UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A Bayesian hierarchical model of local-global processing: visual crowding as a case-study

Permalink
https://escholarship.org/uc/item/98m8569z

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 37(0)

Authors
Zhang, Shunan
Song, Man
Yu, Angela J

Publication Date
2015

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

A Bayesian hierarchical model of local-global processing: visual crowding as a case-study

Shunan Zhang **Man Song** **Angela J. Yu**
(s6zhang@ucsd.edu, feijuejuanling@gmail.com, ajyu@ucsd.edu)
Department of Cognitive Science, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0515

Abstract

We explore the interaction between local-global information processing in visual perception, leveraging a visual phenomenon known as crowding, whereby the perception of a target stimulus is impaired by the presence of nearby flankers. The majority of established models explain the crowding effect in terms of local interactions. However, recent experimental results indicate that a classical crowding effect, the deterioration in the discrimination of a vernier stimulus embedded in a square, is alleviated by the presence of additional flanker squares ("uncrowding"). Here, we propose that crowding and uncrowding arise from cortical inferences about hierarchically organized groups, and formalize this concept using a hierarchical Bayesian model. We show that the model reproduces both crowding and uncrowding for flanked vernier discrimination. More generally, the model provides a normative explanation of how visual information might simultaneously flow bottom-up, top-down, and laterally, to allow the visual system to interactively process local and global features in the visual scene.

Introduction

A common approach to understanding visual perception is to identify the "local" versus "global" processing of visual information, where local processing depends on spatially proximate visual elements, and global processing is more holistic in nature and thus influenced by spatially distal elements. The local versus global distinction has long been an important topic in Gestalt psychology (Wagemans et al., 2012) and neuroscience (Rasche & Koch, 2002). However, recent discoveries in visual perception, in particular in the study of visual crowding (Manassi, Sayim, & Herzog, 2013; Clarke, Herzog, & Francis, 2014), suggest that the distinction may be overly constraining, and that it is necessary to explore the recurrent interactions of local and global processing to fully understand computational principles in vision. Visual crowding is a psychophysical phenomenon in which target discrimination is impaired in the presence of flankers (also known as distractors). Crowding effects have been extensively studied in various settings, including letter recognition, vernier acuity, orientation discrimination, face recognition, etc. It is believed that crowding reflects a crucial bottleneck in peripheral visual discrimination and has a close link with contour integration (May & Hess, 2007; Hess, Dakin, Kapoor, & Tewfik, 2000), segmentation (Wilkinson, Wilson, & Ellemberg, 1997), feature interaction (Van den Berg, Roerdink, & Cornelissen, 2010; Freeman & Pelli, 2007), and object recognition (Levi, 2008; Whitney & Levi, 2011; Pelli, 2008).

Most theories about visual crowding explain it using essentially local mechanisms, where degraded target processing is due to the (inappropriate) pooling of information about the target and the surrounding elements. Examples of such models include spatial pooling (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Dayan & Solomon, 2010), substitution (Huckauf & Heller, 2002; Strasburger, Harvey, & Rentschler, 1991), masking (Tyler & Likova, 2007), and feature integration (Freeman & Pelli, 2007). These models explain some of the key phenomena in crowding, such as the asymmetrical crowding influence between foveal and peripheral stimuli (Dayan & Solomon, 2010). However, recent experimental studies suggest that crowding may involve complex global and feedback processing (Levi & Carney, 2009; Livne & Sagi, 2007; Malania, Herzog, & Westheimer, 2007; Saarela, Sayim, Westheimer, & Herzog, 2009; Manassi et al., 2013), as crowding effects can be significantly reduced or even eliminated (Manassi et al., 2013) when multiple flankers form a perceptual pattern that excludes the target. Previous studies have also found target-flanker proximity to be a major factor that modulates the crowding effect, with a particular separation distance maximizing the crowding effect (Hess et al., 2000). These non-linear effects on crowding suggest that a new approach for computational modeling of visual crowding is needed.

