frontiers in
COMPUTATIONAL NEUROSCIENCE

MINI REVIEW ARTICLE
published: 22 October 2014
doi: 10.3389/fncom.2014.00135

Why vision is not both hierarchical *and* feedforward

Michael H. Herzog* and Aaron M. Clarke

Laboratory of Psychophysics, Brain, Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Edited by:
Antonio J. Rodríguez-Sánchez, University of Innsbruck, Austria

Reviewed by:
Michael Zillich, Vienna University of Technology, Austria
Norbert Küger, The Maersk Mc-Kinney Moller Institute, Denmark

****Correspondence:***
Michael H. Herzog, Laboratory of Psychophysics, Brain, Mind Institute, École Polytechnique Fédérale de Lausanne, EPFL SV BMI LPSY SV-2807, Station 19, CH-1015 Lausanne, Switzerland
e-mail: michael.herzog@epfl.ch

In classical models of object recognition, first, basic features (e.g., edges and lines) are analyzed by independent filters that mimic the receptive field profiles of V1 neurons. In a feedforward fashion, the outputs of these filters are fed to filters at the next processing stage, pooling information across several filters from the previous level, and so forth at subsequent processing stages. Low-level processing determines high-level processing. Information lost on lower stages is irretrievably lost. Models of this type have proven to be very successful in many fields of vision, but have failed to explain object recognition in general. Here, we present experiments that, first, show that, similar to demonstrations from the Gestaltists, figural aspects determine low-level processing (as much as the other way around). Second, performance on a single element depends on all the other elements in the visual scene. Small changes in the overall configuration can lead to large changes in performance. Third, grouping of elements is key. Only if we know how elements group across the entire visual field, can we determine performance on individual elements, i.e., challenging the classical stereotypical filtering approach, which is at the very heart of most vision models.

Keywords: feedback, object recognition, crowding, Verniers, Gestalt

Object recognition traditionally proceeds from the analysis of simple to complex features. The Gestaltists proposed a number of basic rules, such as spatial proximity and good continuation, that underlie the grouping of elements into objects. Whereas the Gestalt rules work well for very basic stimuli, they grossly fail for slightly more complex stimuli. For this reason, research on Gestalt principles almost disappeared after the 1930's. After world war II, the discovery of the receptive field advanced vision science by revealing fundamental principles of retinal and cortical processing, which has led to a core scenario that is often, explicitly or implicitly, behind most models in visual neuroscience and the psychology of perception, and provides the basis for most models in computer vision.

The model is characterized by its hierarchical and feedforward organization (**Figure 1**). Neurons in lower visual areas, with small receptive fields, are sensitive to basic visual features. For example, neurons in V1 respond predominantly to edges and lines. These neurons project to neurons at the next stage of the hierarchy, which code for more complex features. By V4, the neurons are selective for basic shapes, and by IT they respond in a viewpoint-invariant manner to full objects. Decisions making happens in the frontal cortex. This basic scenario has a well-defined set of characteristics. Processing is hierarchical, feedforward, and local on each level, i.e., only neighboring neurons, coding for neighboring parts in the visual field, project to a common higher-level neuron (**Figure 1**). In addition, processing at one stage is fully determined by processing at the previous stage. Information lost at previous stages is irretrievably lost. Processing follows an atomistic, Lego® building block type of encoding. For example, a hypothetical "square neuron" is created by feedforward projections from "lower" neurons coding for vertical and horizontal lines (**Figure 1**; Riesenhuber and Poggio, 1999; Hung et al., 2005; Serre et al., 2005, 2007a,b). Finally, there is an isomorphism between objects of the outer world (e.g., a blue line), basic neural circuitry (analyzing the blue line), and the corresponding percept ("blue line"). And this is exactly the beauty of these models: naturalizing the subjectivity of perception by identifying the basic neural circuits of perception.

Evidence for fast, hierarchical feedforward processing comes from experiments showing that humans can detect animals in a scene in less than 150 ms. Calculations based on neural conduction velocity show that there are only one or two spikes per cortical area before a decision is made, arguing strongly against feedback processing (Thorpe et al., 2001).

