Multi-level Crowding and the Paradox of Object Recognition in Clutter

Mauro Manassi[1] and David Whitney[1,2,3]
[1]Department of Psychology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA 94720, USA
[3]Vision Science Program, University of California, Berkeley, Berkeley, CA 94720, USA
Correspondence: mauro.manassi@berkeley.edu
https://doi.org/10.1016/j.cub.2017.12.051

In everyday life, we are constantly surrounded by complex and cluttered scenes. In such cluttered environments, visual perception is primarily limited by crowding, the deleterious influence of nearby objects on object recognition. For the past several decades, visual crowding was assumed to occur at a single stage, only between low-level features or object parts, thus dismantling, destroying, or filtering object information. A large and converging body of evidence has demonstrated that this assumption is false: crowding occurs at multiple stages of visual analysis, and information passes through crowding at each of these stages. This converging empirical evidence points to a seeming paradox: crowding happens at multiple levels, which would seem to impair object recognition, and yet visual information at each of those levels is maintained intact and influences subsequent higher-level visual processing. Thus, while crowding impairs the *access* we have to visual information at many levels, it does not impair the *representation* of that information. The resolution of this paradox reveals how the visual system strikes a balance between the limits of object selection and the desire to represent multiple levels of visual information throughout cluttered scenes. Understanding crowding is therefore key to resolving the relationship between the richness of object and scene representations and the limits of conscious object recognition.

Introduction

Our visual system is continuously confronted with an enormous amount of information: objects, faces, letters and simple features are constantly present in our visual field. This presents a fundamental limitation to object recognition throughout most of the visual field most of the time, known as visual crowding: objects that can be easily identified in isolation seem jumbled and indistinct in clutter (Figure 1) [1–3]. Crowding is a fundamental limit for visual processing, impeding visually guided actions, reading, and more generally sets the resolution of conscious object individuation [3]. Importantly, crowding is considered a breakdown in object recognition [2], degrading objects to a fuzzy amalgamation and, hence, it is considered an invaluable tool for understanding basic visual processing and visual awareness [4]. Consequently, a great deal of research in the last several decades has focused on uncovering its neural locus and underlying mechanisms.

Proposed mechanisms of crowding can be summarized into three main classes: pooling, substitution and attention. Pooling models explain crowding in terms of low-level averaging [5], population processing [6] or summary statistics [7,8]. Substitution models propose that crowding occurs because of substitution between target and flankers' features [9,10]. The attentional account proposes that crowding is due to poor resolution of attention in the peripheral field [11]. Importantly, crowding is determined by perceptual grouping [12], which is considered a necessary component for any proposed crowding mechanism.

Although most crowding models successfully capture the standard characteristics of crowding [6,7], nearly all implicitly (or explicitly) have two underlying assumptions: first, crowding operates only at one stage, only on relatively low-level visual features; second, very little if any object-specific information gets through the bottleneck of crowding. The first premise is that crowding occurs only at one stage of visual processing, for example between low-level features or object parts, like oriented lines or letter-like features. The second premise is that crowding dismantles, destroys or filters information. This prevents any further potential high-level object processing: crowding, therefore, dramatically reduces the amount of information available past the stage at which crowding happens.

We review here a large and converging body of recent evidence that disproves these two assumptions. First, crowding occurs at multiple levels of visual analysis, including selectively between configural object representations (Figure 2). Second, object-selective information gets through crowding; crowding does not destroy or result in the loss of object specific information (Figure 3). The converging empirical evidence points to a seeming paradox: crowding happens at multiple levels, thus impairing object recognition at several stages of perceptual analysis; yet, represented visual information at each of those levels is maintained intact and carried forward, influencing subsequent visual processing. We argue that any viable model of crowding needs to resolve this seeming paradox. The resolution of the crowding paradox should therefore be a defining litmus test for crowding models. More broadly, because crowding is the fundamental limit on object recognition throughout natural scenes, reconciling the crowding paradox is critical to understanding object and scene perception in general.

What Crowds in Crowding?

Visual crowding has been thoroughly examined using a variety of stimuli, ranging from simple visual features, such as oriented