Here, we propose an alternative theory: the brain simultaneously processes local and global features (Lamme & Roelfsema, 2000), in particular using local features to delineate the hierarchically organized grouping of elements, and in turn using global grouping information to facilitate inferences about local features. The impairment of vernier discrimination in the presence of the square flanker (e.g. Manassi et al., 2013) may be due to the perceptual grouping of the vernier stimulus with the square, such that the straight, well-aligned edges of the square leads to a biased perception of the vernier lines being closer together, thus leading to a deterioration in offset discrimination. When additional square flankers are introduced at regular spacing, the square surrounding the vernier is perceived as being part of the array of squares, and the vernier stimulus is perceived as not belonging to this grouping, and thus features of the square grouping (straight edges) has a smaller influence on the perceived properties of the vernier stimulus (offset spacing). Formally, we introduce a hierarchical Bayesian model that incorporates prior assumptions about grouping of similar features at different spatial and complexity levels, while making inferences about the target, through simultaneous top-down and bottom-up information processing. This grouping-based explanation is consistent with the observation that crowding effects are particularly strong when target and flankers are similar in shape and size (Kooi, Toet, Tripathy, & Levi, 1994), orientation (Levi, Hariharan, & Klein, 2002; Hariharan, Levi, & Klein, 2005) , polarity (Kooi et al., 1994; Chakravarthi & Cavanagh, 2007), spatial frequency (Chung, Levi, & Legge, 2001), color (Kooi et al., 1994; Põder & Wagemans, 2007), depth (Kooi et al., 1994) and motion (Banton & Levi, 1993).

In the following, we first describe the model, then demonstrate how the model explains the crowding and uncrowding effects, and finally conclude with some discussions of related work and future directions.

A Bayesian Model of Visual Crowding in the Vernier Discrimination Task

Visual crowding has been studied intensively in the Vernier discrimination task, a classical perceptual psychophysics paradigm. In the vernier discrimination experiment, e.g. in (Manassi et al., 2013), observers judge whether one line is offset to the left or the right of a second line (Figure 1a); in the crowding version of the experiment, the vernier stimulus is surrounded by a square, and possibly other flankers/distractors (Figure 1b:c). Typically, experimenters identify a discrimination threshold, by measuring the offset distance for which a certain accuracy level is attained (e.g. 75% correct in Manassi et al. (2013)). When vernier offset discrimnation are tested with several stimulus setting – a single vernier, a vernier embedded in one square, and an embedded vernier with horizontally surrounded squares – it was found that, compared with the single vernier condition, the threshold increased (performance deteriorated) in the presence of one flanker, exhibiting the classic crowding effect; but threshold decreased (performance improved), relative to the crowding condition, when the vernier is surrounded by an array of square flankers in addition to the square flanker immediately surrounding the vernier target (Manassi et al., 2013). This latter result, termed "uncrowding", defies any simple local processing account of crowding; instead, it suggests that hierarchical grouping, which probably necessitates global processing, is inherently present in visual crowding.

Here, we propose a normative Bayesian model that makes joint inferences about the identity and grouping of all the objects in the visual scene, and marginalizes over beliefs about the rest of the visual scene to infer offset property about the central target. Below, we first describe

