Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Uncorking the bottleneck of crowding: a fresh look at object recognition

Michael H Herzog and Mauro Manassi

In crowding, the perception of a target deteriorates in the presence of clutter. Crowding is usually explained within the framework of object recognition, where processing proceeds in a hierarchical *and* feedforward fashion from the analysis of low level features, such as lines and edges, to high level features, such shapes and objects. Here, reviewing work of the last two years, we will show evidence that these models fail to explain a large body of findings, which undermine the philosophy of this approach as such. We propose that the configuration of more or less all elements across the entire visual field determines crowding. Wholes, such as objects and shapes, determine performance on their constituting elements. Perceptual grouping and Gestalt, neglected for a long time, are key to understand crowding and object recognition in general.

Addresses
Laboratory of Psychophysics, Brain Mind Institute, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

Corresponding author: Herzog, Michael H. (michael.herzog@epfl.ch)

Current Opinion in Behavioral Sciences 2015, **1**:86–93

This review comes from a themed issue on **Cognitive neuroscience**

Edited by **Cindy Lustig** and **Howard Eichenbaum**

http://dx.doi.org/10.1016/j.cobeha.2014.10.006

Introduction

At the core of most vision research is implicitly or explicitly a hierarchical and feedforward model, in which visual processing proceeds from the analysis of basic features to more and more complex ones (e.g. [1••]). Neurons in the primary visual cortex V1 'extract' edges and lines from the visual images (Figure 1A). Neurons in V2 pool information from V1 neurons coding for more complex features, such illusory contours. This encoding principle proceeds along the visual hierarchy. A hypothetical square neuron is 'created' by projections from neurons coding for its constituting horizontal and vertical lines (Figure 1A).

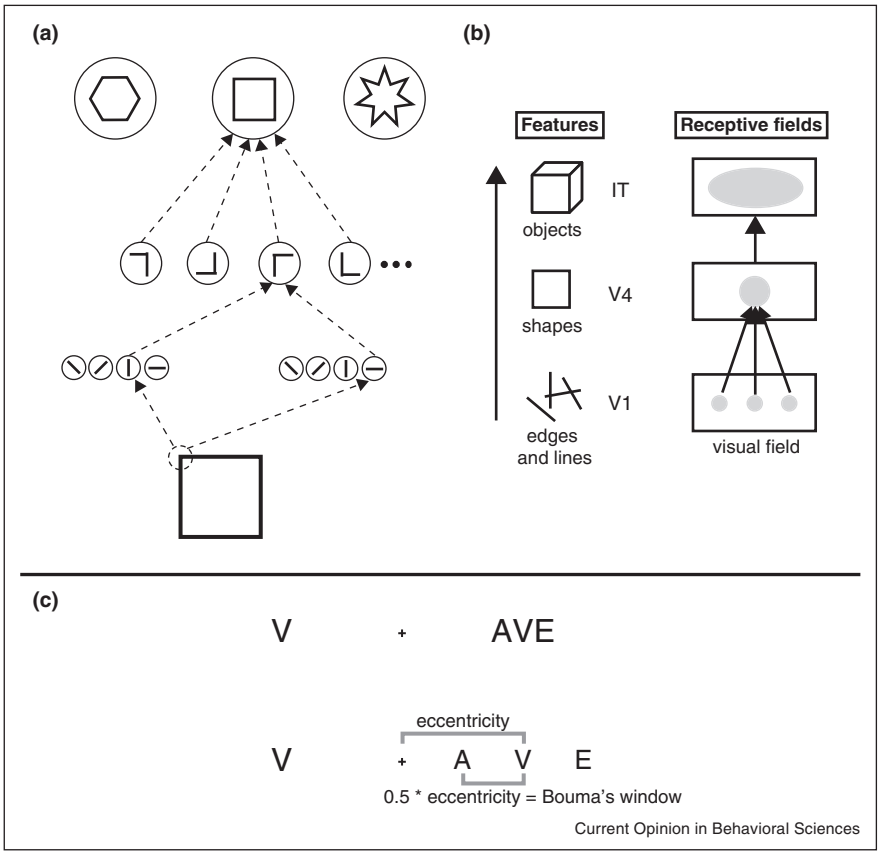
There are three important characteristics. First, processing proceeds from low (lines, edges) to complex (objects, faces) features. As a consequence, if information is lost at the early stages, it is irretrievably lost. In addition, processing at each level is fully determined by processing at the previous level. Second, processing is stereotypical in the sense, that neurons act like filters, which analyse the visual scene in always the same way, that is, independent of the higher level features (Figure 1B). Low determines high level processing and not the other way around. The beauty and main goal of these models is to replace subjective terms, such as grouping and good Gestalt, by a truly mechanistic processing. Third, receptive fields increase along the visual hierarchy because pooling is necessary for object recognition in the sense that a 'square neuron' needs to integrate over larger parts of the visual scene than neurons coding for its constituting lines. For this reason, object recognition becomes difficult when objects are embedded in clutter because object irrelevant elements mingle with relevant ones. This is exactly what crowding is about.

