Modelling binocular disparity processing from statistics in natural scenes

Tushar Chauhan, Yseult Héjja-Brichard, Benoit Cottereau

▶ **To cite this version:**

Tushar Chauhan, Yseult Héjja-Brichard, Benoit Cottereau. Modelling binocular disparity processing from statistics in natural scenes. Vision Research, 2020, 176, pp.27-39. 10.1016/j.visres.2020.07.009 . hal-03009669

HAL Id: hal-03009669
https://hal.science/hal-03009669

Submitted on 19 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vision Research 176 (2020) 27–39

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Modelling binocular disparity processing from statistics in natural scenes

Tushar Chauhan[a,b,*], Yseult Héjja-Brichard[a,b], Benoit R. Cottereau[a,b,*]

[a] Centre de Recherche Cerveau et Cognition, Université de Toulouse, 31052 Toulouse, France
[b] Centre National de la Recherche Scientifique, 31055 Toulouse, France

ARTICLE INFO

Keywords:
Computational neuroscience
Natural scene statistics
Binocular disparity
Binocular vision

ABSTRACT

The statistics of our environment impact not only our behavior, but also the selectivity and connectivity of the early sensory cortices. Over the last fifty years, powerful theories such as efficient coding, sparse coding, and the infomax principle have been proposed to explain the nature of this influence. Numerous computational and theoretical studies have since demonstrated solid, testable evidence in support of these theories, especially in the visual domain. However, most such work has concentrated on monocular, luminance-field descriptions of natural scenes, and studies that systematically focus on binocular processing of realistic visual input have only been conducted over the past two decades. In this review, we discuss the most recent of these binocular computational studies, with particular emphasis on disparity selectivity. We begin with a report of the relevant literature demonstrating concrete evidence for the relationship between natural disparity statistics, neural selectivity, and behavior. This is followed by a discussion of supervised and unsupervised computational studies. For each study, we include a description of the input data, theoretical principles employed in the models, and the contribution of the results in explaining biological data (neural and behavioral). In the discussion, we compare these models to the binocular energy model, and examine their application to the modelling of normal and abnormal development of vision. We conclude with a short description of what we believe are the most important limitations of the current state-of-the-art, and directions for future work which could address these shortcomings and enrich current and future models.

1. Introduction

More than half a century ago, Barlow (1961) postulated that the aim of the early visual cortex is to optimize information processing whilst using the fewest possible resources. Some of the most convincing support for this information-theoretic optimization theory comes from computational studies which showed that the nature of neural activity in the primary visual cortex could be attributed to encoding schemes which extract useful features (luminance, color, contrast, orientation, spatial frequency) from natural visual inputs. These optimisations are based on criteria such as optimal energy consumption (Olshausen & Field, 1996, 2005; Olshausen, 2003), information representation (Atick, 1992; Barlow & Földiák, 1989), bottom-up saliency-based signals (Zhaoping, 2000, 2006), and Bayesian optimisation of psychophysical and perceptual metrics (Knill & Richards, 1996).

Over the last two decades, numerous studies have explored how these optimization processes might be reflected in neural responses in the visual cortex (Olshausen & Field, 1997; Simoncelli & Olshausen, 2001), and how they might impact behavior (Burge & Jaini, 2017; Geisler, 2008). The overwhelming consensus is that natural statistics can, indeed, predict numerous properties of our visual system, from neural responses to behavior. However, a majority of these studies focus on monocular visual properties, whereas numerous species of the animal kingdom have two eyes and therefore experience their surrounding space through a binocular apparatus. In particular, species that have evolved in cluttered environments, like primates, are more likely to have developed a fronto-parallel ocular geometry with strong binocular overlap (Changizi & Shimojo, 2008; see also Langer & Mannan, 2012, for a computational analysis of binocular visibility under clutter). This visual geometry heightens their ability to sense binocular disparity – the small differences between the projections in the two retinae (Fig. 1). Binocular disparity processing is very important for numerous sensory, perceptual, and motor functions. One of the direct consequences of having access to disparity is the ability to estimate the depth of objects and planes in a scene (see e.g. Parker, 2007). In addition to depth estimation, binocular disparity is also known to drive vergence eye movements and accommodation (Masson et al., 1997), to play an important role in the execution of actions such as reaching, grasping and object manipulation (Melmoth & Grant, 2006; Servos & Goodale, 1994; Watt & Bradshaw, 2003), and to give

* Corresponding authors at: CNRS CERCO, UMR 5549, Pavillon Baudot CHU Purpan, BP 25202 31052 Toulouse Cedex, France.
E-mail address: research@tusharchauhan.com (T. Chauhan).

https://doi.org/10.1016/j.visres.2020.07.009
Received 30 December 2019; Received in revised form 19 July 2020; Accepted 20 July 2020
0042-6989/ © 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).

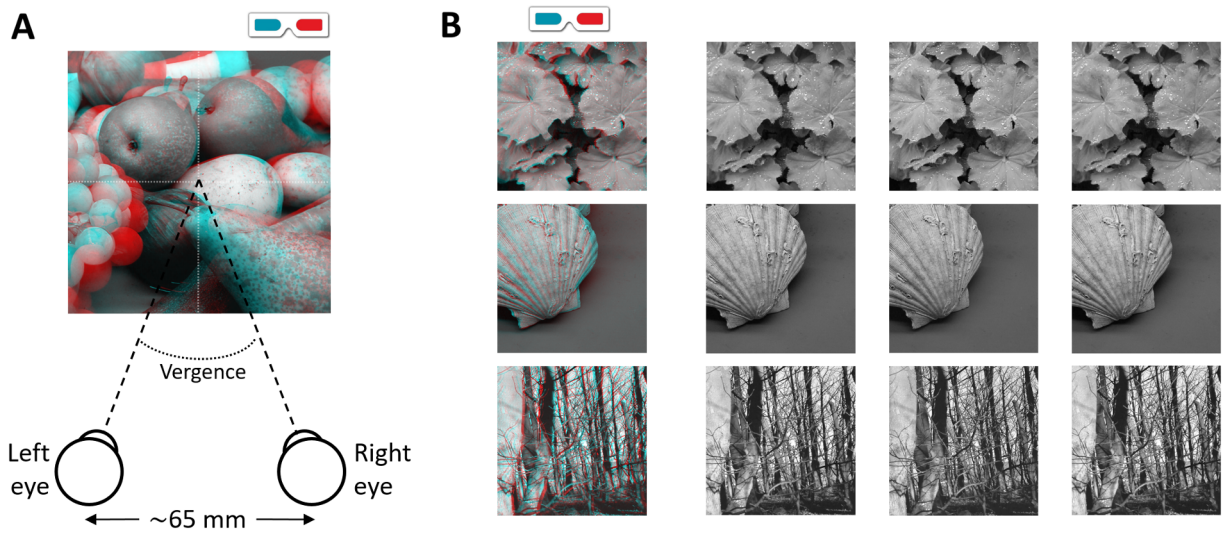

Fig. 1. Binocular vision and stereoscopic datasets. A. Retinal disparity. The two eyes, horizontally separated by about 65 mm on average, converge at the point of fixation. This causes the retinal projections of the same object to differ in the two eyes. B. Stereoscopic datasets. Three sample images (one per row) from the Hunter-Hibbard dataset (Hunter & Hibbard, 2015). The dataset was captured using two cameras which mimic the human visual geometry. The first column shows a red-cyan anaglyph of the scene. The following columns are arranged such that the second and third columns allow for uncrossed fusion, while the third and fourth columns allow for crossed fusion.

important cues for route-planning through a 3D environment (Hayhoe et al., 2009).

Despite the obvious benefits of binocular vision, the relationships that exist between statistics in natural scenes and visual processing and perception are less well understood for binocular than for monocular vision. Nonetheless, over the past two decades, numerous computational studies have tried to better characterize these relationships by proposing models that employ realistic biological constraints (energetic, information-theoretic, perceptual and behavioral) on the activity of neural populations in response to natural binocular stimuli.

The aim of this review is to describe and discuss the results of these studies. We place particular emphasis on computational studies that address binocular disparity, and try to position their results in the context of biological findings at various levels (neural, population coding, and behaviour). In order to do so, we first present properties of natural scenes when observed from a binocular visual system, and their relationship with neural selectivity and depth perception (part I). We then describe computational studies, supervised and unsupervised, which relate these properties to behavioural and neuroscientific measurements (part II). In the final section, we present a discussion which includes a comparison of these models with the binocular-energy model, their relevance and potential application to the study of the normal and abnormal development of vision, and a closer look at some of their limitations and how they can potentially be addressed in future studies (part III).

2. Statistics of binocular disparity in natural scenes, relationship with neural selectivity and depth perception

In this section, we describe how the distribution of binocular disparities in natural visual scenes has been estimated and how some parameters, in particular, are reflected in both cortical and behavioural measurements (2.1). We then describe how the statistics of binocular disparity show biases depending on where in the visual field they are sampled, and present studies which demonstrate a direct link between these biases, neural selectivity, and depth perception (2.2). Finally, we describe studies which have reported statistical relationships between disparity and numerous properties of natural scenes such as luminance, chromaticity or texture (2.3).

2.1. Range of binocular disparities in natural scenes

The first studies which considered the statistics of depth (3D) and binocular disparity in natural scenes were published about twenty years ago. Huang et al. (2000) analyzed depth statistics estimated from laser range data. They were able to show that 3D measures of a scene (such as range) offer a more informative description of its components, such as their structure or their spatial arrangement, as compared to '2D' measures such as luminance intensity, colour and texture. Using a similar range-based approach, Yang and Purves (2003) measured the actual distances from the image plane of all non-occluded points in a series of natural scenes. They found that the distribution of distances between the observer and surfaces in the range-data peaked at around 3 m, decaying exponentially at larger distances. They suggested that this distribution of physical distances in natural scenes could influence depth judgments under viewing conditions where little or no contextual information is available. Under these conditions, objects are typically perceived to be at a distance of 2–4 m, a phenomenon known as specific distance tendency (Gogel, 1965).

Starting from the work of Yang and Purves (2003), Hibbard (2007) attempted to address a major limitation of their study: the failure to account for eye position and therefore oculomotor behaviour, which is necessary to compute binocular disparities. They derived an estimation of the distribution of binocular disparities based on range images and showed a clear effect of fixation. The distribution was found to be broader in the periphery than in a central fixation area. Following Hibbard's study, Liu et al. (2008) further improved the computation of binocular disparities present in natural scenes by taking into account a known fixation behavior: during visual tasks, humans generally tend to fixate on objects relevant to the task. They found that the distribution of binocular disparity at eye level peaks at 0° (i.e., the left and right eye projections have the same retinal coordinates) and spans several degrees. Importantly, this range of disparities corresponds to the measured disparity tuning of neurons in macaque area MT (DeAngelis & Uka, 2003), and is fully within the operational range of human stereopsis determined psychophysically (Landers & Cormack, 1997; Prince & Rogers, 1998; Tyler, 1973).