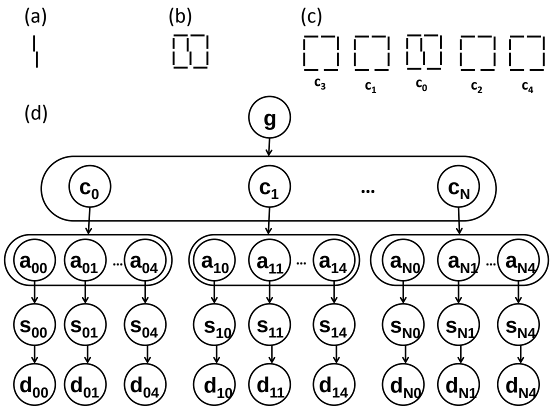

Figure 1: Generative model for the Vernier discrimination task. (a) Model representation of a single vernier. (b) Model representation of a vernier embedded in a square; edges perceived as composition of line segments. We denote the central vernier as a_0, and the four edges as a_1 to a_4. (c) Model representation of an embedded vernier surrounded by additional square flankers. We use location 0 to denote the target location, then location 1 and 2 for its two neighboring locations on the left and the right, and so forth. (d) Graphical illustration of the Bayesian generative model: the global grouping variable, g, specifies whether the objects from all locations are the same. c_n specifies the semantic relationship of the parts that compose the visual object at location n, i.e. either all line segments are unrelated, or they form a square plus an unrelated central edge (if it is present), or all the line segments form a single "window" object. Likewise, $a = 1$ indicates line segments that are well-aligned and belong to the same edge, and $a = 2$ indicates unrelated lines segments.

the *generative model*, which encapsulates all our assumptions about what subjects believe about general organizational principles of the visual scene as well as how noisy sensory observations are generated from the visual scene. We then describe the *recognition mode*, which inverts the generative model, using a combination of Bayes' Rule and other standard statistical tools, to predict how perception arises from noisy sensory data.

Generative Model

In the model, we use the variable g to denote the (potential) global grouping of objects across all N stimulus locations. $g = 1$ means (1) there is at least one object inferred at each location, and (2) the same object is present at all locations (e.g. square present at all locations). Otherwise, $g = 0$. The prior probability of grouping ($g = 1$) is parameterized as p.

We model the visual scene as being made up of an array of N decomposable visual stimuli, $c_0, c_1, ..., c_N$, where c_0 always refers to the central location. $c_n = 0$ denotes that there is no coherent visual object (though there may be semantically unrelated visual input) at location n. $c_n = s$ denotes that there is a square present at location n (and there may or may not be semantically unrelated, additional visual object at this location). $c_n = w$ denotes that there is a "window" object at location n, which consists of a central edge embedded in a square. The distribution of $\mathbf{c}$ conditioned on $g = 1$ is correlated among the different locations:

$$\Pr(c_0 = c_1 = \cdots = c_N = s|g = 1) = \varphi_s$$
$$\Pr(c_0 = c_1 = \cdots = c_N = w|g = 1) = \varphi_w$$
$$\Pr(\mathbf{c}|g = 1) = \frac{1 - (\varphi_s + \varphi_w)}{3^{N+1} - 2}, \text{o.w.}$$

For simplicity, the model only allows the array of objects to all be squares or all be "windows". When there is no global grouping ($g = 0$), the conditional distribution of $\mathbf{c}$ is independent (factorizable):

$$\Pr(\mathbf{c}|g = 0) = \prod_{n=0}^{N} \Pr(c_n)$$

The model assumes that a straight line segment generates noisy sensory evidence of well-aligned shorter line segments (mimicking the neuroscientific observations that visual neurons from retinal ganglion cells to LGN, to V1, are sensitive to oriented line segments of increasing lengths; and the general notion that larger receptive fields depend on pooling responses from neurons with smaller receptive fields). We use the variable a to denote the grouping (alignment) of the two lines that constitute the vernier. $a = 1$ means that the two lines are aligned, $a = 2$ means that the two lines are not aligned, and $a = 0$ denotes the absence of any lines. The prior probabilities of $a = 1$ and $a = 2$ are q_1 and q_2. Let ψ be a large probability, when c_n is a square,

$$\Pr(a_{n0} = 1, a_{n1} = \cdots = a_{n4} = 1|c = s) = \psi \cdot q_1$$
$$\Pr(a_{n0} = 2, a_{n1} = \cdots = a_{n4} = 1|c = s) = \psi \cdot q_2$$
$$\Pr(a_{n0} = 0, a_{n1} = \cdots = a_{n4} = 1|c = s) = \psi \cdot (1 - q_1 - q_2)$$
$$\Pr(a_{n0}, \cdots, a_{n4}|c = s) = \frac{1 - \psi}{3^5 - 3}, \quad \text{o.w.}$$