You can experience crowding for yourself in Figure 1C. When fixating the central cross, it is easy to recognize the single letter V on the left. However, when the V is flanked by other letters, identification is much more difficult (right). Observers perceive the target letter distorted and jumbled with the flanking letters. For this reason, crowding is often seen as a bottleneck or breakdown of object recognition [2••,3].

Because crowding is thought to reflect the above characteristics, crowding is a perfect paradigm to study object recognition. For example, flankers always deteriorate performance because pooling more elements leads to an increase in noise. Bouma [4] showed that when a target is presented at eccentricity e, flankers interfere only when presented within a critical window of the size of $0.5 \times e$ (Bouma's law; Figure 1C). Bouma's law is explained because pooling, particularly for low level features, occurs only within a restricted region [5,6]. Current models propose that features are not simply pooled but merged in textural representations by summary statistics [7,8,9•]. Interactions in Bouma's window are usually thought to be mainly mediated by low-level features because crowding is strong if target and flankers have the same color, and much reduced for different colors [10,11••], in line with current EEG and fMRI studies showing feature-specific suppression in the early visual areas [12–14].

In the following, we will show that crowding strength can weaken if more flankers are presented, crowding occurs

Figure 1



(A) According to hierarchical, feedforward models of object recognition (e.g. [1••]), stimulus processing starts with the analysis of very simple features and proceeds to more and more complex visual representations. A hypothetical 'square' neuron receives input from neurons tuned to angles, which in turn receive inputs from line detectors. Along the hierarchy, processing at each level is fully determined by processing at the previous level. **(B)** Neurons in V1 are sensitive to simple features, such as edges and lines. In higher visual areas, neurons are sensitive to more and more complex features, such as shapes (V4) and objects (IT). Receptive field sizes increase from lower to higher visual areas. **(C)** Crowding. When fixating the central cross, it is easy to recognize the letter V on the left but difficult on the right because of the flanking letters. Crowding is usually thought to occur only for flankers presented within a window of about half the eccentricity of target presentation (Bouma's law). When flankers are placed outside Bouma's window, letter recognition is not compromised.

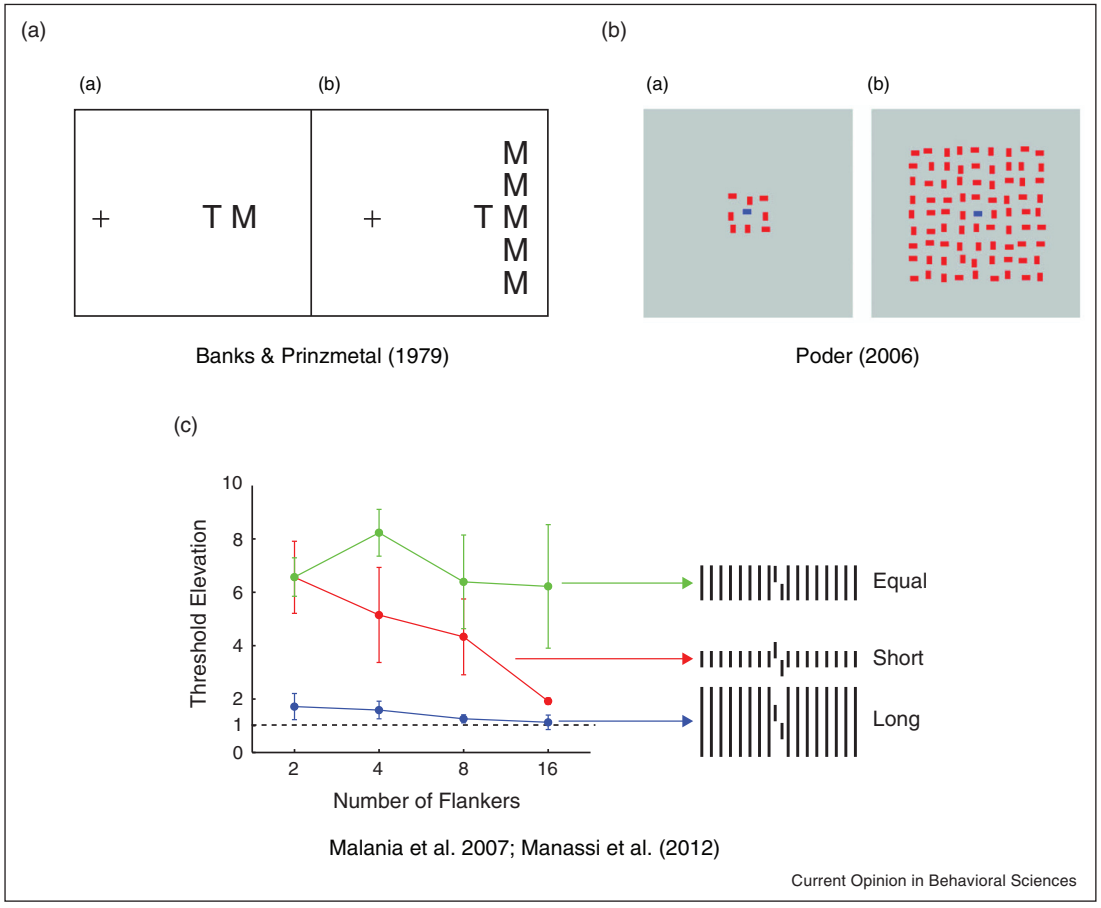
with flankers well beyond Bouma's window, complex features determine low level feature processing, processing is not stereotypically but necessitates a grouping stage, and, finally, information is not lost at early stages. We can uncork the bottleneck of vision simply by adding elements.