More recently, Sprague et al. (2015) simultaneously measured binocular eye position and 3D scene geometry (from stereoscopic cameras) whilst observers performed various everyday tasks such as indoor

and outdoor navigation, social interaction, and near-work (making a sandwich). They computed the disparity distribution in their data and found it to be similar to measurements in the macaque V1 (Prince, Cumming, et al., 2002) – centered around 0 and biased towards near ranges (i.e., closer to the observer than the fixation point). Inspired by this study, Gibaldi, Canessa, and Sabatini (2017) designed a more controlled setup to accurately investigate the role of fixation. They used naturalistic 3D virtual scenes displayed in the peripersonal space of observers and recorded their eye fixation. The measured disparity distribution was then compared to the one obtained from random fixations of a virtual observer, and found to be closer to both neurophysiological data, and the range of disparity predicted by behavioural studies. This study also found the influence of an active fixation strategy to be more important at small eccentricities (central visual field) as compared to the periphery. This suggests that an accurate characterization of disparity statistics under natural viewing conditions should take the position in the visual field into account. In the next subsection, we describe studies that explored this relationship in more detail.

2.2. Statistical relationship between binocular disparity and position within the visual field

In this subsection, we first describe the relationship between binocular disparity distribution in natural scenes and elevation. Next, we explore the statistical properties of local change in disparity (i.e. disparity gradients). Finally, we examine its relationship with eccentricity. One of the first psychophysical studies to demonstrate the role of position in the visual field on the distribution of binocular disparities was conducted by Hibbard and Bouzit (2005). Predicting an effect of elevation on horizontal binocular disparity distribution, they experimentally tested their model by presenting observers with ambiguous stereograms for which binocular matches could result in both crossed and uncrossed disparities (thus, these stimuli could be interpreted as either closer or farther than the fixation cross). They found perceptual biases that were in agreement with their prediction: stimuli were perceived as closer when presented below the fixation point and farther away when above (see Fig. 2C). This suggests that the distribution of horizontal binocular disparities in visual scenes directly influences binocular matching.

A decade later, Sprague et al. (2015) showed that disparity tuning in the primary visual cortex reflects the relationship between horizontal binocular disparity and position within the receptive field. They conducted a *meta*-analysis encompassing five single-unit studies (820 neurons from the macaque V1) and computed the correlation between preferred disparity and receptive field (RF) location. They found that the neurons with RFs in the upper visual field tended to prefer uncrossed disparities, whereas neurons with RFs in the lower visual field preferred crossed disparities (see Fig. 2B). This neural bias was in good agreement with their estimation of binocular disparity distributions under natural viewing, where median values showed a gradient going from crossed disparities in the lower visual field (low elevation) to uncrossed disparities in the upper visual field (high elevation) (see Fig. 2A). In a similar analysis, this result was also confirmed by Gibaldi, Canessa, and Sabatini (2017). Nasr and Tootell (2018) extended our knowledge of this neural selectivity bias to the extrastriate cortex. They scanned human participants at a very high resolution (7 T) whilst showing them random dot stereograms (RDS) that were either in front (near stimuli) or behind (far stimuli) the fixation plane. By localizing horizontal disparity selective columns in areas V2 and V3, and comparing the upper (UVF) versus lower visual field (LVF) representations in these columns, they found that the fMRI signal (BOLD) was stronger for the near stimuli in the LVF representation, and for the far stimuli in the UVF representation. This suggests that disparity encoding in higher visual areas also reflects the biases in the natural statistics of binocular disparities. Interestingly, plausible evidence for a similar bias has been recently reported in the mouse cortex. La Chioma et al. (2019) used drifting vertical gratings and RDS stimuli to assess horizontal disparity tuning in three areas of the mouse cortex: primary visual area V1, rostrolateral area RL (mostly coding for the LVF), and lateromedial area LM (mostly coding for the UVF). They found that more neurons were tuned for crossed disparities in the RL compared to the two other regions. Their results also suggested an effect of elevation on disparity preference. In both V1 and RL, they found LVF-located cells to be, on average, more selective to crossed disparities than UVF-located cells.

The relationship between horizontal disparity and elevation in the visual field can lead to a number of interesting predictions. Here, we outline two such cases which have been studied. First, the relation between horizontal disparity and elevation could affect the empirical horopter, the locus in space that projects on retinal corresponding points where stereoacuity is the finest. Numerous studies have shown that, in humans, the shape of the vertical component of the horopter has a backward tilt, instead of being a vertical plane (E. Cooper et al., 2011; von Helmholtz, 1924; Tyler, 1991), and it has been suggested that this tilt could reflect the distribution of binocular disparities in natural scenes (Sprague et al., 2015). Cooper and Pettigrew (1979) indirectly estimated the tilt of the horopter in cat and owl by mapping the receptive field positions of binocular neurons at different elevations in the visual field. They showed that in these two species, where eye height is closer to the ground, the horopter was much more tilted than in humans, suggesting an adaptation of the visual system to the environment. Second, this relationship between horizontal disparity and elevation can also affect vergence eye movements. For instance, Gibaldi and Banks (2019) suggested that rapid binocular eye movements reflect the distribution of binocular disparities. By having their participants make saccades to eccentric targets on a screen with a 3D setup, they demonstrated that the eyes converged more in the lower visual field and diverged more in the upper visual field, thus reflecting the pattern of crossed/uncrossed disparities in the two hemifields.

The local variations of binocular disparity in natural scenes also have some interesting statistical properties and affect the perceived orientation of surfaces. In an analysis of the distribution of 3D orientations, Burge et al. (2016) found that tilts exhibit a strong cardinal bias: slants about the horizontal axes (tilt = 90°) are most probable, and slants about vertical axes (tilt = 0° and 180°) are the next most probable in the environment. Although they demonstrated that these biases strongly influence tilt estimates, however, the underlying neural mechanisms still remain to be revealed. For instance, single-cell recordings in the macaque caudal intraparietal area (CIP) showed that its neurons were selective to 3D orientations (slants and tilts) but had no biases toward specific values (Rosenberg et al., 2013).

As mentioned briefly above, under naturalistic viewing conditions, there is a relationship between horizontal disparity statistics and eccentricity in the visual field: binocular disparity distribution is broader in peripheral than in central vision (Hibbard, 2007). In addition, because the two eyes are separated along a horizontal and not a vertical axis, the range of vertical disparities in the foveal field of view is expected to be much smaller compared to horizontal disparities. Large vertical disparities are only projected on the retinae in the peripheral field of vision during oblique viewing (Read & Cumming, 2004). Relatively few electrophysiological and behavioural studies have addressed these predictions directly. Broader distributions of horizontal disparity in the periphery compared to the centre are in line with electrophysiological recordings in macaque V1 (Durand et al., 2007) and behavioral measurements of the upper disparity limit in humans (Ghahghaei et al., 2019). Durand et al. (2002, 2007) recorded disparity and orientation preference of V1 cells in macaque, both in the central and peripheral representation of the visual field. Their results revealed a reduced range of vertical disparity encoding in the central but not the peripheral visual field representation. Furthermore, they also found that horizontal and vertical disparities interact. Neurons with foveal receptive fields showed a preference for horizontal disparities, whereas neurons with peripheral receptive fields were found to respond robustly