When c_n is a "window", the conditional probability distribution of $a_{n0}, ... a_{n4}$ has most of its mass (ψ) on $a_{n0} = ... = a_{n4} = 1$, and the rest of the mass split equally to all other conditions ($3^5 - 1$ in total). If there is no object,

$$\Pr(a_{n0}, \cdots, a_{n4}|c = 0) = \prod_{i=0}^{4} \Pr(a_{ni})$$

The offset of the vernier, s, should be close to 0 if the two lines are aligned, but can vary across a broader range of values otherwise. Conditioned on a, s has a Gaussian distribution centered at 0; its variance is close to 0 when $a = 1$, but large when $a = 2$. We assume that there is one population of neurons whose activities are driven by each vernier offset, yielding noisy perception of the offset, d. We assume d has a Gaussian distribution conditioned on s when $s \notin \emptyset$.

$$\Pr(d|s) = \mathcal{N}(s, \sigma_d^2)$$

Recognition Model

Given the noisy input of $\mathbf{d}$ from all existing verniers at all locations, the ideal observer's belief about the true offset of the target vernier, s_{00}, the alignment of the target vernier, a_{00}, the object at the central location, c_0, and the global grouping g, captured by the probability distribution, $P(s_{00}, a_{00}, c_0, g|\mathbf{d})$, is proportional to

$$p(g) \int_{\mathbf{c}} p(\mathbf{c}|g) \int_{\mathbf{a}_n} p(\mathbf{a}_n|c_n) \prod_{i=0}^{4} p(s_{ni}|a_{ni}) p(d_{ni}|s_{ni}) \qquad (1)$$

by Bayes' rule, where $\mathbf{c}$ is shorthand for the object variables across all locations, and $\mathbf{a}_n$ is shorthand for all the vernier alignment variables at one location. This function is proportional to the posterior distribution (without normalizing factor). In practice, we make the observation of the lines' presence/absence noise-free, i.e. subjects should not hallucinate vernier's presence when it is actually absent, or vice versa.

To make a perceptual decision based on the posterior probability, we compute the marginal distribution of s_{00}, by summing over the uncertainty over a_{00}, c_0, and g. How decisions of the offset direction should be made based on the inferred offset size of the vernier is an interesting problem by itself. For the current study, however, we focus on the representational component of the task, and assume a simple, sigmoidal mapping from the posterior mean of the offset, $\hat{s}$, to decision accuracy:

$$a = \frac{1}{1 + e^{-\beta \hat{s}}} \qquad (2)$$

where β is a free parameter that is associated with the internal discriminability.

Simulations

The model infers the offset distance of the target vernier, along a grid of true offset values from .1 to 1.5, for three conditions including 1) single vernier, 2) vernier embedded in a square and 3) additional flankers. For each condition, we ran simulations using the same 10 different values of σ_d^2, iid from $N(.1, .1)$. We generated 500 observations for each offset distance for each particular d. σ_1 and σ_2 were .1 and 1. We denote the posterior mean of s by $\hat{s}_d$; the mean inferred offset, $\hat{s}$, is the average of all $\hat{s}_d$. For Figure 2, $q_1 = .05$, $q_2 = .85$, $\psi = .9999$, $\varphi_w = .7$, $p = .5$, $\beta = -3$. The graphical models were implemented via WinBUGS 1.4.3 (Lunn, Thomas, Best, & Spiegelhalter, 2000), which implements MCMC for posterior inference of hidden variables and model parameters.