A fresh look on crowding and object recognition

More can be better

First, according to pooling models, crowding strength increases if the number of flankers increases because more irrelevant information is pooled. For this reason, almost all experiments on crowding have used only single flankers neighboring the target [37,38]. However, already in 1979, Banks and colleagues showed that crowding is weaker when a target letter is flanked by an array of flanking letters compared to a single letter (Figure 2A, [39]). These results were forgotten for more than 25 years. Recently, we have shown *when* bigger is better (Figure 2B). We presented a vernier stimulus, which consists of two vertical lines slightly offset either to the left or right. Observers indicated the offset direction. When one shorter line to the left and one to the right flanked the vernier, performance strongly deteriorated. Performance improved when further lines were added (Figure 2B, red line). The same pattern of results was found for longer lines (Figure 2B, blue line) but not for lines with the same length as the vernier (Figure 2B, green line). In this case, performance stays roughly on the same level independent of the number of lines. Hence, bigger can be worse and bigger can be better [11••,15,16,41]. The latter case clearly shows that vernier information is not irretrievably lost at the early stages. By

Figure 2



More can be better. **(A)** In a letter identification task, crowding was stronger when one flanking letter was presented next to the target letter (a) compared to a condition with five flanking letters (b). **(B)** Observers indicated the tilt of a blue bar. Performance was much better when many red lines flanked the bar than for fewer lines. **(C)** We presented a vernier, two vertical lines with a small horizontal offset, and asked observers to discriminated its offset. We determined the smallest offset size for which observers reached 75% correct responses (threshold). The dashed line indicates the unflanked vernier threshold. When the vernier was flanked by two lines, thresholds increased. When the number of short and long lines increased, thresholds decreased (red and blue lines). For equal length, thresholds remained on the same level (green line). We propose that crowding decreases the more shorter or longer lines are presented because arrays of short or long flankers ungroup from the vernier much stronger than arrays of flankers.

adding further elements, we can 'uncork' the bottleneck of vision, that is, we can undo crowding.