Computer vision models often follow closely the philosophy of neurobiological feedforward hierarchies. In these, as in neurobiological models, first, basic features are extracted, for example, through V1-style Gabor filtering or Haar wavelets. Often, the downstream hierarchical stages (V2, V4) are collapsed into one processing stage, where a classifier is trained to detect specialized objects such as faces or cars. Similar to IT neurons, these detectors are often scale- and viewpoint-invariant (Biederman, 1987; Ullman et al., 2002; Fink and Perona, 2003; Torralba, 2003; Schneiderman and Kanade, 2004; Viola and Jones, 2004; Felzenszwalb and Huttenlocher, 2005; Fei-Fei et al., 2006; Amit and Trouvé, 2007; Fergus et al., 2007; Heisele et al., 2007; Hoiem et al., 2008; Wu et al., 2010).

Here, we will present experiments from crowding research that challenge classical feedforward hierarchy models. In crowding, target discriminability strongly deteriorates when neighboring elements are presented (**Figure 2**). Crowding is often seen as a breakdown of object recognition and most models

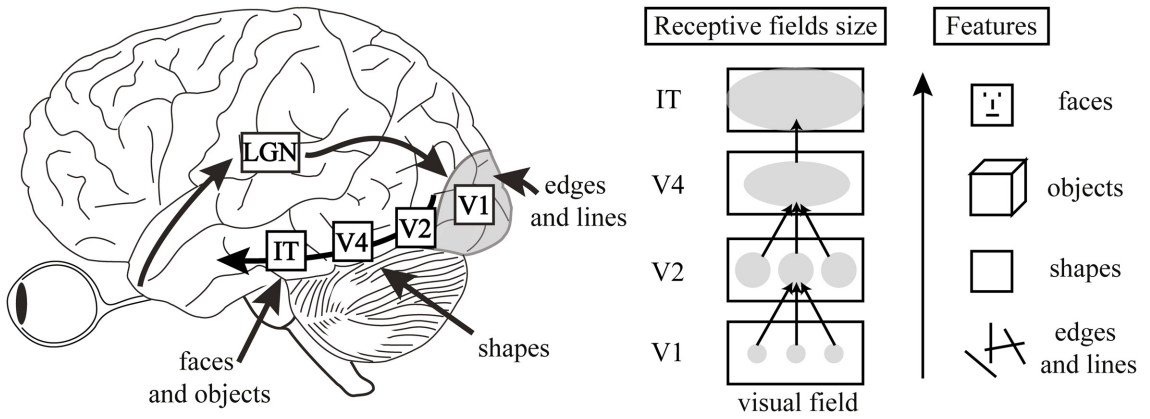

FIGURE 1 | Left: A typical hierarchical, feedforward model, where information processing starts at the retina, proceeds to the LGN, then to V1, V2, V4, and IT. Decisions about stimuli are made in the frontal cortex. **Center:** Lower visual areas have smaller receptive fields, while neurons in higher areas have gradually increasing receptive field sizes, integrating information over larger and larger regions of the visual field. **Right:** Lower visual areas, such as V1, code for basic features such as edges and lines. Higher-level neurons pool information over multiple low-level neurons with smaller receptive fields and code for more complex features. There is thus a hierarchy of features. Figure adapted from Manassi et al. (2013).