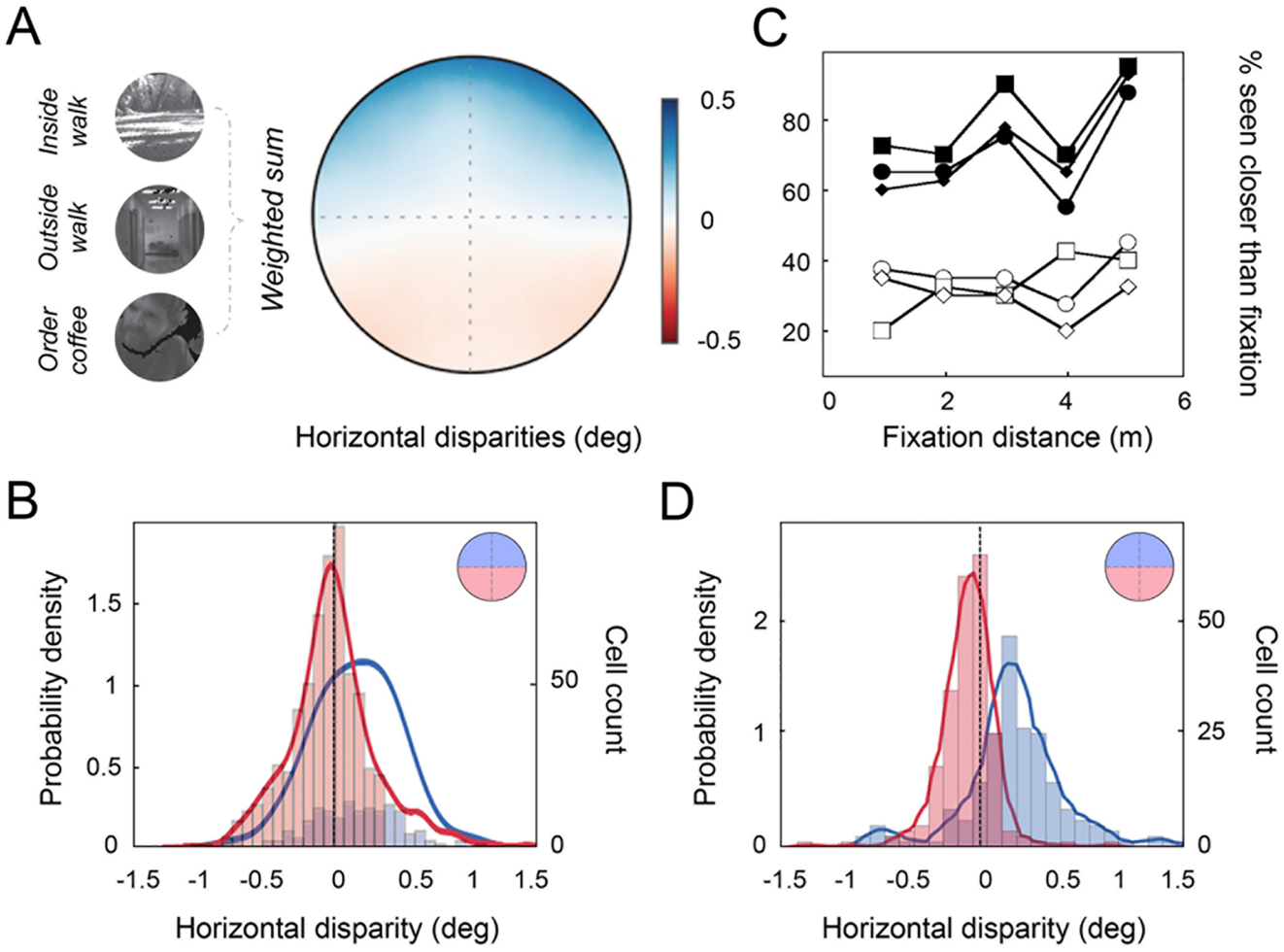

Fig. 2. Horizontal disparity statistics in natural scenes covary with elevation. A. Natural statistics of binocular disparity during everyday tasks. Typical images from the recordings are shown on the left. The rightward disk provides the median horizontal disparity values (weighted across the considered tasks) at different positions of the visual field (disk radius corresponds to 10°). Red and blue values respectively correspond to crossed and uncrossed binocular disparities. Figureadapted from Sprague et al. (2015). B. Electrophysiological measurements. Probability density of horizontal disparity found in a *meta*-sample of ~800 neurons in the macaque primary visual cortex. Neurons with receptive-fields (RFs) in the lower/upper hemifield (respectively in red and blue) are more likely to be most selective to uncrossed/crossed disparities. Figure adapted from Sprague et al. (2015). C. Psychophysical measurements. Hibbard and Bouzit (2005) used ambiguous stimuli that could be interpreted as both in front of, and farther away from fixation. Here, we present the data for one observer (PH), which shows that the stimuli were more likely to be interpreted as being in front of the fixation plane (crossed disparity) when presented in the lower hemifield (black items), and away from fixation (uncrossed disparity) when presented in the upper hemifield (white items). Diamonds, circles and squares respectively correspond to elevations of ± 33.5, ± 50.3, and ± 67 arcmin. Data from Hibbard and Bouzit (2005). D. Computational model. A dataset of natural stereoscopic images was used to train two spike-timing dependent plasticity (STDP) models. The first model (blue) was trained only on the upper visual field, while the second model (red) was trained on the lower visual field. The two solid lines show the distribution of horizontal disparities in the two populations. Notice the similarity with electrophysiological data in B. Figure adapted from Chauhan et al. (2018).

to both horizontal and vertical disparities. This peripheral treatment of vertical disparities was also supported by the subsequent findings from Sprague et al. (2015) and Gibaldi, Canessa, and Sabatini (2017). Both studies showed that preferred vertical disparity is close to zero in the central visual field, and increases with eccentricity along oblique directions. Gibaldi, Canessa, and Sabatini (2017) also suggested that vertical disparities are much less affected by the structure of the environment than horizontal disparities, as they found no significant difference between fixations made by human observers and random fixations.

2.3. Joint statistics of binocular disparity and other visual properties

In the environment, disparity is often correlated with other visual properties such as luminance, chromaticity, texture, orientation, and surface convexity. It is thus reasonable to assume that the joint processing of these visual features is likely to influence disparity estimation and depth perception. In this subsection, we present studies which describe these statistical correlations and their consequences at the neural and behavioural levels.

Potetz and Lee (2003) reported the joint statistics between the range and light intensity of outdoor visual scenes. They showed that although the mean intensity of luminance images tends to be invariant, the same could not be said for range images, for which the average range patch is vertically slanted. Looking at the covariance between luminance and range, they found a negative correlation between both variables, suggesting that brighter pixels tend to be closer to the observer. In a subsequent study (Potetz & Lee, 2006), they showed that this negative correlation was the result of shadows that are present in natural scenes. This relationship between binocular disparities and luminance is also reflected in macaque V1 neuron responses. Samonds, Potetz, and Lee (2012) estimated the luminance and disparity preferences of macaque V1 neurons and found a negative correlation: neurons that preferred light contrast were mostly near-tuned, whereas far-tuned neurons

tended to prefer dark contrast. Interestingly, by estimating the distribution of binocular disparities in natural scenes separately for light increments and decrements, Cooper and Norcia (2014) found differences that agree well with the negative correlation in V1 neuron preferences reported by Samonds, Potetz, and Lee (2012). In the same study, they further designed a psychophysical experiment to test whether human observers use this environmental prior (brighter is closer and darker is farther away). They manipulated luminance in natural images such that the stimuli either agreed (nearer is brighter) or disagreed (nearer is darker) with this prior. They found that observers judged images biased towards the environmental prior to have more depth, suggesting that humans exploit information about correlations between luminance and depth when estimating depth. This relationship between binocular disparity and luminance also holds for their variation across the visual field. For instance, Su et al. (2013) used color images of natural scenes with corresponding ground-truth range maps at a high-definition resolution to demonstrate a covariation between local changes of disparity and luminance.

In the same study (Su et al., 2013), the authors also found that binocular disparity covaries with chromaticity in the environment. They modelled the prior and conditional distributions of luminance, chrominance, and range with a Bayesian stereo algorithm and showed that the resulting binocular disparity maps were closer to the estimated distribution of binocular disparities when both luminance and chrominance were implemented in the algorithm rather than luminance alone. This finding might explain why chromaticity information was found to influence the solving of the stereo correspondence problem in behavioural studies (Jordan et al., 1990; Simmons & Kingdom, 1994). At the neural level, a functional neuroimaging study in non-human primate (Verhoef et al., 2015) revealed the existence of a partial overlap between brain areas responding to binocular disparity and those responding to color in the macaque inferior temporal cortex. EEG measurements in humans have also suggested that the depth illusion obtained from contrast of colour (chromostereopsis) might involve cortical areas that also respond to binocular disparity (Séverac Cauquil et al., 2009). We believe these studies could suggest a joint coding of colour and disparity cues by common neural populations. However, to our knowledge, this joint coding has never been investigated systematically at the neural level.

We saw above that in natural scenes, local changes of disparity and luminance covary. For continuous surfaces, these local changes are also correlated with texture orientation. This relationship might be exploited by the visual system to judge 3D orientation (tilts and slants). Indeed, estimating Bayes optimal values of tilt using three visual cues (disparity, luminance and dominant texture-orientation), Burge et al. (2016) showed that if disparity is the most reliable cue, the precision of the optimal estimate is significantly increased when all three cues are combined in a congruent manner. Interestingly, their results also showed that a linear combination of cues weighted by their relative reliabilities results in tilt estimates which are close to Bayes optimal estimates. Approximate tilt estimation could therefore be achieved by simple linear computations. Several behavioural studies have suggested that the visual system exploits this strategy (Hillis et al., 2004; Knill & Saunders, 2003). At the neural level, fMRI (Murphy et al., 2013) and electrophysiological recordings (Rosenberg & Angelaki, 2014; Sanada et al., 2012) highlighted different visual areas that could be involved in the representation of 3D surface orientation from different cues in the primate brain.

The ability of the visual system to take into account local variations in the relationship between different types of depth cues underlies an important feature of depth perception, namely, figure-ground segregation. There are different figure-ground cues such as convexity, size, or contrast, and a very effective way to detach a figure from its background is the combined use of disparity with a second figure-ground cue. Burge et al. (2010) showed that in a set of natural images, convexity and disparity are statistically correlated such that near regions are more likely to have convex contours. They further demonstrated that human observers exploit this correlation to judge depth separation between near and far regions. For a given disparity value, observers in their study tended to perceive more depth when nearer, occluding regions were convex than when they were concave. At the neural level, it has been shown that figure-ground relationships modulate responses from disparity selective neurons, with an increase in the response amplitude when the figure is nearer than the surround for some brain areas in the human visual cortex (Cottereau et al., 2011, 2012). A similar result has been reported in the macaque where responses of disparity-selective V2 neurons were found to be stronger for the near region of a figure when both disparity and figure-ground cues (contrast borders) were congruent (Qiu & von der Heydt, 2005). Despite these promising results, the neural underpinnings of the joint coding of disparity and convexity remain to be revealed.

3. Modelling population responses

As demonstrated in the previous section, there is overwhelming evidence to suggest that biases in disparity statistics are reflected in the characteristics of both neural populations and behavior. This leads to an important question: whether there exist theoretical and computational principles which can explain the representations of these statistics found in biology (e.g., disparity tuning curves, estimates of horopter, discrimination thresholds etc.). In this section, we explore recent developments in computational modelling which offer a deeper insight into various aspects of this relationship. To varying degrees, these models address the problem of disparity computation in the early visual system, and more crucially, offer plausible hypotheses about why and how these computational systems may emerge in the first place.

3.1. Theoretical background

Most models of neural encoding are framed as generative problems, where the goal of neural representations is to encode various properties of the input with the highest possible benefit. The benefit, in most models, is a trade-off between fidelity and efficiency. While fidelity offers the advantage that the neural population is able to represent the input statistics to a high degree of accuracy, thereby offering maximum possibility for the selection of an appropriate behaviour, efficiency ensures that the incurred energetic costs are as low as possible. As we will see, the exact formulation of these goals depends on the philosophical standpoint of a given model. In doing so, each model emphasises specific constraints and computational goals of the neural population it seeks to describe. To begin, we offer a very general description of this set of models using a single equation. This equation, framed from a neural-networks perspective, is intended to serve as an anchor-point as we go through the various models in subsequent sections (see Fig. 3A for a schematic). Given an input set $\boldsymbol{X}$, each model tries to address its fidelity-efficiency goals by identifying an optimal set of units with RFs ψ, whose connectivity is described by a set of parameters θ. In most cases, this is achieved by solving a minimization problem: $argmin_{\theta,\psi}\Phi$, where the objective function Φ often takes the form:

$$\Phi = F(X, A(\theta, \psi, X), A_{ext}) + \lambda S(X, A(\theta, \psi, X), A_{ext}) \quad (1)$$

Here, A is the activity of the network described by $\{\theta, \psi\}$ in response to the given input X and an external signal A_{ext}, F is the fidelity of the network with respect to the input, and S describes the efficiency constraints imposed on the network. An excellent treatment of this fidelity-efficiency dichotomy is presented by Zhaoping (2006) to explain saliency-driven representations in the visual system.