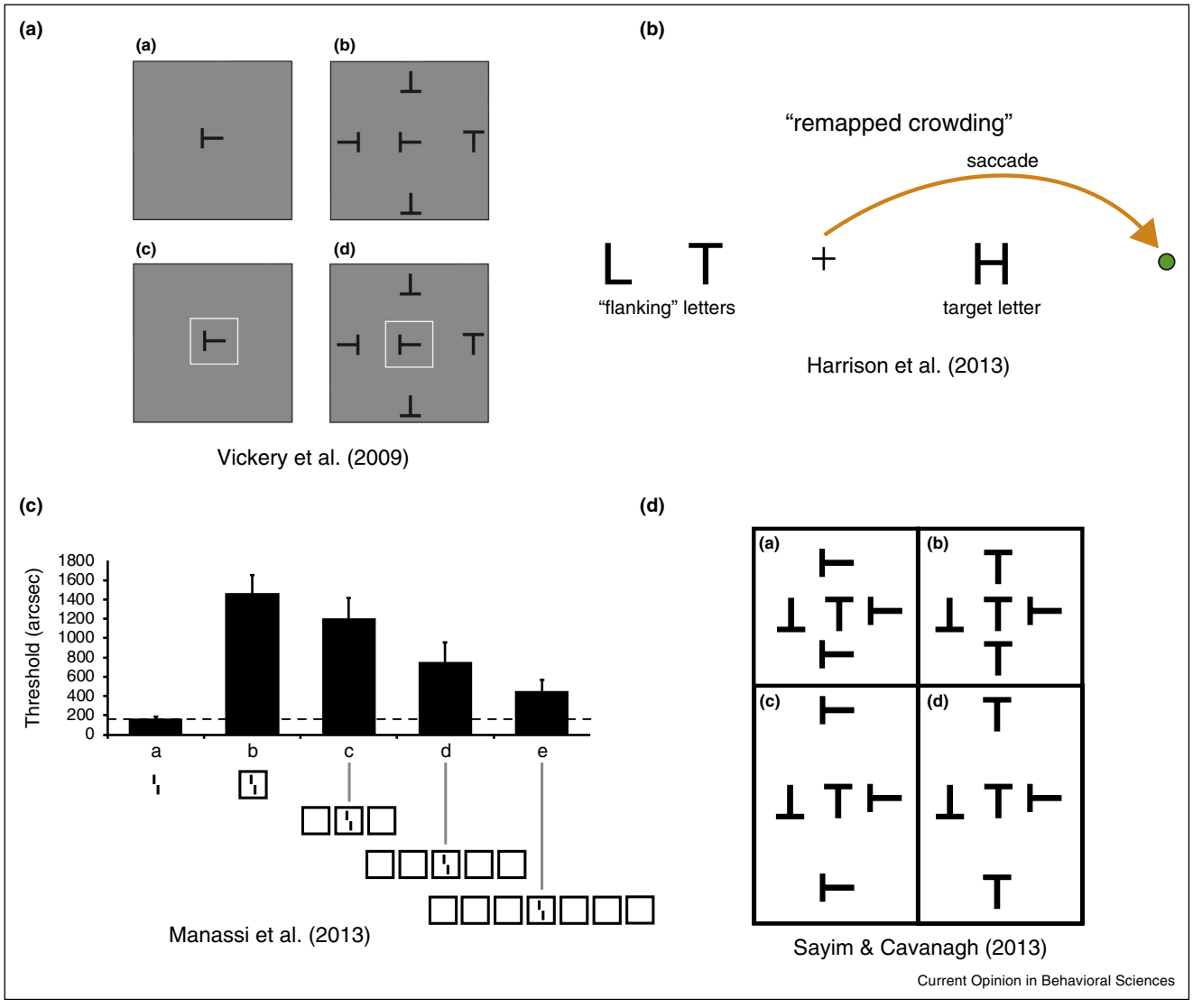
We proposed that grouping explains these results. When single shorter lines are presented they group with the vernier. However, arrays of shorter lines group with each other and do not group with the vernier. For equal length lines, the vernier always groups with the flankers. Hence, crowding is weak when target and flankers do not group with each other. Strong crowding occurs only when target and flankers group.

It may be argued that, for example, adding lines in Figure 2C, [40] simplifies the Fourier spectrum, that is, 'the more the better' argument does not apply. We could not find any evidence that such an approach can succeed [43].

Elements outside Bouma's window can dramatically decrease or increase crowding

Second, because crowding was thought to occur only by flankers presented within Bouma's window, flankers were only presented close to the target. However, crowding extends well beyond Bouma's window. Orientation discrimination of a letter T only slightly deteriorated when flanking Ts were presented outside Bouma's window (Figure 3A, a–b). Crowding was also weak when a square within Bouma's window surrounded the target (Figure 3A, a-c). However, the combination of the two conditions led to 'super-crowding' (Figure 3A-d; [17]). Interestingly, even flankers in the opposite hemifield can deteriorate target perception when an upcoming saccade will place them next to the target (Figure 3B, [18••]). Elements outside Bouma's window can surprisingly even