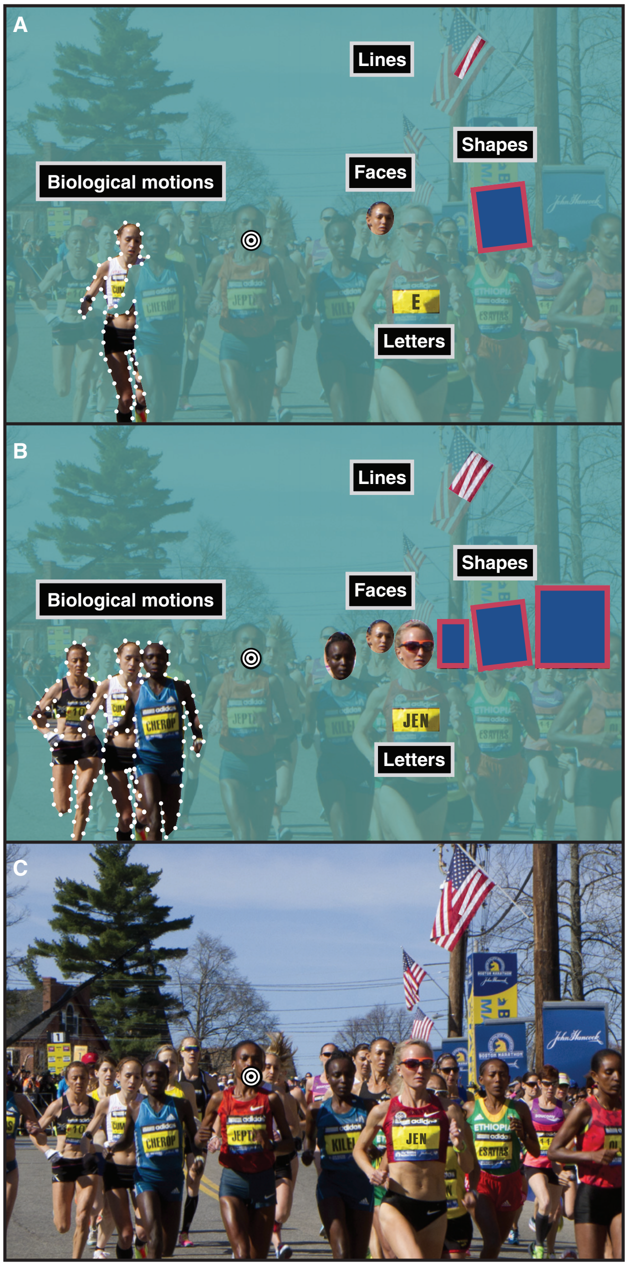


Figure 1. Multi-level crowding in real scenes.
(A) When fixating the bull's eye in the middle, it is relatively easy to identify isolated visual features and objects. This includes features and objects at different levels of visual analysis: the oriented line of the flag (lines), the tilted blue banner (shapes), the letter 'E' on the race bib (letters), the face on the right side (faces), and the (hypothetical) motion direction of the runner on the left side (biological motion). (B) In crowding, nearby flanking objects impair identification of the visual target objects, making them appear jumbled and harder to recognize. While staring at the bull's eye, it is more difficult to recognize the exact orientation of the central line on the flag because of the flanking white stripes (lines), and the orientation and aspect ratio of the central banner because of the flanking banners (shapes). Crowding does not only occur with simple features but also between objects; it is more difficult to read the full name 'JEN' because of flanking 'J' and 'N' on the race bib (letters), and it is harder to recognize the identity on the face because of flanking faces (faces). Crowding occurs also between motion directions; in a real scene, it would be much more difficult to identify the heading direction of the central runner because of flanking runners (biological motion). Hence, crowding happens at many stages of visual processing, from low-level features to high-level object, face, and body representations. (C) Natural scenes are filled with a variety of sorts of clutter, including different visual features, surfaces, objects, faces, etcetera. Here, crowding in real scenes is significant and ubiquitous, impairing our ability to recognize most particular objects. Photo: JD/Flickr.

lines and gratings, to faces, objects, and biological motion. Previous reviews [1,2] have extensively described crowding of low-level features, and there is little debate over the existence of crowding for these types of stimuli. However, crowding has recently been shown to occur also between high-level stimuli. Here, we review recent evidence that crowding occurs at multiple levels of visual analysis, from simple basic features to complex dynamic configurations (Figure 2).

Features: Lines and Orientations

Crowding occurs between low-level features [1], like edges (Figure 2A,B) [13] and orientations (Figure 2C,D) [5]. Low-level features do not crowd by their mere physical properties, but only after perceptual grouping has been computed between target and flankers (Figure 2A,B) [13,14]. These results, in accordance with others, highlight the importance of a grouping stage in crowding, which strongly determines crowding strength [12]. Along the same lines, crowding between low-level features occurs after the *perceived* position of target and flankers is assigned (Figure 2C,D) [15,16].

Object Parts

Martelli *et al.* [17] and Sun and Balas [18] found that crowding can occur between an object's individual parts. Martelli *et al.* [17] measured threshold contrast for identifying parts of a face, and found that face-part recognition (for example, mouth) depends on the spacing between the target part and surrounding parts (Figure 2E,F). Hence, object parts can crowd themselves within an object.