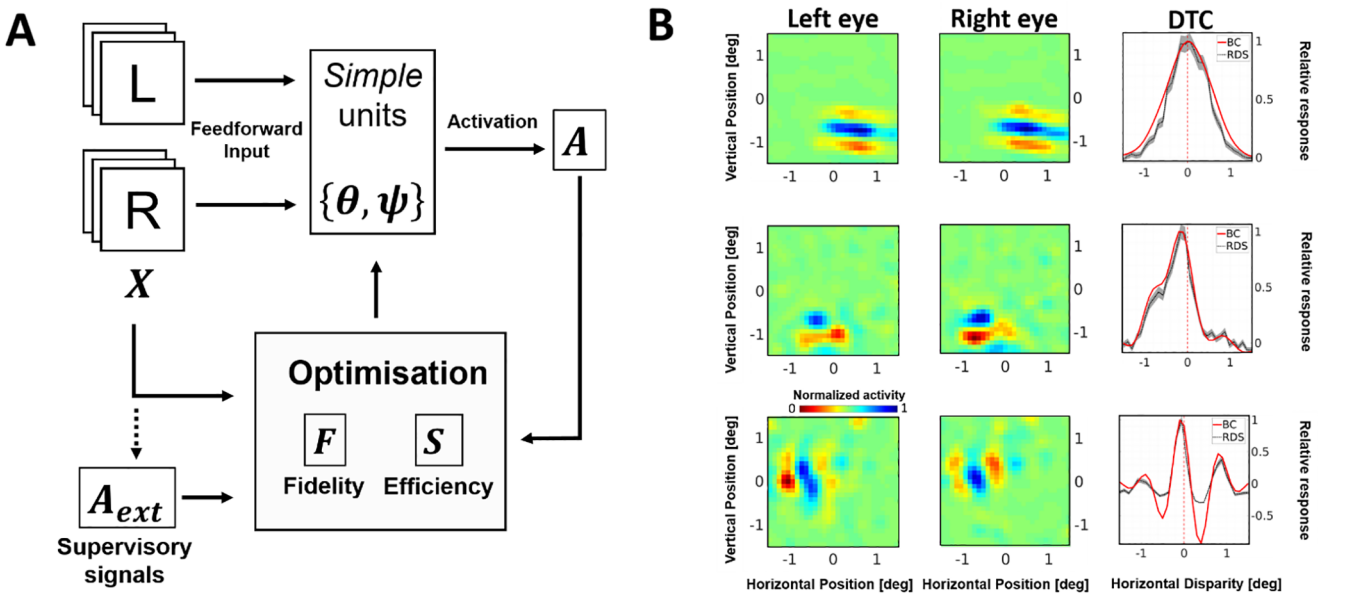

Fig. 3. Optimisation of simple-cell like units. A. A typical feedforward model. Simple units accept binocularly generated inputs which are usually preprocessed using smoothing and blurring operations resembling the processing in the retinogeniculate pathway. Depending on the model, the output activity of simple units is modelled using linear or nonlinear transfer functions. Most procedures, directly or indirectly, employ a balance of fidelity metrics such as reconstruction of the original signal or detection accuracy, and efficiency measures such as sparsity or constraints on the distribution of activations. In addition, supervised and reinforcement learning models also use supervisory signals which are either driven by the input (such as nominal disparity labels) or behaviour policies such as vergence minimization for fixation. The symbols used in the diagram correspond to Eq. (1) in the text. B. Receptive fields and disparity tuning curves. Receptive fields from three representative units (one unit per row) from an STDP-based feedforward model (Chauhan et al., 2018). The units show Gabor-like receptive fields in both eyes. The first neuron is tuned to zero disparity, the second neuron is tuned to small crossed (negative by convention) disparities, and the third neuron shows both position and phase tuning. The disparity tuning curves (DTCs) were estimated using binocular correlation (in red) and random-dot stereogram stimuli (grey).

Besides their position within the aforementioned fidelity-efficiency spectrum, various other schemes can be used to categorise models such as their computational architecture, their complexity, and their accuracy. However, here we choose a simpler and more biologically intuitive criterion to categorise models - supervision. Models which do not require labelled examples or extrinsic intervention from the experimenter during learning are classified as *unsupervised*, while models which require any form of external feedback are termed as *supervised*. Of course, there are some models which tend to employ both supervision and unsupervised learning. For the purposes of this review such hybrid models will be included with the supervised models.

3.2. Unsupervised models

The first set of models we consider are unsupervised. In terms of Eq. (1), this means the term A_{ext} is discounted. These models aim to show the direct impact of natural statistics on the selectivity of neural populations in the early visual cortex, and draw crucial support from the observation that feedforward connections between the lateral geniculate nucleus and the geniculo-recipient layers of the primary visual cortex are exclusively excitatory in nature. Specifically, we limit ourselves to models which describe disparity selectivity in populations of binocular neurons where the input signals originate from two, spatially proximal sensors (the two retinae). The spatial proximity of the sensors is crucial because it introduces correlations between the ocular signals which carry valuable information about the 3-D structure of the scene. The precise nature of these correlations is governed by the acquisition-geometry of the system (the interocular distance, the height of the ocular plane, the degree of orbital convergence etc.), and any given geometry emphasises some correlations over others. This phenomenon is not surprising and has already been described in the sections above using examples such as the cross/uncrossed disparity biases seen in the lower/upper hemifields in natural scenes (see Section 2.2).

Following the seminal work by Barlow (1961, 2001) who proposed a direct link between natural statistics and observed neural selectivity, a plethora of models investigating these links have been proposed. These include models which explain the structure of ON/OFF LGN receptive fields and colour-opponency through decorrelation analyses (Atick et al., 1992; Barlow & Földiák, 1989; Buchsbaum & Gottschalk, 1983), models which show how oriented edges may be the most appropriate filters for neural encoding of natural images (Bell & Sejnowski, 1997; Olshausen & Field, 1996), and models which demonstrate that early visual computations may represent bottom-up, saliency-driven data compression (Zhaoping, 2000, 2006). However, most of these models have focussed on monocular image statistics.

One of the first attempts to address how natural statistics shape binocular neural selectivity was made by Hoyer and Hyvärinen (2000). Their approach, similar in spirit to monocular studies by Bell and Sejnowski (1997) and van Hateren and van der Schaaf (1998), consisted of an initial linear decorrelation of the input, followed by an elimination of higher order correlations through Independent Component Analysis (ICA). They employed a now widely used algorithm for ICA computation (*fastICA*) based on an iterative estimation of input negentropy (Hyvärinen & Oja, 2000). In terms of Eq. (1), *fastICA* imposes no explicit constraints on the efficiency term S, and the fidelity F is implemented through a kurtotic optimisation on the individual filters to make their responses to elements of the input set X as statistically independent as possible. They used patches from a stereo-dataset of 12 natural images acquired using parallel cameras as input to their model. Upon convergence, all units in their studies showed oriented, edge-like RFs in either one or both eyes. Their ocular strength ratio (a measure of ocular dominance) ranged from highly monocular to highly binocular, peaking at an intermediate value. Interestingly, within binocular neurons, orientation and spatial frequency showed a close correspondence between the two eyes (only a qualitative report is provided). Using window-matching of the preferred stimuli for each binocular unit, they were able to identify disparity tuning curves which were tuned excitatory/inhibitory and near/far, similar to those reported in the

monkey visual cortex (Poggio et al., 1985, 1988).

Although highly informative, the study by Hoyer and Hyvärinen (2000) suffered from a crucial limitation - the sampling strategy employed to generate left and right input patches only simulated sampling at fixation. Since disparity statistics in natural scenes show systematic variations with eccentricity (see Section 2.2) this meant that the ICA analysis was performed only on foveally centred patches. Furthermore, the size of the input was only available in pixel values, which made quantitative comparisons with biological data quite approximate. Hunter and Hibbard (2015) addressed this by introducing an extremely important element to the analysis - a well calibrated dataset of natural stereo-images (Hibbard, 2008). They used a custom rig which allowed two calibrated cameras to be arranged in realistic acquisition geometries. They were mounted around 65 mm apart to approximate human interocular distance, and were capable of symmetric vergence (albeit at zero elevation). Neglecting contributions from cyclovergence (which is almost negligible in a symmetric, zero-elevation vergence geometry), this geometry generates retinal projections which are much more realistic than one would acquire using parallel cameras. This allowed the authors to make a thorough comparison of disparity tuning between ICA units and neurons in the primary visual cortex. To do this, they analysed the parameters of Gabor functions fitted to the converged RFs in the two eyes. As expected, the RFs in the two eyes were closely matched, and showed narrowband frequency and orientation tuning. Unlike Hoyer and Hyvärinen (2000), however, their results showed a bimodal distribution of ocular dominance, with neurons either being strongly monocular (25%) or binocular (75%). Furthermore, using the fitted Gabor centres and phases, they were able to report the position and phase disparity distributions for the ICA ensemble. Both horizontal and vertical disparities peaked at zero, with the distribution for horizontal disparities being broader than vertical disparities – both these observations are extensively supported by studies in the macaque V1 area (e.g. see Prince, Cumming, et al., 2002). Curiously, they also report a bimodal distribution of phase disparities with modes at zero and π, i.e., the left and right receptive fields of most neurons in their population were either in-phase or out-of-phase - something that is not observed in real recordings where less than 20% of neurons are tuned inhibitory (see e.g. DeAngelis et al., 1991, or Prince, Cumming, et al., 2002). In a later study (Hunter & Hibbard, 2016), they extended their approach to model representative complex cells by combining output from ICA units using the binocular energy model (BEM) – thus showing that the output of ICA units can drive disparity selective complex cells as well. However, in both studies they note that while ICA can indeed predict a realistic encoding of disparity, this encoding can only partially explain what is observed in real neuronal populations in the primary visual cortex. In a third study (Hunter & Hibbard, 2018), they explored how position in the receptive field can influence the distribution of disparity selectivity in ICA units. They found that ICA ensembles can reproduce many known biases such as an increase in disparity tuning with eccentricity, broader tuning for horizontal compared to vertical disparities, and a preference for crossed or uncrossed disparities depending on whether the receptive field was centred in the lower or upper hemifield (see Section 2.2).

As noted earlier, one of the factors limiting how well ICA-based models explain biological data may be an emphasis on global over local correlations. In biological systems, plasticity is dominated by mechanisms which operate over local synaptic topologies. Recently, Chauhan et al. (2018) proposed a rank-based binocular spiking model to explain how natural disparity statistics may drive the emergence of simple-cell like RFs (Fig. 3B). Their model consisted of an initial decorrelation stage using difference-of-Gaussian filters, followed by a neural network endowed with an abstract spike-timing dependent plasticity (STDP) rule and winner-take-all inhibition. The local rank-based plasticity rule tunes the network to detect the most frequently occurring features in the dataset, and the winner-take-all mechanism enforces sparseness in the converged ensemble. When trained on the same dataset as Hunter and Hibbard (2015) they were able to demonstrate that in addition to realistic binocular RFs, the model showed characteristic sub-optimalities associated with early visual neurons such as symmetrical, and consequently broadly tuned, RFs (Ringach, 2002). Contrary to the ICA model, the units in this model showed a bias for zero phase disparity, which concurs with reports in a number of species such as the macaque (Prince, Cumming, et al., 2002), cat (Anzai et al., 1999; DeAngelis et al., 1991), and the barn-owl (Nieder & Wagner, 2000). Their model was also able to predict biases such as the broadening of population disparity tuning with eccentricity, and the correlation between elevation and disparity (see Fig. 2D and Section 2.2). Furthermore, using a second dataset which was collected using parallel cameras, they showed that learning of biases in naturalistic datasets is not sufficient to predict neural responses to disparity unless a realistic acquisition geometry is also taken into account. While closer to biological data, this model still suffered from a number of limitations such as the lack of retinotopy and inhibitory connections, and the inability to address the emergence of disparity selective complex cells.

Together, these studies show how realistic constraints on data acquisition, information transfer and the formulation of learning rules can lead to units which can predict disparity responses in the early visual cortex. While they are able to address properties at a single-cell level such as disparity tuning curves of tuned excitatory/inhibitory and near/far neurons, their main strength lies in the modelling of population-level characteristics such as the ocular dominance continuum (Prince, Cumming, et al., 2002) and the distributions of position and phase disparity (Anzai et al., 1999; DeAngelis et al., 1991; Nieder & Wagner, 2000; Prince, Cumming, et al., 2002). Furthermore, units predicted by these models are closer to simple-cells. The highly nonlinear nature of excitatory and inhibitory interactions between retinogeniculate inputs, simple-cells, and complex-cells makes it difficult to formulate unsupervised models that can explain the emergence of complex-cells with similar elegance.