Figure 3



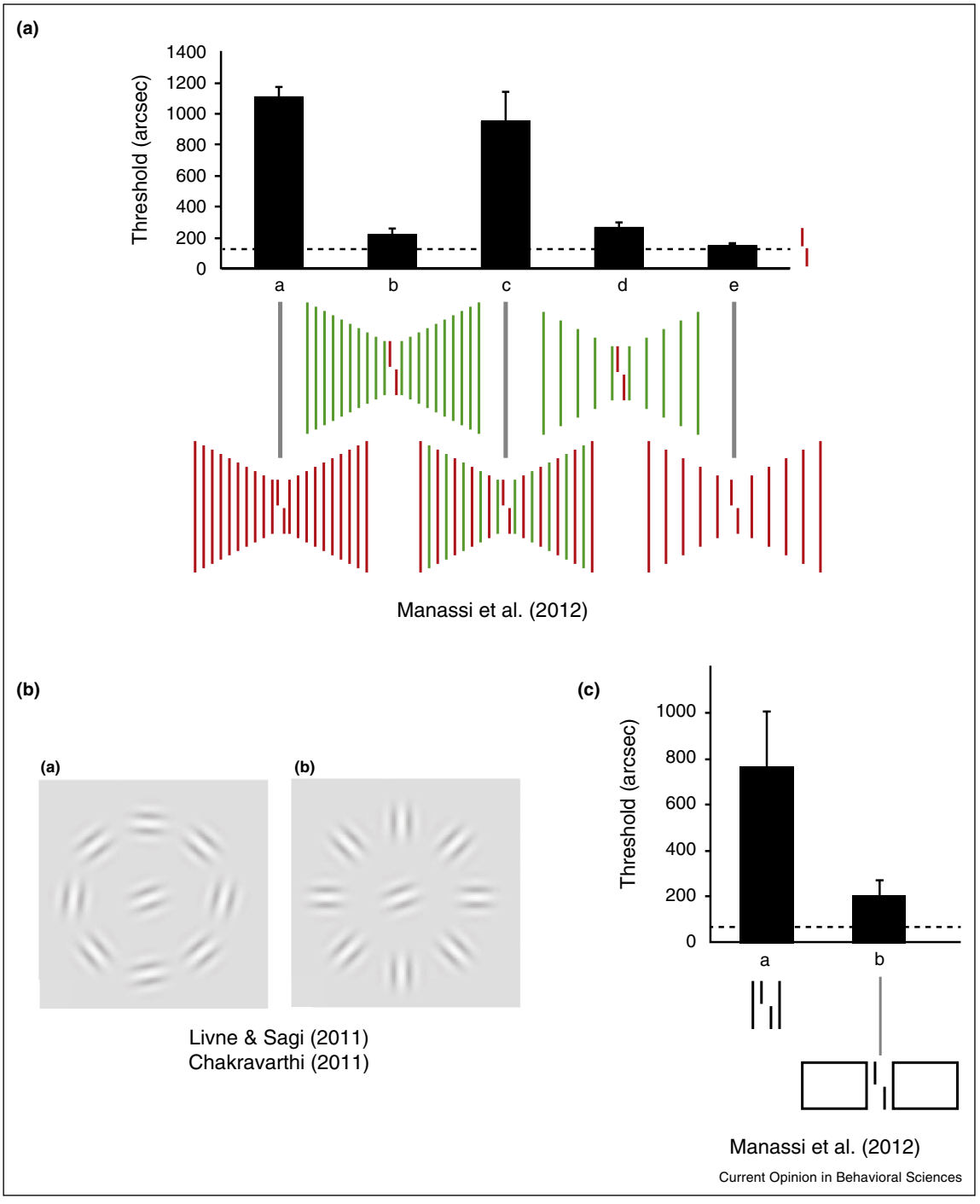
Elements outside Bouma's window can increase or decrease crowding. **(A)** Observers indicated the orientation of a T (a). Crowding is weak when the target was flanked by Ts outside Bouma's window (b) or surrounded by a square (c). However, when combining these two conditions, crowding was strong (d). **(B)** Observers fixated a cross in the middle of the screen. Observers were then asked to make an eye movement (saccade) to the green point on the right. During the saccade two letters were presented briefly in the left visual field and one in the right field. Observers were asked to identify the right letter. Surprisingly, even though the three letters were never presented close to each other, strong crowding occurred. **(C)** A vernier was presented at 9° of eccentricity. Vernier offset discrimination deteriorated when the vernier was surrounded by the outline of a square (a–b). Performance continuously improved when we added further squares (c–e). For three squares on each side of the vernier, crowding almost disappeared (e). These seven squares range from 0.5° to 17.5° of eccentricity whereas Bouma's window ranges only from 4.5° to 13.5°. **(D)** In a T orientation discrimination task, crowding was strong when Ts with random orientation were placed within Bouma's window (a). When Ts along the vertical meridian had the same orientation, performance improved (b). There was no crowding when the randomly oriented Ts were placed outside Bouma's window (c). However, when the flankers had the same orientation as the target, performance improved (d).

decrease crowding strength. We presented a vernier as a target. Performance strongly decreased when the vernier was surrounded by a square. This is a classic crowding effect. Surprisingly, performance improved when more and more squares were added, extending beyond Bouma's window (Figure 3C; [19•], see also [20] in Figure 3D and [21]).

Global configuration rather than low level local interactions determine crowding

Third, because crowding was thought to be specific for low level features, crowding was studied mainly with targets and flankers having, for example, the same orientation or color. However, low level feature similarity is very little predictive for crowding. In Figure 4, we show

Figure 4



Global configuration determines crowding. **(A)** *Global pattern.* A red vernier amongst green lines leads to less crowding than amongst red flankers (a–b). Crowding is strong for a grating with alternating red-green flankers (c) even though crowding is weak for the green and red gratings making up the alternating grating (d,e). **(B)** *Grouping by contour.* In a Gabor orientation discrimination task, crowding was weaker when the flankers were arranged in a smooth contour (a) compared to when they were perpendicular (b). **(C)** *Good Gestalt.* As before, vernier offset discrimination deteriorated when flanked by single lines (a). Crowding strongly reduced when the lines became part of a rectangle even though the lines are at the very same position as in the condition before (b). As we argue, the lines 'loose' their crowding power because they are part of a good Gestalt, which ungroups from the vernier.