Object Configurations: Shapes

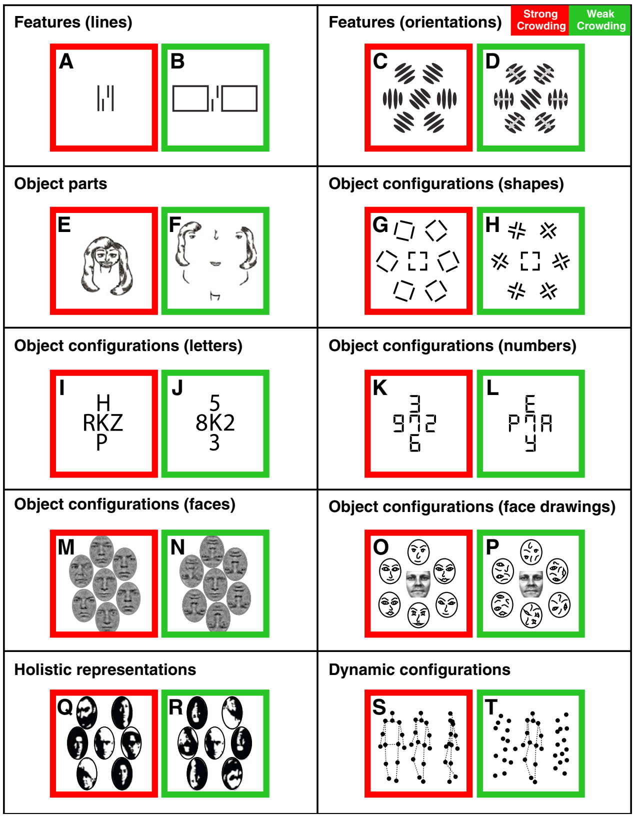
Kimchi and Pirkner [19] showed that a single object (a disconnected square) is crowded by flankers that are similar to the target in global configuration even when the parts are dissimilar (Figure 2G). Conversely, crowding is weaker with flankers that are dissimilar in global configuration but similar in parts (Figure 2H). Hence, they proposed that crowding occurs not only between low-level features (or parts), but also at a level where low-level features are integrated into configural or object representations.

Object Configurations: Letters and Numbers

Huckauf *et al.* [20] found that flanking letters crowd recognition of a target letter more than flanking numbers, meaning that categorical similarity between target and flankers can determine crowding strength (Figure 2I,J). More recently, Reuther and Chakravarthi [21] showed that this effect is not due to local featural differences between the two categories (letters and numbers); using a special font that equates for these differences, they still found that letters and numbers are crowded more by flankers belonging to the same category compared to flankers belonging to different categories (Figure 2K,L). Therefore, crowding between letters and numbers does not occur only at the feature-level but also involves higher-level interactions [20,21].

Object Configurations: Faces and Face Drawings

Louie *et al.* [22] showed that crowding can occur between configural representations of faces. Upright face flankers crowd



Current Biology

Figure 2. Crowding occurs at multiple levels of visual analysis.
Each of the stimulus displays in this figure are normally viewed in the visual periphery. The red boxes indicate conditions that generate strong crowding, where the central target is difficult to recognize due to the presence of flankers. The green boxes indicate conditions that reduce crowding, where the central target object is easier to identify. Features (lines): it is hard to recognize the misalignment between the central lines because of flanking lines (A). However, when the flanking lines are extended to create two rectangles (B), recognition of the central misalignment improves [13,14]. Features (orientations): it is hard to recognize the central orientation because of the flanking orientations (C). However, when the flanking orientations drift away from the target orientation, target discrimination improves (D) [15,16]. Object parts: when face parts are presented close together, it is very difficult to identify the mouth (E). When the same face parts are presented further apart, mouth identification is much easier (F) [17]. (E,F) Republished with permission of the Association for Research in Vision and Ophthalmology, from [17]. Object configurations (shapes): recognizing whether the central shape is square-like or diamond-like is more difficult when the flankers' configurations are similar to the target (flanking squares, G) compared to when flankers' parts are similar (flanking Ls, H) [19]. Object configurations (letters and numbers): identifying the middle letter 'K' is more impaired by flanking letters than by flanking numbers (I,J) [20]. Identifying the middle number '7' is more impaired by flanking numbers than by flanking letters (K,L) [21]. Object configurations (faces): identifying a central target face is crowded by flanking upright faces (M) more than flanking inverted faces (N) [22]. (M,N) Republished with permission of the Association for Research in Vision and Ophthalmology, from [22]. Object configurations (face drawings): crowding strength is higher when flankers are line drawings of faces (O) compared to when the drawings have scrambled face parts (P) [18]. (O,P) Reprinted by permission from Springer Nature [18], copyright 2014. Holistic representations: recognizing whether a two-tone Mooney face is male or female is crowded more with upright flanking Mooney faces (Q) than inverted flanking Mooney faces (R) [25]. (Q,R) Republished with permission of the Association for Research in Vision and Ophthalmology, from [25]. Dynamic configurations: recognizing the heading or action of point-light-walker motion is crowded more when flankers are biological motion stimuli (S) compared to scrambled motion stimuli (T) [26]. (S,T) Republished with permission of the Association for Research in Vision and Ophthalmology, from [26].