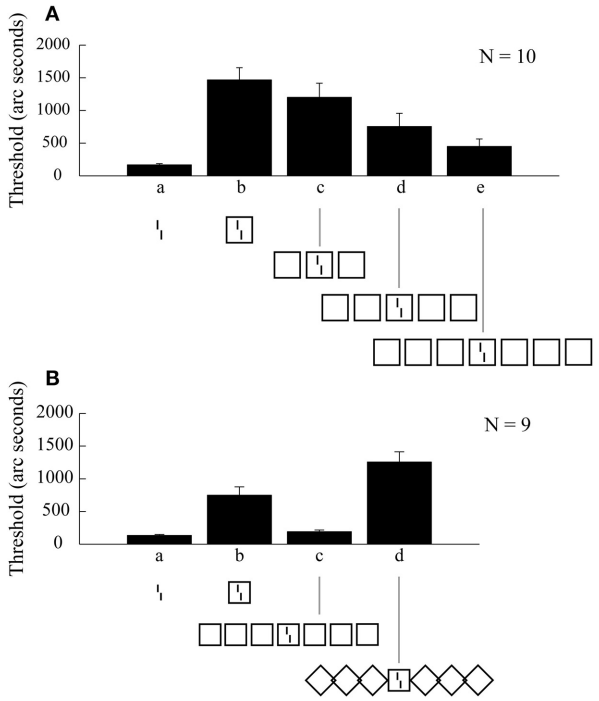

FIGURE 2 | Vernier offset discrimination as a function of stimulus configuration. (A). The reference stimulus is the un-flanked vernier shown in **a**. Enclosing the vernier in a square deteriorates performance. Adding additional squares leads to increasingly better performance. **(B)**. We replicated the results with the squares (b,c). In addition, rotating the flanking squares to form diamonds (d) undoes the effect of grouping and reinstates the crowding effect. From Manassi et al. (2013).

of crowding are very much in the spirit of object recognition models. In pooling models, information from lower-level neurons is pooled by higher-level neurons, to see wholes at the cost of more poorly perceiving the parts. Indeed, observers can clearly *detect* a crowded target, it is only its features and spatial relationships that are jumbled with flanker features (e.g., Pelli et al., 2004). Target-feature perception is lost because target and distracter features are pooled. A prediction made by pooling models is that, because spatial integration is local at each stage, only nearby elements deteriorate target discriminability (Bouma's law). In addition, if more flankers are added within Bouma's window, performance should deteriorate (or at least not improve) because the signal-to-noise ratio decreases. A third prediction is that adding more flankers should deteriorate performance.

In previous experiments, we presented a vernier stimulus, which consists of two vertical lines, offset slightly to the left or right (Manassi et al., 2012, 2013). Observers indicated the offset direction. Verniers were presented in the periphery, 9 degrees (of visual angle) to the right of fixation. Performance strongly deteriorated when the vernier was surrounded by a square (**Figures 2Aa, b**). This is a classic crowding effect and is well-explained by traditional crowding models. Next, Manassi et al. (2012, 2013) presented 2 × 3 neighboring squares (**Figure 2Ae**). According to pooling models, and most object recognition models, more flankers should deteriorate performance. However, the opposite was the case. Crowding almost disappeared. Interestingly, this *un*crowding effect increased with the number of squares that were presented (**Figure 2A**). Importantly, the fixation dot was only 0.5 degrees apart from the left-most square, i.e., the stimulus configuration extended over large parts of the right visual field. Hence, vernier offset discrimination is influenced by elements far outside the integration region predicted by Bouma's law. Second, and more importantly, vernier offset discrimination is influenced by the overall stimulus configuration. This becomes evident when turning the flanking squares by 90° creating diamonds, resulting in the return of the crowding effect (**Figure 2B**). Hence, figural aspects determine basic feature processing (Wolford and Chambers, 1983; Livne and Sagi, 2007; Malania et al., 2007; Sayim et al., 2010).

Our results clearly show that simple pooling models cannot explain crowding and the same seems to be true for most basic models of object recognition. Figural processing determines low-level processing as much as low-level processing determines figural processing. It seems that first the squares are computed from their constituting lines. Next square representations interact with each other and the outputs of this processing determine the vernier offset discriminability. This is reminiscent of the famous quote by Wertheimer that "the whole determines the appearance of the parts" (Wertheimer, 1938). In our example, the whole determines even low-level processing. It also agrees with more modern sentiments suggesting that feedback is crucial for normal vision at all levels of the processing hierarchy (Krüger et al., 2013). We propose that it is only when we know how elements group together that we will be able to accurately predict performance on even the simplest tasks, i.e., without understand grouping across the entire visual field, it is impossible to understand human object recognition.

Note here, that we are not claiming that the visual system is not hierarchical. Nor are we claiming that there is no feedforward sweep through the cortex. We are arguing against models that are both feedforward *and* contain a strict feature hierarchy. For example, classic models posit that low-level features (such as verniers) are encoded at an early cortical level and that shapes (such as squares) are encoded at a later cortical level. Square-square interactions are crucial, as we have shown. However, since there are no feedback connections, the classic models cannot explain how square-square interactions change low-level processing of the vernier. One solution is to give up feedforward processing and have *recurrent* interactions between lower and higher levels of processing.