Finally, any unsupervised approach is based on the inherent assumption that selectivity emerges primarily from the properties of the input. While this may be partially true for the first few geniculo-recipient synapses in layer IV-C, factors such as feedback from proximal layers and corticocortical inhibition make it less likely that this holds for most neurons beyond the very early sensory populations. Since any neural specialisation must, directly or indirectly, support evolutionarily meaningful behaviour (Barlow, 1961), it is likely that neural selectivity in these populations is also shaped by the affordances of behaviour. A second class of models which attempts to address this relationship between encoding and behaviour is described in the next section.

3.3. Supervised models

Supervised models rely on labelled information to learn specific tasks such as detection and discrimination. The term A_{ext} in Eq. (1) is no longer neglected, and during training, is usually a function of the input; i.e., $A_{ext} = A_{ext}(X)$. This allows the inclusion of signals which provide explicit feedback about the model's response to stimulus features such as nominal class-labels, or more complicated functions such as correlates of oculomotor behaviour and grasping. In this section we will specifically concentrate on supervised models which investigate disparity selectivity through the use of natural and naturalistic binocular stimuli.

One of the first models which attempted to make disparity estimations using natural stimuli was proposed by Gray et al. (1998). The model was based on a mixture-of-experts architecture (Jacobs et al., 1991) consisting of separate local disparity, and global gating modules. The input to the network consisted of responses of disparity-energy filters applied to both synthetic (occluded shapes, RDS stimuli) and natural 1-D line-stimuli. The local disparity modules made local binocular energy calculations at various frequencies, while the gating module selected the appropriate combination of disparity modules for

any given stimulus. The model did not impose any direct efficiency constraints (S in Eq. (1)), and the weights were adjusted so as to optimally classify the input disparity (i.e., $A_{ext}(X)$ signal represented an error in the prediction of the class label in the output layer). They showed that such a model can make reliable estimates of disparity even under conditions of transparency and occlusion, while displaying characteristic traits such as stereo-hyperacuity and the ability to predict the effects of low- and band-pass filtering of line targets on disparity discrimination thresholds (Westheimer & McKee, 1980).

Okajima (2004) used an infomax network to investigate the problem of phase versus position encoding of disparity. The model maximised the mutual information between the input class and the network response under a low SNR assumption, and imposed no explicit constraints on the efficiency term S. The network was trained on disparity-labelled Gaussian noise patterns and natural stimuli which were pre-processed using difference-of-Gaussian filters. Analysis of the parameters of Gabor functions fitted to the converged RFs revealed that horizontal disparity was coded by both position and phase. In agreement with experimental observations, it was able to predict a decrease in phase disparity with spatial frequency (Anzai et al., 1999). However, it also predicted a decrease in position disparity with frequency which is not observed in the data. Okajima (2004) also proposed the interesting possibility that the 'supervision' in real neuronal assemblies could take the form of local temporal labelling where inputs within short time-windows are considered as belonging to the same class. Though quite approximate, we believe this is close to the temporal coding idiom of biologically observed mechanisms where locally precise temporal coding modulates synaptic strengthening and, in some cases, weakening.

The two aforementioned models exploited the disparity statistics of natural stimuli only to a limited extent. Burge and Geisler (2014) proposed a supervised scheme which used 1-D line-signals derived exclusively from binocular projections of monocular natural images. They used a Bayesian task-specific optimisation based on accuracy maximization analysis (AMA) (Geisler et al., 2009), to construct a set of filters optimised for disparity detection. Although the sparsity S is not directly constrained, AMA optimisation also models scaled additive noise within individual filters (Burge & Jaini, 2017), which can affect encoding sparsity. The filters were found to possess properties which resemble simple-cells, such as similar preferred frequencies between the two eyes, and a spatial frequency bandwidth of ~1.5 octaves. Like the ICA-based unsupervised models, the final filter-bank also included RFs which consisted of anti-phase filters in the two eyes. Interestingly, since the co-occurrence of dark and bright edges at the same retinal coordinates in the two eyes is a relatively rare occurrence in natural scenes, these units were interpreted as providing information about the stimulus disparity by not responding (see Read & Cumming, 2007, for a discussion of how such neurons can account for responses to anti-correlated random dot stereograms). Considering the goal of the AMA optimisation was to increase the accuracy of disparity-label classification, we believe this suggests that accurate disparity decoding necessitates an encoding ensemble comprising binocular cells with both correlated and anti-correlated RFs. To show how the AMA responses could be used to decode disparity in novel inputs, the filtering was followed by a Bayesian optimal, maximum-a posteriori (MAP) decoder. The MAP decoder was found, to a qualitative agreement, to predict a number of psychophysical results such as the exponential decay in thresholds with an increase in disparity (McKee et al., 1990; Stevenson et al., 1992), and the patterns of sign-confusion for small disparities (Landers & Cormack, 1997). Notably, they were able to show that this decoder can be implemented by operations resembling the binocular energy model (see Section 4.1).

In a more recent study, Goncalves and Welchman (2017) delved deeper into the question of the aforementioned non-responding units. They trained a binocular 3-layered CNN consisting of a convolutional ReLU layer (called *simple units*), followed by a max-pooling layer, and finally a softmax output layer (called *complex units*), by back-propagating errors in the classification of stimuli as near or far. The training stimuli were generated by projecting a dataset of luminance-field images of natural scenes (using the accompanying depth-map) on to various depth-planes and simulating disparity by horizontally shifting the projected image. By allowing both positive (excitatory) and negative (inhibitory) weights in their network, they were able to show that a complex unit trained to detect a given disparity developed stronger connections (both excitatory and inhibitory) with *simple units* which responded to similar disparities. Crucially, the connections were strongly excitatory when the left and right RFs of the *simple unit* were correlated, and strongly inhibitory when they were anti-correlated. Through this model, they were able to predict the attenuated responses for anti-correlated RDS stimuli (compared to correlated RDS) recorded in complex cells of the macaque (Cumming & Parker, 1997; Ohzawa et al., 1990; Samonds, Potetz, Tyler, & Lee, 2013). This suggests a more important role for corticocortical inhibition in disparity selectivity (e.g., see Read & Cumming, 2007, for an interesting phenomenological model of phase-disparity selective 'lie detector' neurons).

The supervised models covered so far use categorical disparity labels under a strict classification paradigm. Under this paradigm, supervision is either interpreted as a task-specific feedback signal (Burge & Geisler, 2014) delivered at the end of each learning step, or a form of temporal, localised labelling (Okajima, 2004). However, another plausible source of such supervisory signals could simply be reactive cortical feedback pertaining to time-continuous sensorimotor demands. These demands, in turn, may either be goal-oriented (such as grasping, haptic affordances) or volitional (such as vergence eye-movements, accommodation). In these cases, the input to the model interacts continuously with its output (active sequential learning), and supervisory signals are evaluative, as opposed to purely instructive – thus making them better suited to a reinforcement learning framework. In fact, hybrid models which explicitly address this point of view by combining the learning of disparity with intrinsic supervisory feedback, are being increasingly used in robotics and computer vision (Gibaldi, Canessa, Solari, & Sabatini, 2015; Konda & Memisevic, 2014; Lelais et al., 2019). Although more directly applicable in the context of adaptive robotics, these models offer valuable insights into how motor behaviour can interact with disparity encoding.

Here, we present one of the first such studies by Zhao et al. (2012) which specifically demonstrated how vergence control and efficient disparity encoding can be learnt simultaneously by combining efficient coding and reinforcement learning. They used translational shifts at various disparities (say d_{input}) to generate a binocular dataset from a database of natural monocular images. The input to the model was then generated by displaying patches from randomly selected stereo-images in blocks of 10 frames. For any given block (the image did not change within the 10-frame block), fixation was simulated by sampling patches from random locations within the image. The patches were not exactly centre-matched, and the distance between their centres was thus used as a measure of vergence (say v). In this scenario, the binocular fixation would be maintained when the retinal disparity ($d_{retinal} = d_{input} - v$) would be zero. The model was divided into two stages. The first, unsupervised stage of the model computed a convolutional, sparse dictionary (simple units) using a two-stage process similar to Olshausen and Field (1996). The activity of the simple units was pooled using a squared nonlinearity to generate complex unit activations. This was followed by a second stage of reinforcement learning which used a modified natural actor-critic algorithm (Bhatnagar et al., 2009) to determine vergence behaviour policies. Running the first stage of the model resulted in Gabor-like simple units which were disparity selective. Interestingly, the most active units were tuned excitatory, while the least active units were either tuned inhibitory or near/far tuned. Running the second stage of the model using converged simple units (from the first stage) resulted in the development of vergence behaviour which strongly reflected the past exposure of the simple units.

However, when both the first and second stage of the model were run simultaneously, both optimum realistic simple units, and optimum vergence behaviour were learnt such that when presented with an input at a given disparity d_{input}, the model robustly adjusted its vergence behaviour to maintain fixation (i.e., $d_{retinal} = d_{input}$). This study shows that not only is the joint learning of oculomotor and visual features highly effective, but that one may facilitate the other. In subsequent work, the robustness of this model was further verified and then demonstrated using a robotic system (Lonini, Forestier, et al., 2013; Lonini, Zhao, et al., 2013).

Together, supervised models demonstrate how the inclusion of feedback signals can enrich the interpretations of system level models. Contrary to what one might expect, these models do not diminish the role of fundamental information-theoretic principles such as the fidelity and efficiency of the resulting encoding which form the core of bottom-up, unsupervised models. Rather, through supervision, they add a behavioural, top-down context to how the early visual system may extract useful features from natural stimuli. Certain testable predictions about disparity selective neural populations can already be made with the current models. The most notable amongst these is the inhibitory influence of non-responding units with anti-correlated RFs in the two eyes, which have now been predicted by multiple studies (supervised models: Burge & Geisler, 2014; Goncalves & Welchman, 2017; unsupervised model: Hunter & Hibbard, 2016). Although such units have been reported in the literature, current reports estimate their population to be far below the model predictions. Indeed, tuned inhibitory cells represent about 15 percent of the disparity selective neurons recorded in cats and macaques (DeAngelis et al., 1991; Poggio et al., 1988; Prince, Pointon, et al., 2002) whereas the modeling studies mentioned above found around 35–40 percent of neurons to have anti-correlated receptive fields. Addressing this discrepancy is important (see Read & Cumming, 2017) and presents a real, and feasible challenge to both computational and experimental neuroscientists.