upright target face recognition more than inverted flankers (Figure 2M,N). Moreover, Sun and Balas [18] also found crowding between configural face representations; identifying the sex of a face (male/female) was crowded more by line drawings of faces than scrambled faces (Figure 2O,P). Kalpadakis-Smith, *et al.* [23] suggested that it may be the similarity between object parts that drives these kinds of crowding effects, but Sun and Balas [18] and other results showed that what counts as similar is based on how face-like or configurally correct the face is. Because only high-level visual areas represent configural face properties [24], face crowding must be mediated by a high level of visual processing, beyond that for feature-selective crowding.

Holistic Representations

Face-selective crowding even happens between Mooney faces, two-tone stimuli that require holistic recognition before any features of the face can be identified: upright Mooney face flankers crowd upright Mooney face recognition more than inverted flankers (Figure 2Q,R) [25].

Dynamic Configurations

Ikeda *et al.* [26] found evidence for crowding between high-level motion representations. Three point-light walker (human biological motion) stimuli were presented and observers were asked to indicate the motion direction of the central walker. Crowding was stronger with flanking point-light walker stimuli compared to when the flankers were scrambled (Figure 2S,T), despite equating the local motion signals. The results indicate that crowding can occur selectively between high-level representations of human motion. Interestingly, crowding was modulated also during action discrimination, that is, when target and flankers perform different actions [27].

Taken together, these results provide a broad and consistent range of empirical evidence that crowding can occur within different levels of representation, from low-level features to high-level dynamic objects (Figure 2). Crowding at lower levels does not predict or explain the crowding measured with high-level objects, configurations, faces, or biological motion. Importantly, however, the existence of high-level crowding does not deny nor preclude the existence of low-level crowding; crowding can occur between low-level features, as well as between high-level representations. For example, crowding among faces can occur between contours, shapes, face parts, and holistic face representations [17,19,22,25].

What Gets Through Crowding?

Standard crowding models assume a loss of information at relatively early stages of visual processing. According to this assumption, very little object-level information can survive in a crowded scene; what remains is an indistinguishable jumble of stimulus features from the target and flankers. But if crowding occurs within different levels of representation (Figure 2), one could infer that information about the crowded objects might be at least partially preserved. This inference is supported by direct experimental evidence using a variety of techniques, which shows that target-specific information survives at multiple stages of crowding. Here we review that evidence (Figure 3).

Orientation and Ensemble Orientation

Three different psychophysical techniques show that orientation information passes through crowding (Figure 3A,B). First, negative aftereffects: when adapting to an object, its features can bias