how 'global' and figural aspects determine crowding [11••,22–24]. As a first example: in accordance with previous results and models, performance strongly deteriorated when a red vernier was flanked by red lines (Figure 4A,a). There was only little deterioration for green flankers (Figure 4A,b). However, when flankers alternated in color, performance was as much deteriorated as with the red flankers (Figure 4A,c). This effect cannot be explained by the red lines in the alternating pattern because, when presented alone, they led to very little crowding, and so did the green lines (Figure 4A,d–e). Hence, when crowding is probed with simple feature differences, indeed, it appears to be that crowding is specific to low-level features. However, using slightly more complex features disproves this thinking. Second example: observers discriminated the tilt of a Gabor patch surrounded by flanking Gabors of various orientations. When these Gabors made up a smooth contour, crowding was much weaker than when the very same Gabors were making up a star like pattern. Hence, it is the overall configuration of the flankers, which matters (Figure 4B, [42]). The third example shows how good Gestalt determines crowding. Performance strongly deteriorated when a vernier was flanked by two lines, well in accordance with previous findings. However, when rectangles were flanking the vernier, crowding was weak, even though the same flanking lines from the previous condition were at the very same positions (Figure 4C, [11]). Hence, crowding is not restricted to low level features interactions. Surprisingly, even high level features such as good Gestalt (rectangles) trump low level ones (simple lines). Particularly, these results are hard to explain with hierarchical, feedforward models. When the vernier is processed at early stages and there are no feedback connections how can then high level features, such as the shape of the rectangles, determine vernier processing? It seems that we need to give up either the feedforward or the hierarchy assumption.

It seems the best predictor for crowding is grouping between the target and flankers. Crowding is strong in the alternating pattern because the vernier fits in the overall configuration very well and thus groups with all elements. Because the flanking Gabors make up a smooth contour, the central Gabor does not group with the flankers. In the last example, crowding is weak when the very same lines become part of a good Gestalt and thus ungroup from the vernier. These results are in line with physiological evidence that crowding occurs in late rather than early visual processing [25•,26], reflecting recurrent processing related to the global spatial layout of the entire stimulus configuration.

Particularly, the results on stimulus configuration have strong philosophical implications for object recognition in general. The philosophy of hierarchical, feedforward models is that the complex problem of vision can be broken down into a cascade of simple and independent processing stages. Analysis starts with basic feature detection by stereotyped filtering (Figure 1B). For example, a vertical line presented alone is processed in the same way as when embedded into context. Only later stages will take contextual information into account by pooling. As a square is nothing else as four lines, encoding of a square is nothing else combining the outputs of line detectors. Such models are aimed to eliminate and thus explain the inherent subjective aspects of perception. Such models are highly desirable from a mathematical point of view avoiding, for example, the use of analytically insolvable differential equations, which easily come into play when processing is recurrent. However, the crowding results of the last years show that visual processing is more complex. It seems that a grouping stage cannot be avoided. First, we need to know how elements group before we know which elements interfere with each. This grouping is flexible in the sense that small changes in the configuration, invisible to low level features analysis, can lead to strong changes in crowding strength. Hence, high level determines low level processing as much as the other way around.

Applications

Understanding crowding is not only crucial for basic vision research but also for many other fields where crowding is used as tool or in clinical research. A better understanding of crowding is, for example, important for amblyopia [27], dyslexia [28,29] and aging [30•,31].

Crowding is often used to render a target invisible in consciousness research. Many studies rely on the above hierarchical, feedfoward models assuming that crowding is a low-level bottleneck and thus crowding can be used to study which features are filtered out at the early stage of vision and which features are passed on for conscious perception. Unconscious processing of orientation [32], objects [33] and facial expressions [34,35] were shown to pass through the bottleneck of crowding, placing its cortical mechanism higher and higher along the visual hierarchy.

However, as we have shown, crowding is not a bottleneck of low level vision. Quite to the contrary, crowding strength depends on the overall stimulus configuration and, hence, high level processing. In addition, we can render a target easily and flexibly visible by adding elements [11••,15].