Evidence for recurrent processing comes from timing experiments on the dynamics of grouping in crowding (Manassi and Herzog, 2013; Manassi, 2014). A vernier target was flanked by either two vertical lines, or by two vertical lines that formed the edges of two cuboids. In both cases, the vertical lines were identical and only the surrounding context differed—the lines grouped with the vernier, but when they were part of the cuboids, the lines segmented from the vernier. Vernier offset discrimination thresholds were measured as a function of stimulus presentation time for seven fixed durations ranging from 20 ms to 640 ms. Under brief presentation times ($\leq$120 ms) performance in the two stimulus conditions did not significantly differ. Beyond 160 ms, however, performance with the cuboids was significantly better than with the lines. These results indicate that perceptual grouping evolves with time, even for such basic stimuli as verniers. Current models of vernier offset discrimination show that this task can be achieved in a feedforward way by reading out the responses of orientation-tuned V1 neurons (Wilson, 1986)—a process that takes on the order of 50 ms (Cottaris and De Valois, 1998; Gershon et al., 1998). Sending spikes to additional synapses requires at least 10 ms per spike. Thus, the long time required for vernier discrimination in the cuboid flanker condition to be differentiated from line flanker condition ($\geq$160 ms, i.e., more than double the arrival time of the stimulus at V1) indicates significant additional cortical processing for perceptual grouping. Since 160 ms $-$50 ms $=$ 110 ms, at least 11 additional synaptic connections could be activated. Recent electrophysiological evidence suggests that the additional time can be accounted for by feedback connections from the lateral occipital cortex to earlier cortical areas, the result of which is the promotion of perceptual grouping (Shpaner et al., 2013).

Our results are not restricted to crowding but occur in many other visual paradigms including overlay masking (Saarela and Herzog, 2008, 2009), backward masking (Herzog, 2007; Hermens and Herzog, 2007; Dombrowe et al., 2009), letter recognition (Saarela et al., 2010), in haptics (Overvliet and Sayim, 2013), and in audition (Oberfeld et al., 2012).

Why does the processing of an element's basic features depend on remote elements? Vision is ill-posed. For example, the light (luminance) that arrives at the retina is a product of the light shining on an object (illuminance) and the material properties of the object (reflectance). Hence, on the photoreceptor level, it is impossible to determine whether or not a banana is yellow and ready to eat. The brain tries to solve this problem by discounting the illuminance, taking contextual information into account. This becomes obvious in the case of computing material properties. Glossy objects, for example, reflect bright spots (specularities) in regions of high curvature. Removal or addition of an object's specularities completely changes the object's perceived material, in spite of the fact that the rest of the object remains the same. To compute the material properties, integrating information across the visual field is crucial: where is the illuminance coming from? What is the shape of the object?

Key then, is that without knowing the whole one cannot know the parts. To the best of our knowledge, very few models adopt this approach of including recurrent processing and effectively integrating information over large parts of the visual field. Not surprisingly, these models are highly effective at modeling human data, not only from crowding, but also from many other areas of cognitive science, hinting at their general ability to explain cortical processing. For example, they effectively explain data pertaining to attention (Tsotsos, 1995; Tsotsos et al., 1995; Cutzu and Tsotsos, 2003; Bruce and Tsotsos, 2005, 2009; Rodriguez-Sanchez et al., 2007), and visual object learning (Bengio et al., 2013; Goodfellow et al., 2013; Salakhutdinov et al., 2013). They also do well at scene-segmentation, where successful models typically use a global approach, such as coarse-to-fine image pyramids (Estrada and Elder, 2006) or normalized cuts over extended graphs (Malik et al., 1999, 2001; Shi and Malik, 2000; Ren and Malik, 2002; Martin et al., 2004), which are leveraged to produce human-like scene segmentations. Here, again, computations are not purely local and feedforward, but rather global and iterative. Grossberg has also produced similar models in terms of their ability to do grouping that extends over a scene (Grossberg and Mignolla, 1985; Dresp and Grossberg, 1997), as has Francis (Francis et al., 1994; Francis and Grossberg, 1995). Future work will show whether these models can explain our particular crowding results.

