vlaskamp, B.N. and Hooge, I.T. (2006) Crowding degrades saccadic search performance. *Vision Res.* 46, 417–425
- 78 Cohen, E.H. et al. (2007) The relationship between spatial pooling and attention in saccadic and perceptual tasks. *Vision Res.* 47, 1907–1923
- 79 Meegan, D.V. and Tipper, S.P. (1998) Reaching into cluttered visual environments: spatial and temporal influences of distracting objects. *Q. J. Exp. Psychol. A* 51, 225–249
- 80 Habak, C. et al. (2009) Preservation of shape discrimination in aging. *J. Vis.* 9(12), 1–18
- 81 Farzin, F. et al. (2010) Spatial resolution of conscious visual perception in infants. *Psychol. Sci.* 21, 1502–1509
- 82 Atkinson, J. et al. (1988) Visual acuity testing of young children with the Cambridge Crowding Cards at 3 and 6 m. *Acta Ophthalmol. (Copenh.)* 66, 505–508
- 83 Jeon, S.T. et al. (2010) Developmental changes during childhood in single-letter acuity and its crowding by surrounding contours. *J. Exp. Child Psychol.* 107, 423–437
- 84 Keita, L. et al. (2010) Far visual acuity is unremarkable in autism: do we need to focus on crowding? *Autism. Res.* 333–341
- 85 Martelli, M. et al. (2009) Crowding, reading, and developmental dyslexia. *J. Vis.* 9(4), 14 1–18
- 86 Levi, D.M. and Klein, S.A. (1985) Vernier acuity, crowding and amblyopia. *Vision Res.* 25, 979–991
- 87 Tripathy, S.P. and Cavanagh, P. (2002) The extent of crowding in peripheral vision does not scale with target size. *Vision Res.* 42, 2357–2369
- 88 Schwartz, E.L. (1980) Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Res.* 20, 645–669
- 89 Levi, D.M. et al. (1985) Vernier acuity, crowding and cortical magnification. *Vision Res.* 25, 963–977
- 90 Zhang, J.Y. et al. (2009) Legibility of Chinese characters in peripheral vision and the top-down influences on crowding. *Vision Res.* 49, 44–53
- 91 Grainger, J. et al. (2010) Crowding affects letters and symbols differently. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 673–688
- 92 Scolari, M. et al. (2007) Spatial attention, preview, and popout: which factors influence critical spacing in crowded displays? *J. Vis.* 7(2), 7 1–23
- 93 Vickery, T.J. et al. (2009) Supercrowding: weakly masking a target expands the range of crowding. *J. Vis.* 9(2), 12 1–15
- 94 Chung, S.T. (2007) Learning to identify crowded letters: does it improve reading speed? *Vision Res.* 47, 3150–3159
- 95 Sun, G.J. et al. (2010) Ideal observer analysis of crowding and the reduction of crowding through learning. *J. Vis.* 10, 1–14
- 96 Latham, K. and Whitaker, D. (1996) Relative roles of resolution and spatial interference in foveal and peripheral vision. *Ophthalmic Physiol. Opt.* 16, 49–57
- 97 Huckauf, A. (2007) Task set determines the amount of crowding. *Psychol. Res.* 71, 646–652
- 98 Bichot, N.P. and Desimone, R. (2006) Finding a face in the crowd: parallel and serial neural mechanisms of visual selection. *Prog. Brain Res.* 155, 147–156
- 99 Gallant, J. (2003) Neural mechanisms of natural scene perception. In *The Visual Neurosciences* (LM, C. and JS, W., eds), pp. 1590–1602, MIT Press
- 100 Chelazzi, L. (1999) Serial attention mechanisms in visual search: a critical look at the evidence. *Psychol. Res.* 62, 195–219