4. Discussion

4.1. Comparison with the binocular energy model

A phenomenological model which has had considerable success in explaining numerous characteristics of complex-cells (most notably, their responses to random-dot stereograms) is the binocular energy model (BEM) (Fleet et al., 1996; Haefner & Cumming, 2008; Lippert & Wagner, 2001; Ohzawa et al., 1990; Read et al., 2003). BEM, proposed by Ohzawa et al. (1990), derives from a set of spatiotemporal energy models first proposed to explain motion detection (Adelson & Bergen, 1985). It typically involves an initial linear filtering stage which models simple-cell responses, followed by a nonlinear combination. Outputs from quadrature sets of simple-cell filters are then summed to obtain complex-cell responses.

Here, we draw attention to two studies which, within the framework of natural statistics-driven modelling, were able to draw interesting conclusions regarding the interactions between simple and complex cells predicted by BEM. Hibbard (2008) applied the BEM to natural stereoscopic images and found that while qualitative trends such as an increase in the range of encoded disparity with eccentricity can be predicted, BEM is not able to provide accurate quantitative predictions about neural tuning based on natural disparity statistics. Using a Bayesian inference paradigm, Burge and Geisler (2014) showed that if simple units are optimised for disparity detection, their responses show an approximately Gaussian distribution, thus allowing for the derivation of a Bayesian-optimal decoder which has a quadratic form similar to a BEM unit. Both these studies suggest that BEM, while originally proposed as a purely mechanistic model to explain complex-cell responses, remains, up to some degree, compatible with the natural statistics of disparity.

However, it must be noted that there are numerous known criticisms of the BEM which are also valid for models of disparity selective complex cells based on natural statistics. All these approaches are based on hierarchical cascades of computation and are therefore unable to satisfactorily explain the role of recurrent and inhibitory connections in real recordings. Indeed, in the cortex, synaptic connections to disparity-selective complex cells are unlikely to be purely feedforward, and include lateral interactions between complex cells, intra- and interlaminar inhibition mediated by interneurons with vastly differing spatiotemporal properties, and direct thalamic inputs to some complex cells in L2/3, L5 and L6 (see, e.g., Bardy et al., 2006; Ferster & Lindström, 1983; Hoffmann & Stone, 1971; Livingstone & Tsao, 1999; Malpeli, 1983; McGuire et al., 1984; Tanaka, 1985, for an interesting overview of the debate over the years). Models based on recurrent connectivity and intradendritic activity have shown that it is important to consider the dynamics introduced by such non-hierarchical interactions (Archie & Mel, 2000; Chance et al., 1999; Samonds, Potetz, Tyler, & Lee, 2013; Tao et al., 2004), and a concrete theory about constraints which drive the structure and function of the complex-cell circuitry still remains elusive.

4.2. Can current computational models explain binocular disparity selectivity development and/or refinement through visual experience?

A very interesting, and perhaps also provocative claim that can be made by experience driven computational models of the early visual system is that in addition to neural selectivity, they may also be able to address plasticity during the critical period. Over the past decade, several studies (Hsu & Dayan, 2007; Hunt et al., 2013; Klimmasch et al., 2018; Saxe et al., 2011) have shown that unsupervised models trained with modified inputs can reproduce what is observed in animal models trained under abnormal rearing conditions (Freeman & Pettigrew, 1973; Wiesel & Hubel, 1963). For example, Hunt et al. (2013) used three different generative models and showed that all of them captured the changes of binocular selectivity observed in kittens reared under six different rearing conditions. Notably, they showed that asymmetries in inter-ocular correlation across orientations led to orientation-specific binocular receptive fields. More recently, Cloherty et al. (2016) used a computational approach based on Hebbian plasticity to predict how rearing animals with visual inputs biased towards vertical orientations in one eye and horizontal orientations in the other eye (cross-rearing) could change the spatial relationship between pinwheel and ocular dominance regions. These predictions were subsequently verified in cats reared under similar conditions. In one of our previous studies (Chauhan et al., 2018), we proposed that a model based on STDP could capture the progressive development of binocular disparity selectivity in early visual cortex (see e.g. movie 1 in this publication).

Due to their ability to simulate abnormal viewing conditions, these computational approaches could also constitute an interesting tool to better understand developmental pathologies such as amblyopia which are associated with numerous deficits in binocular functions (see, e.g., Levi et al., 2015, for a detailed amblyopia-specific review). Indeed, studies based on unsupervised learning have shown that neural ensembles trained on visual inputs that are randomized between the two eyes do not develop selectivity to binocular inputs. Instead, such stimuli lead to mostly monocular RFs which do not respond well to binocular disparity (Chauhan et al., 2018; Hunter & Hibbard, 2015).

Are the mechanisms described above enough to fully characterize the development and refinement of binocular disparity selectivity in early visual cortex? In numerous species, receptive fields at birth already show some preliminary forms of responsiveness to visual features such as orientation and spatial frequency (see, e.g., Wiesel & Hubel, 1974). For binocular disparity, despite the fact that selectivity undergoes some critical refinement during early life (Freeman & Pettigrew, 1973; Norcia et al., 2017; Pettigrew et al., 1973; Pettigrew, 1974; Tao et al., 2014), it was shown that an initial form of binocular correlation exists in young macaque monkeys as early as the sixth postnatal day

(Chino et al., 1997). In humans, it was recently found that binocular disparity could be used to trigger vergence eye movements in 5- to 10-week old infants (Seemiller et al., 2018). Although such studies do not exclude the possibility that disparity selectivity is acquired through visual experience in the very first moments of life (see, e.g., Li et al., 2006, who demonstrated that motion direction selectivity in the ferret is not present at eye opening but can develop within a few hours), they suggest that more comprehensive models of binocular disparity development should take into account prenatal processes such as those triggered by retinal waves (Ackman et al., 2012). Interestingly, previous work has shown that unsupervised models such as those described in this review could also capture prenatal mechanisms of synaptic refinement (Albert et al., 2008; Butts et al., 2007). We believe that combining such innate developmental mechanisms with experience driven learning could lead to models which better characterise both normal and abnormal development of binocular disparity selectivity.

4.3. Perspectives for computational modelling of disparity selectivity

In the preceding sections we have remarked on some of the limitations of current computational models which address disparity processing using natural statistics. Here, we briefly comment on two additional shortcomings, and how we believe they could be addressed. The first shortcoming is not related to computation, but the availability of datasets which realistically approximate retinal input. When one takes into account the various degrees of freedom of movement (orbital movement of the eyes within their sockets, and the movement of the head), and the curvature of the human retinae, the human visual geometry is indeed complicated. Consequently, ecologically valid datasets which closely replicate retinal input are very challenging to collect. Here, we note some of the more comprehensive datasets available in the public domain.

Hibbard (2008) used two cameras with realistic fixations restricted to a straight-ahead, zero-elevation plane to collect a relatively large dataset of indoor and outdoor scenes (about 120 images in total). In an even more realistic acquisition, Sprague et al. (2015) used head-mounted cameras and an eye-tracking system to collect not only a binocular video dataset, but also eye-fixation data. This was supplemented by a projective model which translated the dataset to realistic retinal coordinates. While suitable for unsupervised learning of binocular disparity, both the aforementioned studies lacked distance-specific labelling which may be required for ground truth labelling and supervised algorithms. Adams et al. (2016), in a very different approach, collected LIDAR range-data and high dynamic range spherical imagery from locations sampling 25 indoor and outdoor categories. This dataset allowed for distance- labelling of pixels using a single centre-of-projection. In an even more accurate LIDAR dataset, Burge et al. (2016) co-registered LIDAR images with independent centres-of-projection for the two cameras – thus making it possible to accurately distance-label each pixel from each camera. However, both the LIDAR datasets lack eye-fixation data which could be useful for models which require precise retinal projections such as those exploring binocular saliency maps.

Of these, only the first dataset has, as yet, been substantially exploited by binocular computational models. Most datasets used in current studies are either generated artificially by pixel shifting, or acquired using unrealistic camera geometries which do not reflect realistic retinal acquisition. While this allows the input data to be highly curated (which is especially useful for supervised learning) it also limits the comparative power of the models with respect to characterising real neuronal populations. Future work towards the collection of realistic stereo-datasets using both traditional stereo-camera rigs and light-range and LIDAR imaging, and the development of realistic retinal projection models which can interpret these datasets, could greatly boost the quality of inputs used in the computational modelling of binocular vision (see, e.g., Ehinger et al., 2017; Iyer & Burge, 2018). Furthermore, it is important to make such resources available in the public domain so that they can be used to compare models on an equal footing.

A second limitation of the current models is the lack of dynamics. Most of the approaches described in this review were based on natural stereoscopic images whereas our visual environment is dynamic – both because the objects in the surrounding space are moving, and because we are moving (our eyes, our head and our body). Thus, it seems very important for future computational models of stereoscopic vision to take this temporal aspect into account. Some of the motion properties in natural scenes are statistically correlated with binocular disparity, and therefore directly relevant for depth perception (see Section 2.3 for static visual properties that are correlated with binocular disparity in natural scenes). For example, motion parallax is a powerful depth cue (Rogers & Graham, 1979) based on velocity gradient that was proposed to be jointly coded with binocular disparity in macaque area MT in order to extract the 3D structure of the scene (Kim et al., 2015; Nadler et al., 2013). The same type of co-occurrence exists between binocular disparity and optic flow (Ito & Shibata, 2005), and could be used by our nervous system during navigation (Cardin & Smith, 2011). By training on monocular natural videos, unsupervised models based on ICA (van Hateren & Ruderman, 1998) and sparse coding (Olshausen, 2003) have reported converged, simple-cell like neural populations which show realistic spatiotemporal tuning and are selective to motion direction. Future studies should build on these approaches to derive models that are able to capture the statistical correlation that exist between binocular disparity and motion properties in dynamic natural scenes. In fact, joint-coding models spanning multiple domains (including luminance, contrast, colour) could perhaps provide a more realistic description of the early visual system, its relationship with behaviour, and the part that natural statistics play in shaping them both.

5. Conclusions

In this review, we described and discussed recent studies that characterise how binocular disparity statistics in natural scenes can influence neural responses in early visual cortex. We presented different computational approaches that permit to better understand how the underlying mechanisms emerge, possibly through visual experience during development. Finally, we compared these computational approaches to more classical models of binocular disparity selectivity and proposed directions for future studies in this field of research.

Acknowledgments

We thank Suzanne McKee, Jean-Baptiste Durand and Yves Trotter for their valuable comments on the manuscript. This research was supported by a grant from the ‘Agence National de la Recherche’ (ANR-16-CE37-0002-01, ANR JCJC 3D3M) awarded to Benoit R. Cottereau and a grant from the ‘Fondation pour la Recherche Médicale’ (Grant FRM: SPF20170938752) awarded to Tushar Chauhan.

References

Ackman, J. B., Burbridge, T. J., & Crair, M. C. (2012). Retinal waves coordinate patterned activity throughout the developing visual system. *Nature, 490*(7419), 219–225. https://doi.org/10.1038/nature11529.

Adams, W., Elder, J., Graf, E., Leyland, J., Lugtigheid, A., & Muryy, A. (2016). The southampton-york natural scenes (SYNS) dataset: statistics of surface attitude. *Scientific Reports, 6*, 35805. https://doi.org/10.1038/srep35805.