In summary, there is a wealth of evidence suggesting that cortical processing is not purely hierarchical *and* feed-forward. In order to know how the visual system processes fine-grained

information at a particular location it is necessary to integrate information about the surrounding context over the entire visual field. Grouping and segmentation are crucial to understanding vision, and must be understood on a global scale.

FUNDING

Grant number 320030_135741.

ACKNOWLEDGMENTS

Aaron Clarke was funded by the Swiss National Science Foundation (SNF) project "Basics of visual processing: what crowds in crowding?" (Project number: 320030_135741).

REFERENCES

Amit, Y., and Trouvé, A. (2007). Pop: Patchwork of parts models for object recognition. *Int. J. Comput. Vis.* 75, 267–282. doi: 10.1007/s11263-006-0033-9

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: a review and new perspectives. *Patt. Analys. Mach. Intell. IEEE Trans.* 35, 1–30. doi: 10.1109/TPAMI.2013.50

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychol. Rev.* 94, 115. doi: 10.1037/0033-295X.94.2.115

Bruce, N., and Tsotsos, J. (2005). Saliency based on information maximization. *Adv. Neural Inf. Process. Syst.* 18, 1–8. Available online at: http://papers.nips.cc/paper/2830-saliency-based-on-information-maximization

Bruce, N. D., and Tsotsos, J. K. (2009). Saliency, attention, and visual search: an information theoretic approach. *J. Vis.* 9:5. doi: 10.1167/9.3.5

Cottaris, N. P., and De Valois, R. L. (1998). Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature* 395, 896–900. doi: 10.1038/27666

Cutzu, F., and Tsotsos, J. K. (2003). The selective tuning model of attention: psychophysical evidence for a suppressive annulus around an attended item. *Vis. Res.* 43, 205–219. doi: 10.1016/S0042-6989(02)00491-1

Dombrowe, I., Hermens, F., Francis, G., and Herzog, M. H. (2009). The roles of mask luminance and perceptual grouping in visual backward masking. *J. Vis.* 9:22. doi: 10.1167/9.11.22

Dresp, B., and Grossberg, S. (1997). Contour integration across polarities and spatial gaps: from local contrast filtering to global grouping. *Vis. Res.* 37, 913–924. doi: 10.1016/S0042-6989(96)00227-1

Estrada, F. J., and Elder, J. H. (2006). "Multi-scale contour extraction based on natural image statistics," in *Proc. IEEE Workshop on Perceptual Organization in Computer Vision* (New York, NY:IEEE), 183.

Fei-Fei, L., Fergus, R., and Perona, P. (2006). One-shot learning of object categories. *Patt. Analys. Mach. Intell. IEEE Trans.* 28, 594–611. doi: 10.1109/TPAMI.2006.79

Felzenszwalb, P. F., and Huttenlocher, D. P. (2005). Pictorial structures for object recognition. *Int. J. Comput. Vis.* 61, 55–79. doi: 10.1023/B:VISI.0000042934.15159.49

Fergus, R., Perona, P., and Zisserman, A. (2007). Weakly supervised scale-invariant learning of models for visual recognition. *Int. J. Comput. Vis.* 71, 273–303. doi: 10.1007/s11263-006-8707-x

Fink, M., and Perona, P. (2003). "Mutual boosting for contextual inference," in *NIPS*, Vol. 2 (Vancouver, BC), 7.

Francis, G., and Grossberg, S. (1995). *Cortical Dynamics of Boundary Segmentation and Reset: Persistence, Afterimages, and Residual Traces.* Technical report, Boston University Center for Adaptive Systems and Department of Cognitive and Neural Systems.

Francis, G., Grossberg, S., and Mingolla, E. (1994). Cortical dynamics of feature binding and reset: control of visual persistence. *Vis. Res.* 34, 1089–1104. doi: 10.1016/0042-6989(94)90012-4

Gershon, E. D., Wiener, M. C., Latham, P. E., and Richmond, B. J. (1998). Coding strategies in monkey v1 and inferior temporal cortices. *J. Neurophysiol.* 79, 1135–1144.