Summary and outlook

Crowding is usually explained by hierarchical, feedforward processing, where (1) more flankers always deteriorate performance, (2) only nearby elements interfere with a target (Bouma's window), and (3) interference occurs mainly within feature specific 'channels'. These characteristics have shaped crowding research for the last

40 years. However, research of the last years has shown that none of these characteristics is met in crowding. More can be better. Elements far outside Bouma's window can strongly in- or decrease crowding. Crowding strength seems to depend on *all* elements in the *entire* visual field and, on top of it, on the *overall* configuration of the elements. Moreover, crowding is not an inevitable bottleneck. Adding elements can 'uncork' the bottle. Clearly local, hierarchical approaches fail to explain these results. The same holds true for object recognition in general. Subtle changes, wherever in the visual field, can strongly change object recognition. 'Basic' vernier acuity and Gabor detection cannot be explained by local models. It seems we cannot break down visual processing into small retinotopic, independent processing units and, when we have understood their exact characteristics, put them into a hierarchical, feedforward framework.

It seems we are back to the days of the Gestaltists with all its issues. For example, grouping is not a mechanism to explain why crowding occurs. Why is there suppression or feature jumbling? It may be that, for example, pooling operates within groups rather than within Bouma's window. However, why should the human brain give up good resolution in certain conditions (with two flankers) but not in others (with many flankers)? We think that large scale, recurrent and, particularly, normative models are crucial to answer these questions [36]. These topics are not only crucial for basic research but are also of clinical research and for all of us. For example, it is not the right spacing but the right grouping that speeds up or slows down reading of this article.

Conflict of interest statement

Nothing declared.

Acknowledgement

This work was supported by the Swiss National Science Foundation (SNF) Project 'Basics of visual processing: what crowds in crowding?'.

References

1. •• DiCarlo JJ, Zoccolan D, Rust NC: **How does the brain solve visual object recognition?** *Neuron* 2012, **73**:415-434.
This review proposed that object recognition occurs through a cascade of reflexive, largely feedforward computations along the ventral visual pathway.

2. •• Strasburger H, Rentschler I, Jüttner M: **Peripheral vision and pattern recognition: a review**. *J Vis* 2011, **11**.
An extensive review on peripheral vision.

3. Whitney D, Levi DM: **Visual crowding: a fundamental limit on conscious perception and object recognition**. *Trends Cogn Sci* 2011, **15**:160-168.

4. Bouma H: **Interaction effects in parafoveal letter recognition**. *Nature* 1970, **226**:177-178.

5. Pelli DG, Palomares M, Majaj NJ: **Crowding is unlike ordinary masking: distinguishing feature integration from detection**. *J Vis* 2004, **4**:1136-1169 12.

6. Zahabi S, Arguin M: **A crowdful of letters: disentangling the role of similarity, eccentricity and spatial frequencies in letter crowding**. *Vis Res* 2014, **97**:45-51.

7. Balas B, Nakano L, Rosenholtz R: **A summary-statistic representation in peripheral vision explains visual crowding**. *J Vis* 2009, **9**.

8. Freeman J, Simoncelli EP: **Metamers of the ventral stream**. *Nat Neurosci* 2011, **14**:1195-1201.

9. • Nandy AS, Tjan BS: **Saccade-confounded image statistics explain visual crowding**. *Nat Neurosci* 2012, **15**:463-469.
This study developed a model for crowding based on saccade-confounded image statistics.

10. Kooi FL, Toet A, Tripathy SP, Levi DM: **The effect of similarity and duration on spatial interaction in peripheral vision**. *Spat Vis* 1994, **8**:255-279.

11. •• Manassi M, Sayim B, Herzog MH: **Grouping, pooling, and when bigger is better in visual crowding**. *J Vis* 2012, **12**:1-14.
Pooling and substitution models are inadequate to predict crowding strength. Instead, perceptual grouping between target and flankers was shown to be a better predictor.

12. Millin R, Arman AC, Chung ST, Tjan BS: **Visual crowding in v1**. *Cereb Cortex* 2013.

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

14. Kwon M, Bao P, Millin R, Tjan BS: **Radial-tangential anisotropy of crowding in the early visual areas**. *J Neurophysiol* 2014 http://dx.doi.org/10.1152/jn.00476.2014.

15. Levi DM, Carney T: **Crowding in peripheral vision: why bigger is better**. *Curr Biol* 2009, **19**:1988-1993.

16. Chanceaux M, Grainger J: **Constraints on letter-in-string identification in peripheral vision: effects of number of flankers and deployment of attention**. *Front Psychol* 2013, **4** 119–119.

17. Vickery TJ, Shim WM, Chakravarthi R, Jiang YV, Luedeman R: **Supercrowding: weakly masking a target expands the range of crowding**. *J Vis* 2009, **9**:1-15 12.

18. •• Harrison WJ, Retell JD, Remington RW, Mattingley JB: **Visual crowding at a distance during predictive remapping**. *Curr Biol* 2013, **23**:793-798.
This study showed evidence for remapped crowding, with target and flankers in separate hemifields.

19. • Manassi M, Sayim B, Herzog MH: **When crowding of crowding leads to uncrowding**. *J Vis* 2013, **13**:1-10.
Contrary to hierarchical feeforward models of object recognition, figural processing can determine low-level processing in crowding.