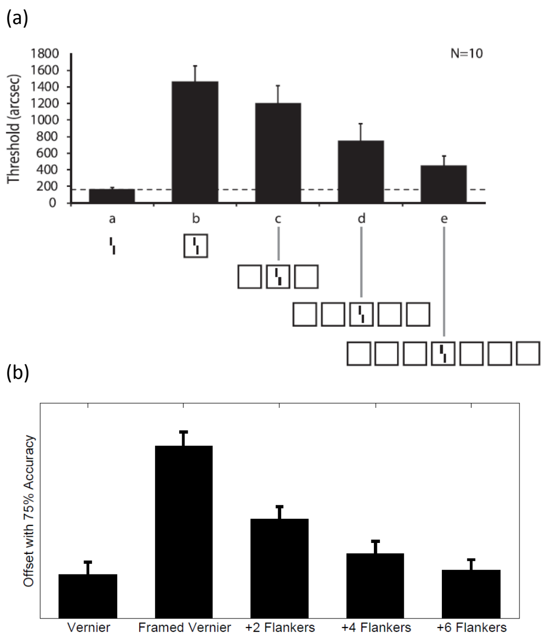

Figure 2: Model simulation results consistent with human performance (top plot) as reported in (Manassi et al., 2013). The exact forms of the stimuli used in (Manassi et al., 2013) are shown along with the horizontal axis in the top plot. Each bar shows the offset distance for which 75% correct responses occurred. Thresholds increased when the vernier was embedded in a square, but gradually decreased when in creasing the number of flanking squares. In the bottom plot, thresholds calculated from model simulations for 75% accuracy, showing a similar pattern as in the behavioral data. Error bars calculated across simulation results using the 10 different σ_d values.

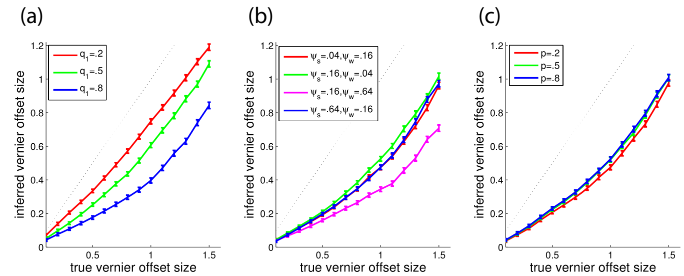

Figure 3: (a) Single vernier discrimination task: inferred vernier offset distances are more under-estimated when the prior probability of grouping of the lines (q_1) increases. (b) Vernier embedded in a square: effects of varying the prior probability of a square (φ_s), a "window" (φ_w), or no object ($1-\varphi_s-\varphi_w$). (c) Embedded vernier with flankers: inferred vernier offset distances increase (but still under-estimated) when the prior probability of a global group (p) increases.

Results

By simulating our hierarchical Bayesian model (details in the previous section), we show that our model is able to recover the observed patterns in the behavioral experiment as reported in (Manassi et al., 2013). As shown in Figure 2, when we simulate our model at 75% discrimination accuracy for the different experimental conditions, the predicted model discrimination threshold shows a similar pattern as human behavioral data (Manassi et al., 2013).

Having seen that our model can reproduce the behavioral pattern, we then explore further how the model works by simulating the inferred offset distance for each of the experimental conditions (single vernier, framed vernier, additional flankers), under different parameter settings (see Figure 3), in particular for different priors of the grouping variable at each level. When there is a single vernier stimulus, the model predicts that the inferred offset distance should be in general smaller than the true offset distance, due to the influence of the of the grouping variable (Figure 3a). As the prior probability of the two lines being grouped together into a single well-aligned object increases, the perceived offset distance becomes smaller. When the prior probability of such a grouping goes toward 0, or when the true offset distance gets very large, the under-estimation of offset distance also diminishes toward 0 (results not shown).

When the vernier is embedded in a square frame, we assume that the visual system can either perceive a "window", a "square" (plus any semantically unrelated visual input), or "no object" (completely incoherent visual input) at this central location. As illustrated in Figure 3b, the inferred vernier offset distance is smallest, and thus the discrimination performance is the worst, when the prior probability of "window" is high (magenta). This is because when there is a relatively high prior probability of perceiving a single "window" object, the target vernier "inherits" the "well-aligned-ness" of the edges of the surrounding square, with q_1 (in Figure 3a) no longer being a model parameter but a hidden parameter whose distribution is influenced top-down by the perception of the "window" (see previous section for details).