Goodfellow, I. J., Courville, A., and Bengio, Y. (2013). Scaling up spike-and-slab models for unsupervised feature learning. *Pattern Anal. Mach. Intell. IEEE Trans.* 35, 1902–1914. doi: 10.1109/TPAMI.2012.273

Grossberg, S., and Mignolla, E. (1985). Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol. Rev.* 92, 173–211. doi: 10.1037/0033-295X.92.2.173

Heisele, B., Serre, T., and Poggio, T. (2007). A component-based framework for face detection and identification. *Int. J. Comput. Vis.* 74, 167–181. doi: 10.1007/s11263-006-0006-z

Hermens, F., and Herzog, M. H. (2007). The effects of the global structure of the mask in visual backward masking. *Vis. Res.* 47, 1790–1797. doi: 10.1016/j.visres.2007.02.020

Herzog, M. H. (2007). Spatial processing and visual backward masking. *Adv. Cogn. Psychol.* 3, 85–92. doi: 10.2478/v10053-008-0016-1

Hoiem, D., Efros, A. A., and Hebert, M. (2008). Putting objects in perspective. *Int. J. Comput. Vis.* 80, 3–15. doi: 10.1007/s11263-008-0137-5

Hung, C. P., Kreiman, G., Poggio, T., and DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science* 310, 863–866. doi: 10.1126/science.1117593

Krüger, N., Janssen, P., Kalkan, S., Lappe, M., Leonardis, A., Piater, J., et al. (2013). Deep hierarchies in the primate visual cortex: What can we learn for computer vision? *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1847–1871. doi: 10.1109/TPAMI.2012.272

Livne, T., and Sagi, D. (2007). Configuration influence on crowding. *J. Vis.* 7:4. doi: 10.1167/7.2.4

Malania, M., Herzog, M. H., and Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *J. Vis.* 7, 1–7. doi: 10.1167/7.2.1

Malik, J., Belongie, S., Leung, T., and Shi, J. (2001). Contour and texture analysis for image segmentation. *Int. J. Comput. Vis.* 43, 7–27. doi: 10.1023/A:1011174803800

Malik, J., Belongie, S., Shi, J., and Leung, T. (1999). "Textons, contours and regions: Cue integration in image segmentation," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Vol. 2 (Kerkyra: IEEE), 918–925. doi: 10.1109/ICCV.1999.790346

Manassi, M. (2014). *Crowding, Grouping and Object Recognition*. Ph. D. thesis, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

Manassi, M., and Herzog, M. (2013). Crowding and grouping: how much time is needed to process good gestalt? *Perception* 42, 229. doi: 10.1068/v130301

Manassi, M., Sayim, B., and Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *J. Vis.* 12:13. doi: 10.1167/12.10.13

Manassi, M., Sayim, B., and Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *J. Vis.* 13:10. doi: 10.1167/13.13.10

Martin, D. R., Fowlkes, C. C., and Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. *Pattern Anal. Mach. Intell. IEEE Trans.* 26, 530–549. doi: 10.1109/TPAMI.2004.1273918

Oberfeld, D., Stahn, P., and Kuta, M. (2012). Binaural release from masking in forward-masked intensity discrimination: evidence for effects of selective attention. *Hear. Res.* 294, 1–9. doi: 10.1016/j.heares.2012.09.004

Overvliet, K., and Sayim, B. (2013). Contextual modulation in haptic vernier offset discrimination. *Percept. 42 ECVP Abstr. Suppl.* 175. Available online at: http://www.perceptionweb.com/abstract.cgi?id=v130322

Pelli, D. G., Palomares, M., and Majaj, N. J. (2004). Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J. Vis.* 4:12. doi: 10.1167/4.12.12

Ren, X., and Malik, J. (2002). A probabilistic multi-scale model for contour completion based on image statistics. *Comput. Vis. ECCV 2002*, 7, 312–327. doi: 10.1007/3-540-47969-4_21

Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025. doi: 10.1038/14819

Rodriguez-Sanchez, A. J., Simine, E., and Tsotsos, J. K. (2007). Attention and visual search. *Int. J. Neural Syst.* 17, 275–288. doi: 10.1142/S0129065707001135