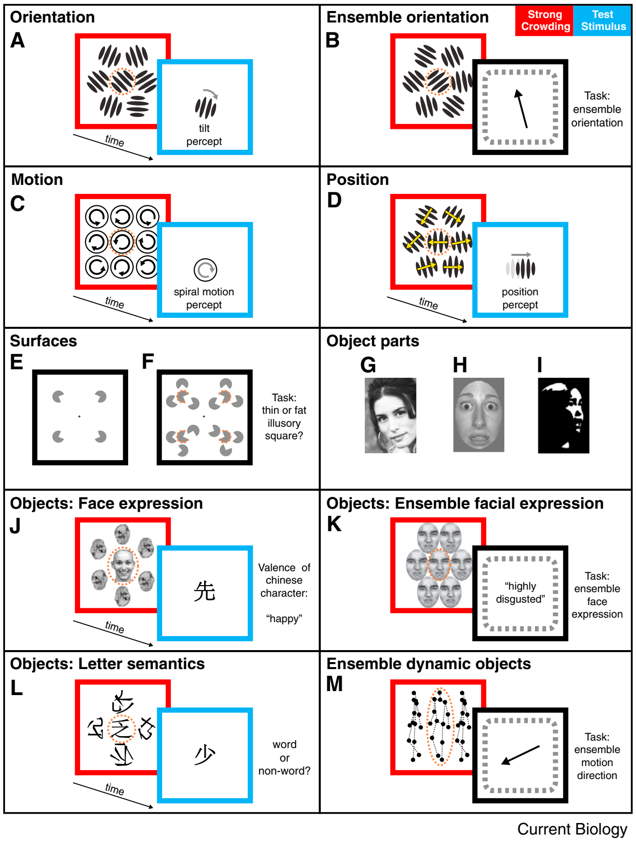


Figure 3. Information passes through crowding at multiple levels of visual analysis.
The red boxes indicate conditions that generate strong crowding, where the central target (highlighted by a dashed orange circle line) is difficult to recognize due to the presence of flankers. The blue boxes indicate stimuli that follow the crowded objects. Orientation: following adaptation to a crowded orientation, aftereffects occur, such as the appearance of an illusory tilt away from the adapting orientation (A) [11]. Ensemble orientation: when observers are asked to report the ensemble orientation of multiple gratings, the crowded orientation still contributes to the perceived ensemble orientation (B) [5]. Motion: when adapting to a crowded array of randomly moving spirals, the motion of a crowded spiral can bias spiral motion perception of a following ambiguous spiral in the opposite direction of the adaptor (C) [30]. Position: when adapting to a crowded motion array, the position of a subsequent test stimulus can be shifted in the opposite motion direction of the adaptor (D) [32]. Surfaces: observers are asked to indicate whether the perceived illusory rectangle is thin or fat (panel E shows a thin illusory rectangle). Observers are able to discriminate the thin shape of the illusory shape even when the pac-men inducers are crowded (F) [33]. Object parts: faces (G) (adapted from [34]), facial expressions (H) (adapted from [35]) and Mooney faces (I) (adapted from [36]) can be recognized in the visual periphery. Hence, face recognition must allow face parts to survive crowding within a face. Facial expression: after viewing a crowded facial expression, the pleasantness of a following Chinese character is rated in accordance with the previous face expression (J) (adapted from [37]). Ensemble facial expression: despite being unrecognizable because of crowding, the central face expression still contributes to the perceived average ensemble face expression in a group of faces (K) (adapted from [38]). Letter semantics: after viewing a crowded Chinese character, reaction times in recognizing whether the following character is a word or not are faster when the two characters are semantically related (L) [39]. Ensemble dynamic objects: when observers are asked to report the walking direction of the central figure, the reported walking direction reflects to some extent the average walking direction of the target and two flankers (M) (republished with permission of the Association for Research in Vision and Ophthalmology, from [26]).

the percept of a subsequent object in the opposite direction of the adaptor. For example, adapting to the orientation of a crowded object biases subsequent percepts away from the adapting orientation (Figure 3A) [11,28]. Second, reaction times: short or long adaptation to a crowded orientation can speed up or slow down reaction times when identifying the orientation of a subsequent grating (Figure 3A) [29]). Third, ensemble representations: although observers cannot correctly report the orientation of a crowded orientation, they can still reliably report the average ensemble orientation. Hence, the crowded orientation, although unrecognizable, still contributes to the ensemble orientation (Figure 3B) [5].

Motion

Crowded visual motion information can also bias subsequent percepts. Adaptation to complex types of motion, like spiral motion, biases the perceived motion of a subsequently viewed directionally ambiguous stimulus in the opposite direction (Figure 3C) [30,31].

Position

Crowded motion can shift the perceived position of a subsequent stationary object. Adapting to an array of randomly moving gratings can shift the perceived position of a subsequent grating in a direction opposite the previous motion at the same location (Figure 3D) [32].

Surfaces

Illusory surface information passes through crowding. By presenting four Pac-men at four corners, Lau and Cheung [33] created an illusory thin/fat Kanizsa surface (Figure 3E). Even when the crowded Pac-men had unrecognizable rotations, observers were able to discriminate the illusory surface's shape (Figure 3F). Hence, illusory surface information can pass through crowding, independent of the crowding that impairs recognition of the individual inducers.

Object Parts

Faces [34], expressions [35] and Mooney faces [36] can be recognized in the peripheral visual field (Figure 3G–I). In order to allow peripheral face recognition, facial features must be processed to a level of configural representation, without crowding themselves. Hence, although face parts can crowd each other [17], the representation of these object parts nevertheless survives crowding and allows configural/holistic object recognition.

Objects: Facial Expression and Ensemble Facial Expression

When observers are presented with a crowded face expression, they rate the pleasantness of a subsequent neutral Chinese character to be more consistent with the emotion of the crowded face (Figure 3J) [37]. Along the same lines, a crowded face can nonetheless influence the perceived ensemble expression in a group of faces (Figure 3K) [38]. These results hold even when crowding is complete (the faces are indiscriminable or incorrectly recognized). They also hold selectively for upright faces with intact configural information. Thus, it is not an errant feature or a variation in the strength of crowding that explains the results; instead, the findings suggest that representations of whole upright faces can survive crowding to influence subsequent perceptual judgments.

Objects: Letter Semantics

Even semantic information can pass through crowding. A crowded and unrecognizable Chinese character that is followed by another target character results in faster recognition of the target character when both items are semantically related (Figure 3L) [39]. Zhou *et al.* [40] provided also a neural correlate for this semantic priming effect: semantically related items elicit a smaller N400 electroencephalogram (EEG) component compared to unrelated items. Along the same lines, crowded symbols like sequences of geometric shapes and arrows can bias reaction times in following stimuli [4].