As the prior of the presence of a global group is increased, the model increasingly under-estimates the offset magnitude, and the probability that the central square is grouped with the flanker squares becomes higher (Figure 3c).

Discussion

We introduced a Bayesian hierarchical model for visual scene processing, which makes simultaneous inferences about (relatively) global grouping membership at different levels of organizational hierarchy, as well as (relatively) local visual features. The model explains the crowding phenomenon as follows: when the target vernier is presented along with a square frame, it may be perceptually grouped with the square flanker and thus partially inherits the features of other group members (straight edges), making the perceived offset distance smaller and thus the discrimination threshold higher. On the other hand, when additional squares are presented as flankers, inferences about global group membership result in perceptual grouping of all the squares, and the "decoupling" of the vernier from that group, resulting in a reduced propagation of features from the square frame to the target stimulus, and thus a lower discrimination threshold. Our model constitutes a new approach for explaining the counter-intuitive phenomenon of uncrowding, using bidirectional information flow in the simultaneous processing of global and local features of the visual scene.

Our model is related to other mechanistic models that involve the feed-forward and feedback loops in visual perception (e.g. Lamme & Roelfsema, 2000), but our model is distinctive in being a normative Bayesian generative model (a "computational" model in the parlance of Marr's three levels of analyses (Marr, 1982)). It is also related to other Bayesian models of visual processing (Yu & Dayan, 2005; Dayan & Solomon, 2010), although those models all essentially utilize localized pooling operations and thus cannot be easily modified to accommodate the non-local uncrowding effects. At an abstract level, our account of uncrowding is somewhat analogous to the Bayesian account of multimodal cue combination (Kording et al., 2007), whereby the diminishing of auditory-visual cue integration at larger spatial separations is explained in terms of a higher-level perception that the cues may not originate from the same object source. In terms of the crowding literature, the current model is closest in spirit to the compatibility bias model we proposed earlier to explain flanker congruency effects in the Eriksen task (Yu, Dayan, & Cohen, 2009).

Our model makes several predictions that can be tested in future experimental work. For example, in the uncrowding condition, vernier discriminability depends on the relative probability of the central square plus venier being perceived as a single or two separate visual objects as a function of the surrounding flankers. If instead of using square flankers, one used "window" flankers (squares with a vertical bisecting segment), then our model would predict an increased probability of perceiving the vernier stimulus and the square frame as being the same object, and thus result in an even worse discrimination threshold than without the additional flankers. Another subtle prediction our model makes, is that the deterioration in vernier discriminability in the presence of the square flanker is due to a systematic bias (under-estimation) of the offset distance, as opposed to an increase in the variance of perceived offset distance. We therefore predict, that if subjects were asked to compare two vernier stimuli, one alone and the other framed by a square, then they should consistently report that the flanked vernier has a smaller offset distance, even when the two are the same.

This paper primarily focused on the representational component of the vernier task, which deals with how information about the vernier stimulus and the rest of the visual scene are propagated, integrated, and distributed; the paper did not focus much on the decision component, which has to do with the transformation from the visual representation to the behavioral output, such as when and how to respond. It is a reasonable approximation in that people's performance on the visual discrimination task, in terms of the percentage of correct decisions, is generally well-captured by a sigmoidal psychometric function of the inferred stimulus magnitude. An alternative approach might be to assume that observers may make a Vernier decision based on a few samples drawn from the posterior distribution over offset magnitude, similar to previous models of human probabilistic decision-making in other contexts (Vul, Goodman, Griffiths, & Tenenbaum, 2009). More importantly, the current model lacks a principled way to explain how perceptual representation and response tendencies evolve over longer viewing time, especially in light of recent experimental results showing that contextual effects in vernier crowding depends on viewing time (Manassi, Clarke, Chicherov, & Herzog, 2014). Future work are needed to explore more explicitly the decisional and the temporal aspect of vernier discrimination.