Saarela, T. P., and Herzog, M. H. (2008). Time-course and surround modulation of contrast masking in human vision. *J. Vis.* 8, 1–10. doi: 10.1167/8.3.23

Saarela, T. P., and Herzog, M. H. (2009). Size tuning and contextual modulation of backward contrast masking. *J. Vis.* 9, 1–12. doi: 10.1167/9.11.21

Saarela, T. P., Westheimer, G., and Herzog, M. H. (2010). The effect of spacing regularity on visual crowding. *J. Vis.* 10, 1–7. doi: 10.1167/10.10.17

Salakhutdinov, R., Tenenbaum, J. B., and Torralba, A. (2013). Learning with hierarchical-deep models. *Pattern Anal. Mach. Intell. IEEE Trans.* 35, 1958–1971. doi: 10.1109/TPAMI.2012.269

Sayim, B., Westheimer, G., and Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychol. Sci.* 21, 641–644. doi: 10.1177/0956797610368811

Schneiderman, H., and Kanade, T. (2004). Object detection using the statistics of parts. *Int. J. Comput. Vis.* 56, 151–177. doi: 10.1023/B:VISI.0000011202.85607.00

Serre, T., Oliva, A., and Poggio, T. (2007a). A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6424–6429. doi: 10.1073/pnas.0700622104

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. (2007b). Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.* 29, 411–426. doi: 10.1109/TPAMI.2007.56

Serre, T., Wolf, L., and Poggio, T. (2005). "Object recognition with features inspired by visual cortex," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005*, Vol. 2 (San Diego, CA: IEEE), 994–1000.

Shi, J., and Malik, J. (2000). Normalized cuts and image segmentation. *Pattern Anal. Mach. Intell. IEEE Trans.* 22, 888–905. doi: 10.1109/34.868688

Shpaner, M., Molholm, S., Forde, E., and Foxe, J. J. (2013). Disambiguating the roles of area v1 and the lateral occipital complex (loc) in contour integration. *Neuroimage* 69, 146–156. doi: 10.1016/j.neuroimage.2012.11.023

Thorpe, S., Delorme, A., and Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Netw.* 14, 715–725. doi: 10.1016/S0893-6080(01)00083-1

Torralba, A. (2003). Contextual priming for object detection. *Int. J. Comput. Vis.* 53, 169–191. doi: 10.1023/A:1023052124951

Tsotsos, J. K. (1995). Toward a computational model of visual attention. *Early Vis. Beyond* 1, 207–218.

Tsotsos, J. K., Culhane, S. M., Kei Wai, W. Y., Lai, Y., Davis, N., and Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artif. Intell.* 78, 507–545. doi: 10.1016/0004-3702(95)00025-9

Ullman, S., Vidal-Naquet, M., and Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat. Neurosci.* 5, 682–687. doi: 10.1038/nn870

Viola, P., and Jones, M. J. (2004). Robust real-time face detection. *Int. J. Comput. Vis.* 57, 137–154. doi: 10.1023/B:VISI.0000013087.49260.fb

Wertheimer, M. (1938). *Gestalt Theory*. London: Hayes Barton Press.

Wilson, H. R. (1986). Responses of spatial mechanisms can explain hyperacuity. *Vis. Res.* 26, 453–469. doi: 10.1016/0042-6989(86)90188-4

Wolford, G., and Chambers, L. (1983). Lateral masking as a function of spacing. *Percept. Psychophys.* 33, 129–138. doi: 10.3758/BF03202830

Wu, Y. N., Si, Z., Gong, H., and Zhu, S.-C. (2010). Learning active basis model for object detection and recognition. *Int. J. Comput. Vis.* 90, 198–235. doi: 10.1007/s11263-009-0287-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 April 2014; accepted: 03 October 2014; published online: 22 October 2014.
Citation: Herzog MH and Clarke AM (2014) Why vision is not both hierarchical and feedforward. Front. Comput. Neurosci. ***8****:135. doi: 10.3389/fncom.2014.00135*
This article was submitted to the journal Frontiers in Computational Neuroscience.