Ensemble Dynamic Objects

Finally, dynamic configural information can also survive crowding. When observers are asked to indicate the motion direction of a point-light-walker stimulus flanked by two additional biological motion stimuli, the reported direction reflects, to some extent, an integration of the three directions from the target and two flankers (Figure 3M) [26].

Beyond perception, crowded and indistinguishable stimuli can also influence our actions. For example, a three-dimensional disk crowded by flanking disks may be unrecognizable, but observers can still scale their grasp to the size of the target [41].

Importantly, it might be argued that the information passing though crowding seems so because of incomplete crowding or stimulus fluctuations. In order to ensure complete indiscriminability under crowding conditions, in all the mentioned studies observers were at chance-level in recognizing target features and objects. Even when considering only trials where observers did not correctly recognize the crowded target, visual information still biases subsequent objects [38]. In addition, in order to make sure that observers never looked directly at the stimuli (weakening crowding), some studies substituted the crowded stimulus with irrelevant content as soon as the observer's gaze diverged from the predefined fixation location [4,37].

Taken all together, these results show that crowded object information is not lost at early stages of visual processing, but it is largely preserved at low-level, mid-level, and high-level stages.

Multi-level Crowding in Scenes

Standard models assume that crowding occurs at a single stage, mostly between low-level features [1,2]. We have reviewed converging evidence that crowding can occur selectively at multiple stages of visual analysis, from low-level features (orientation, motion, position) to high-level representations (objects, letters, faces and biological motion stimuli; Figure 2). When looking back at the real scene in Figure 1, several kinds of crowding can be identified: crowding between lines, shapes, letters, faces, and human bodies.

If crowding can occur among features or objects at many stages of visual processing, what determines the appearance of crowding in a natural scene? The level at which crowding manifests in our experience depends on several factors, like the sorts of stimuli, the complexity of the scene, and the layers of information present. For example, even within a single object (such as the runners in Figure 1) there may be layers of crowding, ranging from feature crowding to configural crowding. In addition, task and attention play a crucial role; for example, it is easier to discriminate an object when it is defined by its category rather than by its form [42], or a number when it is defined by its relative magnitude rather than its exact digit [43]. Similarly, the degree to which attention is devoted to a crowded object can strongly determine crowding strength [44], whereas the role of awareness is still under debate [45]. Hence, crowding is not a unitary process which affects all features and objects in the same way, but different kinds of crowding arise depending on multiple factors including the goals of the observer: holistic or configural crowding occurs when we are looking for the identity of friend in a crowd of people, but also object-part-based crowding can occur when we look for a particular smile.

Finally, standard models of crowding assume a loss of information at early stages. However, the reviewed array of evidence shows that crowded information, although unrecognizable, can pass through crowding at multiple levels of visual analysis (Figure 3). Further evidence against this loss of information account comes from the fact that, depending on the global stimulus configuration or the number of flankers, low-level crowded information can be fully retrieved (see [12] for a review of recent results). Taken together, these findings show that crowding is not a low-level bottleneck that irretrievably dismantles, destroys, or filters low level information [12] or high-level object-specific information (Figure 3).

Implications for the Neural Locus of Crowding

Crowding occurs at several levels across the visual hierarchy, from low-level features like orientation and lines to holistic and dynamic object representations. This helps explain the seemingly conflicting findings that crowding may occur in early visual areas, like V1 [46,47], but also in later ones [48,49]. In fact, crowding may be a persistent outcome of visual processing at several stages along the visual hierarchy, from early to late visual areas. Thus, although there is a dearth of physiology experiments on crowding of high-level objects, such as bodies and faces, the psychophysical results reviewed here predict that, with appropriately controlled high-level stimuli and tasks, crowding should occur in object, face, and body selective cortical regions in addition to early visual areas.

Reconciling the Seeming Paradox and Implications for Crowding Models

The results reviewed here revealed a seeming paradox. Crowding occurs at multiple levels (Figure 2), and it impairs recognition of individual features or objects. Despite the inability to identify these crowded features and objects, these same features and objects can influence or be incorporated into subsequent visual percepts (Figure 3), indicating that the visual system does maintain high-fidelity representations of crowded features and objects. Reconciling the paradox is simply a matter of acknowledging that crowding is a problem of *access* and not representation: crowding occurs at multiple levels of analysis, but crowding does not dismantle or destroy object representations at any level. In accordance with this view, crowding was recently proposed as a valid tool to investigate non-conscious processing [4].

