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**

For a complete overview see the Issue and the Editorial

Available online 25th October 2014

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).

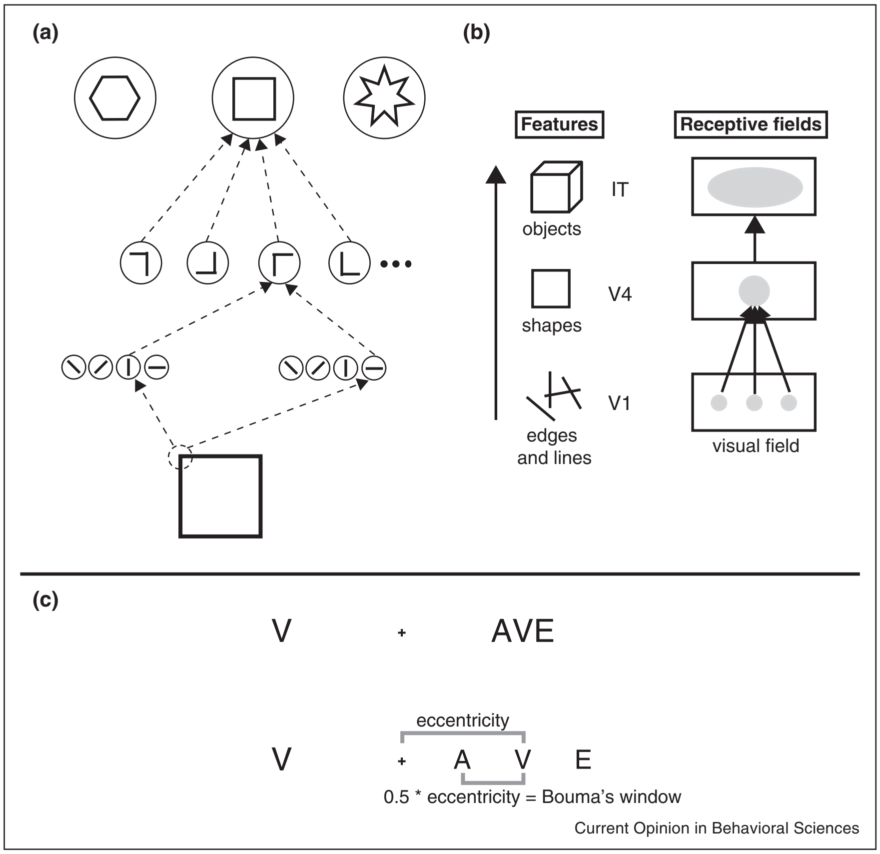
There are three important characteristics. First, processing proceeds from low (lines, edges) to complex (objects, faces) features. As a consequence, if information is lost at the early stages, it is irretrievably lost. In addition, processing at each level is fully determined by processing at the previous 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