Finally, the model presented here exemplifies a new normative approach for capturing and explaining local-global interactions in visual processing. In particular, it points out the important role played by perceptual grouping at multiple levels of hierarchical organization. Currently, our model focuses on capturing the global representation for the specific vernier crowding task. But this hierarchical framework can be extended to explain other visual phenomena, such as many of the Gestalt principles for grouping. A fruitful direction of future research would be to extend the model to account for broader classes of visual perceptual phenomena that involve complex interactions among stimuli in the visual scene.

References

Banton, T., & Levi, D. M. (1993). Spatial localization of motion-defined and luminance-defined contours. *Vision research*, *33*(16), 2225–2237.

Chakravarthi, R., & Cavanagh, P. (2007). Temporal properties of the polarity advantage effect in crowding. *Journal of Vision*, *7*(2), 11.

Chung, S. T., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision research*, *41*(14), 1833–1850.

Clarke, A. M., Herzog, M. H., & Francis, G. (2014). Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception. *Frontiers in psychology*, *5*.

Dayan, P., & Solomon, J. A. (2010). Selective bayes: Attentional load and crowding. *Vision research*, *50*(22), 2248–2260.

Freeman, J., & Pelli, D. G. (2007). An escape from crowding. *Journal of Vision*, *7*(2), 22.

Hariharan, S., Levi, D. M., & Klein, S. A. (2005). Crowding?in normal and amblyopic vision assessed with gaussian and gabor cs. *Vision research*, *45*(5), 617–633.

Hess, R. F., Dakin, S. C., Kapoor, N., & Tewfik, M. (2000). Contour interaction in fovea and periphery. *JOSA A*, *17*(9), 1516–1524.

Huckauf, A., & Heller, D. (2002). What various kinds of errors tell us about lateral masking effects. *Visual Cognition*.

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial vision*, *8*(2), 255–279.

Kording, K. P., Beierholm, U., Ma, W., Quartz, S., Tenenbaum, J., & Shams, L. (2007). Causal inference in cue combination. *PLOSOne*, *2*(9), e943.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences*, *23*(11), 571–579.

Levi, D. M. (2008). Crowdingan essential bottleneck for object recognition: A mini-review. *Vision research*, *48*(5), 635–654.

Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better. *Current Biology*, *19*(23), 1988–1993.

Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in amblyopic vision. *Vision research*, *42*(11), 1379–1394.

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, *7*(2), 4.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). Winbugs-a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing*, *10*(4), 325–337.

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision*, *7*(2), 1.

Manassi, M., Clarke, A., Chicherov, V., & Herzog, M. H. (2014, May). *Crowding, grouping, timing.* Vision Sciences Society. (Abstract for Vision Sciences Society 14th Annual Meeting)

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of vision*, *13*(13), 10.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

May, K. A., & Hess, R. F. (2007). Ladder contours are undetectable in the periphery: A crowding effect? *Journal of Vision*, *7*(13), 9.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature neuroscience*, *4*(7), 739–744.

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current opinion in neurobiology*, *18*(4), 445–451.

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision*, *7*(2), 23.

Rasche, C., & Koch, C. (2002). Recognizing the gist of a visual scence: possible perceptual and neural mechanisms. *Neurocomputing*.

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision*, *9*(2), 5.

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, *49*(6), 495–508.

Tyler, C. W., & Likova, L. T. (2007). Crowding: A neuroanalytic approach. *Journal of vision*, *7*(2), 16.

Van den Berg, R., Roerdink, J. B., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS computational biology*, *6*(1), e1000646.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? optimal decisions from very few samples. In *Proceedings of the 31st annual conference of the cognitive science society.* Amsterdam, Netherlands.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization. *Psychol. Bull.*, *138*, 1172-1217.

Whitney, D., & Levi, D. M. (2011). Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends in cognitive sciences*, *15*(4), 160–168.

Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *JOSA A*, *14*(9), 2057–2068.

Yu, A. J., & Dayan, P. (2005). Inference, attention, and decision in a Bayesian neural architecture. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17.* Cambridge, MA: MIT Press.

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: toward a rational bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 700.