20. Sayim B, Cavanagh P: **Grouping and crowding affect target appearance over different spatial scales**. *PLOS ONE* 2013, **8**.

21. Harrison WJ, Bex PJ: **Integrating retinotopic features in spatiotopic coordinates**. *J Neurosci* 2014, **34**:7351-7360.

22. Livne T, Sagi D: **How do flankers' relations affect crowding?** *J Vis* 2010, **10**:1-14 1.

23. Chakravarthi R, Pelli DG: **The same binding in contour integration and crowding**. *J Vis* 2011, **11**:1-12 10.

24. Saarela TP, Westheimer G, Herzog MH: **The effect of spacing regularity on visual crowding**. *J Vis* 2010, **10**:1-7 17.

25. •• Anderson EJ, Dakin SC, Schwarzkopf DS, Rees G, Greenwood JA: **The neural correlates of crowding-induced changes in appearance**. *Curr Biol* 2012, **22**:1199-1206.
This fMRI study shows that crowding is a multistage process, involving early and late cortical visual areas.

26. Chicherov V, Plomp G, Herzog MH: **Neural correlates of visual crowding**. *Neuroimage* 2014, **93(Pt 1)**:23-31.

27. Greenwood JA, Tailor VK, Sloper JJ, Simmers AJ, Bex PJ, Dakin SC: **Visual acuity, crowding, and stereo-vision are linked**

in children with and without amblyopia. *Invest Ophthalmol Vis Sci* 2012, **53**:7655-7665.

28. Moll K, Jones M: **Naming fluency in dyslexic and nondyslexic readers: differential effects of visual crowding in foveal parafoveal and peripheral vision**. *Q J Exp Psychol (Hove)* 2013, **66**:2085-2091.

29. Callens M, Whitney C, Tops W, Brysbaert M: **No deficiency in left-to-right processing of words in dyslexia but evidence for enhanced visual crowding**. *Q J Exp Psychol (Hove)* 2013, **66**:1803-1817.

30. • Scialfa CT, Cordazzo S, Bubric K, Lyon J: **Aging and visual crowding**. *J Gerontol B Psychol Sci Soc Sci* 2013, **68**:522-528.
This study showed that there is no difference in crowding strength with aging.

31. McGowan VA, White SJ, Jordan TR, Paterson KB: **Aging and the use of interword spaces during reading: evidence from eye movements**. *Psychon Bull Rev* 2014, **21**:740-747.

32. He S, Cavanagh P, Intriligator J: **Attentional resolution and the locus of visual awareness**. *Nature* 1996, **383**:334-337.

33. Faivre N, Kouider S: **Multi-feature objects elicit nonconscious priming despite crowding**. *J Vis* 2011, **11**.

34. Kouider S, Berthet V, Faivre N: **Preference is biased by crowded facial expressions**. *Psychol Sci* 2011, **22**:184-189.

35. Fischer J, Whitney D: **Object-level visual information gets through the bottleneck of crowding**. *J Neurophysiol* 2011, **106**:1389-1398.

36. Foley NC, Grossberg S, Mingolla E: **Neural dynamics of object-based multifocal visual spatial attention and priming: object cueing, useful-field-of-view, and crowding**. *Cogn Psychol* 2012, **65**:77-117.

37. • Ester EF, Klee D, Awh E: **Visual crowding cannot be wholly explained by feature pooling**. *J Exp Psychol Hum Percept Perform* 2014, **40**:1022-1033.
This study showed that crowding cannot be fully explained by compulsory pooling, but also a mechanism of probabilistic substitution must be taken into account.

38. • Freeman J, Chakravarthi R, Pelli DG: **Substitution and pooling in crowding**. *Atten Percept Psychophys* 2012, **74**:379-396.
Simple substitution models cannot account for crowding.

39. Banks WP, Larson DW, Prinzmetal W: **Asymmetry of visual interference**. *Percept Psychophys* 1979, **25**:447-456.

40. Põder E: **Crowding, feature integration, and two kinds of "attention"**. *J Vis* 2006, **6**:163-169.

41. Malania M, Herzog MH, Westheimer G: **Grouping of contextual elements that affect vernier thresholds**. *J Vis* 2007, **7**:1-7.

42. Livne T, Sagi D: **Configuration influence on crowding**. *J Vis* 2007, **7**:1-12 1.

43. Clarke AM, Herzog MH, Francis G: **Visual crowding illustrates the inadequacy of local versus global and feedforward versus feedback distinctions in modelling visual perception**. *Frontiers in Psychology* 2014, **5(1193)** http://dx.doi.org/10.3389/fpsyg.2014.01193 http://www.frontiersin.org/perception_science/10.3389/fpsyg.2014.01193/abstract.