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *JOSA A, 2*(2), 284–299. https://doi.org/10.1364/JOSAA.2.000284.

Albert, M., Schnabel, A., & Field, D. (2008). Innate visual learning through spontaneous activity patterns. *PLoS Computational Biology, 4*(8), Article e1000137. https://doi.org/10.1371/journal.pcbi.1000137.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999). Neural mechanisms for encoding binocular disparity: Receptive field position versus phase. *Journal of Neurophysiology, 82*(2), 874–890.

Archie, K., & Mel, B. (2000). A model for intradendritic computation of binocular

disparity. *Nature Neuroscience, 3*(1), 54–63. https://doi.org/10.1038/71125.
Atick, J. (1992). Could information theory provide an ecological theory of sensory processing? *Network: Computation in Neural Systems, 3*(2), 213–251. https://doi.org/10.1088/0954-898X_3_2_009.
Atick, J., Li, Z., & Redlich, N. (1992). Understanding retinal color coding from first principles. *Neural Computation, 4*(4), 559–572. https://doi.org/10.1162/neco.1992.4.4.559.
Bardy, C., Huang, J. Y., Wang, C., FitzGibbon, T., & Dreher, B. (2006). 'Simplification' of responses of complex cells in cat striate cortex: Suppressive surrounds and 'feedback' inactivation. *The Journal of Physiology, 574*(3), 731–750. https://doi.org/10.1113/jphysiol.2006.110320.
Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. *Sensory Communication* (pp. 217–234). The MIT Press. https://doi.org/10.7551/mitpress/9780262518420.003.0013.
Barlow, H. (2001). The exploitation of regularities in the environment by the brain. *Behavioral and Brain Sciences, 24*(04), 602–607. https://doi.org/10.1017/S0140525X01000024.
Barlow, H., & Földiák, P. (1989). Adaptation and decorrelation in the cortex. In R. Durbin, C. Miall, & G. Mitchison (Eds.). *The computing neuron* (pp. 54–72). Inc: Addison-Wesley Longman Publishing Co.
Bell, A., & Sejnowski, T. (1997). The "independent components" of natural scenes are edge filters. *Vision Research, 37*(23), 3327–3338. https://doi.org/10.1016/S0042-6989(97)00121-1.
Bhatnagar, S., Sutton, R. S., Ghavamzadeh, M., & Lee, M. (2009). Natural actor–critic algorithms. *Automatica, 45*(11), 2471–2482. https://doi.org/10.1016/j.automatica.2009.07.008.
Buchsbaum, G., & Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society of London Series B Biological Sciences, 220*(1218), 89–113. https://doi.org/10.1098/rspb.1983.0090.
Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *Journal of Neuroscience, 30*(21), 7269–7280. https://doi.org/10.1523/JNEUROSCI.5551-09.2010.
Burge, J., & Geisler, W. (2014). Optimal disparity estimation in natural stereo images. *Journal of Vision, 14*(2), 1. https://doi.org/10.1167/14.2.1.
Burge, J., & Jaini, P. (2017). Accuracy maximization analysis for sensory-perceptual tasks: computational improvements, filter robustness, and coding advantages for scaled additive noise. *PLOS Computational Biology, 13*(2), Article e1005281. https://doi.org/10.1371/journal.pcbi.1005281.
Burge, J., McCann, B., & Geisler, W. (2016). Estimating 3D tilt from local image cues in natural scenes. 2 2 *Journal of Vision, 16*(13), https://doi.org/10.1167/16.13.2.
Butts, D., Kanold, P., & Shatz, C. (2007). A burst-based, "Hebbian" learning rule at retinogeniculate synapses links retinal waves to activity-dependent refinement. *PLoS Biology, 5*(3), Article e61. https://doi.org/10.1371/journal.pbio.0050061.
Cardin, V., & Smith, A. (2011). Sensitivity of human visual cortical area V6 to stereoscopic depth gradients associated with self-motion. *Journal of Neurophysiology, 106*(3), 1240–1249.
Chance, F., Nelson, S., & Abbott, L. (1999). Complex cells as cortically amplified simple cells. *Nature Neuroscience, 2*(3), 277–282. https://doi.org/10.1038/6381.
Changizi, M. A., & Shimojo, S. (2008). "X-ray vision" and the evolution of forward-facing eyes. *Journal of Theoretical Biology, 254*(4), 756–767. https://doi.org/10.1016/j.jtbi.2008.07.011.
Chauhan, T., Masquelier, T., Montlibert, A., & Cottereau, B. (2018). Emergence of binocular disparity selectivity through Hebbian learning. *The Journal of Neuroscience, 38*(44), 9563–9578. https://doi.org/10.1523/JNEUROSCI.1259-18.2018.
Chino, Y., Smith, E., Hatta, S., & Cheng, H. (1997). Postnatal development of binocular disparity sensitivity in neurons of the primate visual cortex. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 17*(1), 296–307. https://doi.org/10.1523/JNEUROSCI.17-01-00296.1997.
Cloherty, S. L., Hughes, N. J., Hietanen, M. A., Bhagavatula, P. S., Goodhill, G. J., & Ibbotson, M. R. (2016). Sensory experience modifies feature map relationships in visual cortex. *ELife, 5*, Article e13911. https://doi.org/10.7554/eLife.13911.
Cooper, E., Burge, J., & Banks, M. (2011). The vertical horopter is not adaptable, but it may be adaptive. *Journal of Vision, 11*(3), https://doi.org/10.1167/11.3.20.
Cooper, E., & Norcia, A. (2014). Perceived depth in natural images reflects encoding of low-level luminance statistics. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 34*(35), 11761–11768. https://doi.org/10.1523/JNEUROSCI.1336-14.2014.
Cooper, M., & Pettigrew, J. (1979). A neurophysiological determination of the vertical horopter in the cat and owl. *Journal of Comparative Neurology, 184*(1), 1–25. https://doi.org/10.1002/cne.901840102.
Cottereau, B., McKee, S., Ales, J., & Norcia, A. (2011). Disparity-tuned population responses from human visual cortex. *Journal of Neuroscience, 31*(3), 954–965. https://doi.org/10.1523/JNEUROSCI.3795-10.2011.
Cottereau, B., McKee, S., Ales, J., & Norcia, A. (2012). Disparity-specific spatial interactions: evidence from EEG source imaging. *Journal of Neuroscience, 32*(3), 826–840. https://doi.org/10.1523/JNEUROSCI.2709-11.2012.
Cumming, B., & Parker, A. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature, 389*(6648), 280–283. https://doi.org/10.1038/38487.
DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1991). Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature, 352*(6331), 156–159. https://doi.org/10.1038/352156a0.
DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology, 89*(2), 1094–1111. https://doi.org/10.1152/jn.00717.2002.
Durand, J.-B., Celebrini, S., & Trotter, Y. (2007). Neural bases of stereopsis across visual field of the alert macaque monkey. *Cerebral Cortex, 17*(6), 1260–1273. https://doi.org/10.1093/cercor/bhl050.
Durand, J.-B., Zhu, S., Celebrini, S., & Trotter, Y. (2002). Neurons in parafoveal Areas V1 and V2 encode vertical and horizontal disparities. *Journal of Neurophysiology, 88*(5), 2874–2879. https://doi.org/10.1152/jn.00291.2002.
Ehinger, K., Adams, W., Graf, E., & Elder, J. (2017). Local depth edge detection in humans and deep neural networks. *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 2681–2689). .
Ferster, D., & Lindström, S. (1983). An intracellular analysis of geniculo-cortical connectivity in area 17 of the cat. *The Journal of Physiology, 342*(1), 181–215. https://doi.org/10.1113/jphysiol.1983.sp014846.
Fleet, D., Wagner, H., & Heeger, D. (1996). Neural encoding of binocular disparity: Energy models, position shifts and phase shifts. *Vision Research, 36*(12), 1839–1857. https://doi.org/10.1016/0042-6989(95)00313-4.
Freeman, R., & Pettigrew, J. (1973). Alteration of visual cortex from environmental asymmetries. *Nature, 246*(5432), 359–360. https://doi.org/10.1038/246359a0.
Geisler, W. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59*, 167–192. https://doi.org/10.1146/annurev.psych.58.110405.085632.
Geisler, W., Najemnik, J., & Ing, A. (2009). Optimal stimulus encoders for natural tasks. 17 17 *Journal of Vision, 9*(13), https://doi.org/10.1167/9.13.17.
Ghahghaei, S., McKee, S., & Verghese, P. (2019). The upper disparity limit increases gradually with eccentricity. 3 3 *Journal of Vision, 19*(11), https://doi.org/10.1167/19.11.3.
Gibaldi, A., & Banks, M. S. (2019). Binocular eye movements are adapted to the natural environment. *Journal of Neuroscience, 39*(15), 2877–2888. https://doi.org/10.1523/JNEUROSCI.2591-18.2018.
Gibaldi, A., Canessa, A., & Sabatini, S. P. (2017). The active side of stereopsis: fixation strategy and adaptation to natural environments. *Scientific Reports, 7*(1), 1–18. https://doi.org/10.1038/srep44800.
Gibaldi, A., Canessa, A., Solari, F., & Sabatini, S. P. (2015). Autonomous learning of disparity–vergence behavior through distributed coding and population reward: Basic mechanisms and real-world conditioning on a robot stereo head. *Robotics and Autonomous Systems, 71*, 23–34. https://doi.org/10.1016/j.robot.2015.01.002.
Gogel, W. C. (1965). Equidistance tendency and its consequences. *Psychological Bulletin, 64*(3), 153–163. https://doi.org/10.1037/h0022197.
Goncalves, N., & Welchman, A. (2017). "What not" detectors help the brain see in depth. *Current Biology*. https://doi.org/10.1016/j.cub.2017.03.074.
Gray, M. S., Pouget, A., Zemel, R. S., Nowlan, S. J., & Sejnowski, T. J. (1998). Reliable disparity estimation through selective integration. *Visual Neuroscience, 15*(3), 511–528. https://doi.org/10.1017/S0952523898153129.
Haefner, R., & Cumming, B. (2008). Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model. *Neuron, 57*(1), 147–158. https://doi.org/10.1016/J.NEURON.2007.10.042.
Hayhoe, M., Gillam, B., Chajka, K., & Vecellio, E. (2009). The role of binocular vision in walking. *Visual Neuroscience, 26*(1), 73–80. https://doi.org/10.1017/S0952523808080838.
von Helmholtz, H. (1924). *Handbuch der physiologischen optik*. Optical Society of America.
Hibbard, P. (2008). Binocular energy responses to natural images. *Vision Research, 48*(12), 1427–1439. https://doi.org/10.1016/j.visres.2008.03.013.
Hibbard, P. (2007). A statistical model of binocular disparity. *Visual Cognition, 15*(2), 149–165. https://doi.org/10.1080/13506280600648018.
Hibbard, P., & Bouzit, S. (2005). Stereoscopic correspondence for ambiguous targets is affected by elevation and fixation distance. *Spatial Vision, 18*(4), 399–411. https://doi.org/10.1163/1568568054389589.
Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12), 967–992. https://doi.org/10.1167/4.12.1.
Hoffmann, K., & Stone, J. (1971). Conduction velocity of afferents to cat visual cortex: A correlation with cortical receptive field properties. *Brain Research, 32*(2), 460–466. https://doi.org/10.1016/0006-8993(71)90340-4.
Hoyer, P., & Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Network: Computation in Neural Systems, 11*(3), 191–210. https://doi.org/10.1088/0954-898X_11_3_302.
Hsu, A. S., & Dayan, P. (2007). An unsupervised learning model of neural plasticity: Orientation selectivity in goggle-reared kittens. *Vision Research, 47*(22), 2868–2877. https://doi.org/10.1016/j.visres.2007.07.023.
Huang, J., Lee, A., & Mumford, D. (2000). Statistics of range images. Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662), 1, 324–331 vol.1. https://doi.org/10.1109/CVPR.2000.855836.
Hunt, J., Dayan, P., & Goodhill, G. (2013). Sparse coding can predict primary visual cortex receptive field changes induced by abnormal visual input. *PLoS Computational Biology, 9*(5), Article e1003005. https://doi.org/10.1371/journal.pcbi.1003005.
Hunter, D., & Hibbard, P. (2015). Distribution of independent components of binocular natural images. *Journal of Vision, 15*(13), 6. https://doi.org/10.1167/15.13.6.
Hunter, D., & Hibbard, P. (2016). Ideal binocular disparity detectors learned using independent subspace analysis on binocular natural image pairs. *PLOS ONE, 11*(3), Article e0150117. https://doi.org/10.1371/journal.pone.0150117.
Hunter, D., & Hibbard, P. (2018). The effect of image position on the Independent Components of natural binocular images. *Scientific Reports, 8*(1), 449. https://doi.org/10.1038/s41598-017-18460-1.
Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*(4–5), 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5.