Pooling, substitution, grouping, and attention are powerful mechanisms in explaining low-level crowding, and there is independent evidence that each exists, but the architecture of these mechanisms needs to be rethought to account for crowding at different stages along the visual hierarchy and the fact that crowded information is not completely lost. Pooling models can be a valuable mechanism to explain low-level crowding between simple features, for example in terms of averaging [5], population processing [6], or summary statistics [7,8]. But, these models must be substantially revised to allow crowded information to pass through the pooling stage, and to allow pooling selectively between high-level representations. Substitution models [9,10] and grouping accounts [12] must be revised to accommodate multi-level and object-level crowding, in addition to allowing information at each stage to pass through crowding. Attentional models [11], which are based on a single attentional bottleneck, must implement multiple attentional bottlenecks at several levels of visual processing.

More generally, the results reviewed here provide a general framework within which new crowding and object recognition models can be evaluated and tested; we propose that the seeming paradox — crowding happens at multiple levels, but represented visual information at each level can influence subsequent visual processing — is an inescapable litmus test to define and test the validity of future models. In line with this framework, Chaney *et al.* [50] recently proposed a hierarchical sparse selection model, where crowding is not due to degraded visual representations in the brain, but to impoverished sampling or selection of those representations, which can happen at multiple levels. In line with this view, crowding occurs between stimuli that are represented at the same stage (and are thus functionally similar; see Figure 2, red boxes). When stimuli are represented at different stages (and are thus functionally dissimilar), the target will be isolated at its own stage and, hence, crowding will be weaker (Figure 2, green boxes). Importantly, object representations are preserved at each relevant processing stage and, hence, they can influence following stimuli (Figure 3). The same architecture could in principle account for grouping results [12,13], if grouping occurs at multiple levels of visual processing.

Conclusion

Taken together, the results we have reviewed reinforce the idea that there is more to crowding than early sensory interactions. We propose that future crowding models should aim at explaining: first, how crowding occurs at several levels of visual processing; and second, how crowded object information is preserved at low-level, mid-level, and high-level stages. Reconciling the seeming paradox that complex information can sometimes get through the bottleneck of crowding will be the new challenge for models of crowding and object recognition in general.

ACKNOWLEDGMENTS

The authors would like to thank Michael Herzog for his feedback on earlier drafts of this article.

REFERENCES

1. Levi, D.M. (2008). Crowding–an essential bottleneck for object recognition: a mini-review. Vision. Res. *48*, 635–654.
2. Pelli, D.G., and Tillman, K.A. (2008). The uncrowded window of object recognition. Nat. Neurosci. *11*, 1129–1135.
3. Whitney, D., and Levi, D.M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. Trends. Cogn. Sci. *15*, 160–168.
4. Faivre, N., Berthet, V., and Kouider, S. (2014). Sustained invisibility through crowding and continuous flash suppression: a comparative review. Front. Psychol. *5*, 475.
5. Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., and Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. Nat. Neurosci. *4*, 739–744.
6. Harrison, W.J., and Bex, P.J. (2015). A unifying model of orientation crowding in peripheral vision. Curr. Biol. *25*, 3213–3219.
7. Balas, B., Nakano, L., and Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. J. Vis. *9*, 1–18.
8. Freeman, J., and Simoncelli, E.P. (2011). Metamers of the ventral stream. Nat. Neurosci. *14*, 1195–1201.
9. Ester, E.F., Klee, D., and Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. J. Exp. Psychol. Hum. Percept. Perform. *40*, 1022–1033.
10. Huckauf, A., and Heller, D. (2002). What various kinds of errors tell us about lateral masking effects. Vis. Cogn. *9*, 889–910.
11. He, S., Cavanagh, P., and Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. Nature *383*, 334–337.
12. Herzog, M.H., Sayim, B., Chicherov, V., and Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. J. Vis. *15*, 1–18.
13. Manassi, M., Sayim, B., and Herzog, M.H. (2012). Grouping, pooling, and when bigger is better in visual crowding. J. Vis. *12*, 1–14.
14. Sayim, B., Westheimer, G., and Herzog, M.H. (2010). Gestalt factors modulate basic spatial vision. Psychol. Sci. *21*, 641–644.
15. Maus, G.W., Fischer, J., and Whitney, D. (2011). Perceived positions determine crowding. PLoS One *6*, e19796.
16. Dakin, S.C., Greenwood, J.A., Carlson, T.A., and Bex, P.J. (2011). Crowding is tuned for perceived (not physical) location. J.Vis. *11*, 1–13.
17. Martelli, M., Majaj, N.J., and Pelli, D.G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. J. Vis. *5*, 58–70.
18. Sun, H.M., and Balas, B. (2015). Face features and face configurations both contribute to visual crowding. Atten. Percept. Psychophys. *77*, 508–519.
19. Kimchi, R., and Pirkner, Y. (2015). Multiple level crowding: crowding at the object parts level and at the object configural level. Perception *44*, 1275–1292.
20. Huckauf, A., Heller, D., and Nazir, T.A. (1999). Lateral masking: limitations of the feature interaction account. Percept. Psychophys. *61*, 177–189.
21. Reuther, J., and Chakravarthi, R. (2014). Categorical membership modulates crowding: evidence from characters. J. Vis. *14*, 1–13.
22. Louie, E.G., Bressler, D.W., and Whitney, D. (2007). Holistic crowding: selective interference between configural representations of faces in crowded scenes. J. Vis. *7*, 1–11.
23. Kalpadakis-Smith, A., Goffaux, V., and Greenwood, J.A. (2017). Crowding is not holistic: Visual similarity determines the effect of clutter on faces. PsyArXiv. https://doi.org/10.17605/OSF.IO/2Z8UN.
24. Kanwisher, N., and Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. Philos. Trans. R. Soc. Lond. B. *361*, 2109–2128.
25. Farzin, F., Rivera, S.M., and Whitney, D. (2009). Holistic crowding of Mooney faces. J. Vis. *9*, 1–15.
26. Ikeda, H., Watanabe, K., and Cavanagh, P. (2013). Crowding of biological motion stimuli. J. Vis. *13*, 1–6.
27. Ikeda, H., and Watanabe, K. (2016). Action congruency influences crowding when discriminating biological motion direction. Perception *45*, 1046–1059.
28. Rajimehr, R., Montaser-Kouhsari, L., and Afraz, S.R. (2003). Orientation-selective adaptation to crowded illusory lines. Perception *32*, 1199–1210.
29. Faivre, N., and Kouider, S. (2011). Increased sensory evidence reverses nonconscious priming during crowding. J. Vis. *11*, 1–13.

30. Aghdaee, S.M. (2005). Adaptation to spiral motion in crowding condition. Perception *34*, 155–162.

31. Rajimehr, R., Vaziri-Pashkam, M., Afraz, S.R., and Esteky, H. (2004). Adaptation to apparent motion in crowding condition. Vis. Res. *44*, 925–931.

32. Whitney, D. (2005). Motion distorts perceived position without awareness of motion. Curr. Biol. *15*, R324–326.

33. Lau, J.S., and Cheung, S.H. (2012). Illusory contour formation survives crowding. J. Vis. *12*, 1–12.

34. Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., and Thorpe, S.J. (2016). Finding faces, animals, and vehicles in far peripheral vision. J. Vis. *16*, 1–13.

35. Bayle, D.J., Schoendorff, B., Henaff, M.A., and Krolak-Salmon, P. (2011). Emotional facial expression detection in the peripheral visual field. PLoS One *6*, e21584.

36. McKone, E. (2004). Isolating the special component of face recognition: peripheral identification and a Mooney face. J. Exp. Psychol. Learn. Mem. Cogn. *30*, 181–197.

37. Kouider, S., Berthet, V., and Faivre, N. (2011). Preference is biased by crowded facial expressions. Psychol. Sci. *22*, 184–189.

38. Fischer, J., and Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. J. Neurophysiol. *106*, 1389–1398.

39. Yeh, S.L., He, S., and Cavanagh, P. (2012). Semantic priming from crowded words. Psychol. Sci. *23*, 608–616.

40. Zhou, J., Lee, C.-L., and Yeh, S.-L. (2016). Word meanings survive visual crowding: evidence from ERPs. Lang. Cogn. Neurosci. *31*, 1167–1177.

41. Chen, J., Sperandio, I., and Goodale, M.A. (2015). Differences in the effects of crowding on size perception and grip scaling in densely cluttered 3-D scenes. Psychol. Sci. *26*, 58–69.

42. Huckauf, A. (2007). Task set determines the amount of crowding. Psychol. Res. *71*, 646–652.

43. Huckauf, A., Knops, A., Nuerk, H.C., and Willmes, K. (2008). Semantic processing of crowded stimuli? Psychol. Res. *72*, 648–656.

44. Yeshurun, Y., and Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. J. Vis. *10*, 1–12.

45. Shin, K., Chung, S.T.L., and Tjan, B.S. (2017). Crowding, visual awareness, and their respective neural loci. J. Vis. *17*, 1–18.

46. Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., and Fang, F. (2014). Attention-dependent early cortical suppression contributes to crowding. J. Neurosci. *34*, 10465–10474.

47. Millin, R., Arman, A.C., Chung, S.T., and Tjan, B.S. (2014). Visual crowding in V1. Cereb. Cortex *24*, 3107–3115.

48. Anderson, E.J., Dakin, S.C., Schwarzkopf, D.S., Rees, G., and Greenwood, J.A. (2012). The neural correlates of crowding-induced changes in appearance. Curr. Biol. *22*, 1199–1206.

49. Freeman, J., Donner, T.H., and Heeger, D.J. (2011). Inter-area correlations in the ventral visual pathway reflect feature integration. J. Vis. *11*, 1–23.

50. Chaney, W., Fischer, J., and Whitney, D. (2014). The hierarchical sparse selection model of visual crowding. Front. Integr. Neurosci. *8*, 73.