Ito, H., & Shibata, I. (2005). Self-motion perception from expanding and contracting optical flows overlapped with binocular disparity. *Vision Research, 45*(4), 397–402. https://doi.org/10.1016/j.visres.2004.11.009.
Iyer, A., & Burge, J. (2018). Depth variation and stereo processing tasks in natural scenes. 4 4 *Journal of Vision, 18*(6), https://doi.org/10.1167/18.6.4.
Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3*(1), 79–87. https://doi.org/10.1162/neco.1991.3.1.79.
Jordan, J. R., Geisler, W., & Bovik, A. C. (1990). Color as a source of information in the stereo correspondence process. *Vision Research, 30*(12), 1955–1970. https://doi.org/10.1016/0042-6989(90)90015-D.
Kim, H. R., Angelaki, D. E., & DeAngelis, G. C. (2015). A functional link between MT neurons and depth perception based on motion parallax. *Journal of Neuroscience, 35*(6), 2766–2777. https://doi.org/10.1523/JNEUROSCI.3134-14.2015.
Klimmasch, L., Schneider, J., Lelais, A., Shi, B. E., & Triesch, J. (2018). An active efficient coding model of binocular vision development under normal and abnormal rearing conditions. In P. Manoonpong, J. C. Larsen, X. Xiong, J. Hallam, & J. Triesch (Eds.). *From Animals to Animats 15* (pp. 66–77). Springer International Publishing. https://doi.org/10.1007/978-3-319-97628-0_6.
Knill, D., & Richards, W. (Eds.). (1996). Perception as Bayesian Inference. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511984037.
Knill, D., & Saunders, J. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43*(24), 2539–2558. https://doi.org/10.1016/s0042-6989(03)00458-9.
Konda, K., & Memisevic, R. (2014). A unified approach to learning depth and motion features. Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing, 1–7.
La Chioma, A., Bonhoeffer, T., & Hübener, M. (2019). Area-specific mapping of binocular disparity across mouse visual cortex. *Current Biology, 29*(17), 2954–2960.e5. https://doi.org/10.1016/j.cub.2019.07.037.
Landers, D. D., & Cormack, L. K. (1997). Asymmetries and errors in perception of depth from disparity suggest a multicomponent model of disparity processing. *Perception & Psychophysics, 59*(2), 219–231. https://doi.org/10.3758/BF03211890.
Langer, M. S., & Mannan, F. (2012). Visibility in three-dimensional cluttered scenes. *Journal of the Optical Society of America A, 29*(9), 1794–1807. https://doi.org/10.1364/JOSAA.29.001794.
Lelais, A., Mahn, J., Narayan, V., Zhang, C., Shi, B. E., & Triesch, J. (2019). Autonomous development of active binocular and motion vision through active efficient coding. *Frontiers in Neurorobotics, 13*. https://doi.org/10.3389/fnbot.2019.00049.
Levi, D., Knill, D., & Bavelier, D. (2015). Stereopsis and amblyopia: A mini-review. *Vision Research, 114*, 17–30. https://doi.org/10.1016/j.visres.2015.01.002.
Li, Y., Fitzpatrick, D., & White, L. (2006). The development of direction selectivity in ferret visual cortex requires early visual experience. *Nature Neuroscience, 9*(5), 676–681. https://doi.org/10.1038/nn1684.
Lippert, J., & Wagner, H. (2001). A threshold explains modulation of neural responses to opposite-contrast stereograms. *NeuroReport, 12*(15), 3205.
Liu, Y., Bovik, A. C., & Cormack, L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision, 8*(11), 19.1-14. https://doi.org/10.1167/8.11.19.
Livingstone, M., & Tsao, D. (1999). Receptive fields of disparity-selective neurons in macaque striate cortex. *Nature Neuroscience, 2*(9), 825–832. https://doi.org/10.1038/12199.
Lonini, L., Forestier, S., Teulière, C., Zhao, Y., Shi, B., & Triesch, J. (2013). Robust active binocular vision through intrinsically motivated learning. *Frontiers in Neurorobotics, 7*, 20. https://doi.org/10.3389/fnbot.2013.00020.
Lonini, L., Zhao, Y., Chandrashekhariah, P., Shi, B. E., & Triesch, J. (2013). Autonomous learning of active multi-scale binocular vision. *IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL), 2013*, 1–6. https://doi.org/10.1109/DevLrn.2013.6652541.
Malpeli, J. (1983). Activity of cells in area 17 of the cat in absence of input from layer a of lateral geniculate nucleus. *Journal of Neurophysiology, 49*(3), 595–610. https://doi.org/10.1152/jn.1983.49.3.595.
Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature, 389*(6648), 283–286. https://doi.org/10.1038/38496.
McGuire, B., Hornung, J., Gilbert, C., & Wiesel, T. (1984). Patterns of synaptic input to layer 4 of cat striate cortex. *Journal of Neuroscience, 4*(12), 3021–3033. https://doi.org/10.1523/JNEUROSCI.04-12-03021.1984.
McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30*(11), 1763–1779. https://doi.org/10.1016/0042-6989(90)90158-H.
Melmoth, D. R., & Grant, S. (2006). Advantages of binocular vision for the control of reaching and grasping. *Experimental Brain Research, 171*(3), 371–388. https://doi.org/10.1007/s00221-005-0273-x.
Murphy, A., Ban, H., & Welchman, A. (2013). Integration of texture and disparity cues to surface slant in dorsal visual cortex. *Journal of Neurophysiology, 110*(1).
Nadler, J. W., Barbash, D., Kim, H. R., Shimpi, S., Angelaki, D. E., & DeAngelis, G. C. (2013). Joint Representation of Depth from Motion Parallax and Binocular Disparity Cues in Macaque Area MT. *Journal of Neuroscience, 33*(35), 14061–14074. https://doi.org/10.1523/JNEUROSCI.0251-13.2013.
Nasr, S., & Tootell, R. B. H. (2018). Visual field biases for near and far stimuli in disparity selective columns in human visual cortex. *NeuroImage, 168*, 358–365. https://doi.org/10.1016/j.neuroimage.2016.09.012.
Nieder, A., & Wagner, H. (2000). Horizontal-disparity tuning of neurons in the visual forebrain of the behaving barn owl. *Journal of Neurophysiology, 83*(5), 2967–2979.
Norcia, A. M., Gerhard, H. E., & Meredith, W. J. (2017). Development of relative disparity sensitivity in human visual cortex. *Journal of Neuroscience, 37*(23), 5608–5619. https://doi.org/10.1523/JNEUROSCI.3570-16.2017.
Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors. *Science, 249*(4972), 1037–1041. https://doi.org/10.1126/science.2396096.
Okajima, K. (2004). Binocular disparity encoding cells generated through an Infomax based learning algorithm. *Neural Networks, 17*(7), 953–962. https://doi.org/10.1016/j.neunet.2004.02.004.
Olshausen, B. (2003). Learning sparse, overcomplete representations of time-varying natural images. Proceedings 2003 International Conference on Image Processing (Cat. No.03CH37429), 1, I–41. https://doi.org/10.1109/ICIP.2003.1246893.
Olshausen, B., & Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*(6583), 607–609. https://doi.org/10.1038/381607a0.
Olshausen, B., & Field, D. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research, 37*(23), 3311–3325. https://doi.org/10.1016/S0042-6989(97)00169-7.
Olshausen, B., & Field, D. (2005). How close are we to understanding V1? *Neural Computation, 17*(8), 1665–1699. https://doi.org/10.1162/0899766054026639.
Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience, 8*(5), 379–391. https://doi.org/10.1038/nrn2131.
Pettigrew, J. (1974). The effect of visual experience on the development of stimulus specificity by kitten cortical neurones. *The Journal of Physiology, 237*(1), 49–74. https://doi.org/10.1113/jphysiol.1974.sp010469.
Pettigrew, J., Olson, C., & Barlow, H. (1973). Kitten visual cortex: Short-term, stimulus-induced changes in connectivity. *Science, 180*(4091), 1202–1203. https://doi.org/10.1126/science.180.4091.1202.
Poggio, G., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual cortex: Binocular correlation and disparity selectivity. *Journal of Neuroscience, 8*(12) http://www.jneurosci.org/content/8/12/4531.short.
Poggio, G., Motter, B., Squatrito, S., & Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Research, 25*(3), 397–406. https://doi.org/10.1016/0042-6989(85)90065-3.
Potetz, B., & Lee, T.-S. (2003). Statistical correlations between two-dimensional images and three-dimensional structures in natural scenes. *JOSA A, 20*(7), 1292–1303. https://doi.org/10.1364/JOSAA.20.001292.
Potetz, B., & Lee, T.-S. (2006). Scaling Laws in Natural Scenes and the Inference of 3D Shape. Advances in Neural Information Processing Systems, 18, 1089–1096.
Prince, S. J. D., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology, 87*(1), 209–221. https://doi.org/10.1152/jn.00466.2000.
Prince, S., Pointon, A., Cumming, B., & Parker, A. (2002). Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. *J Neurophysiol, 87*(1), 191–208.
Prince, S., & Rogers, B. (1998). Sensitivity to disparity corrugations in peripheral vision. *Vision Research, 38*(17), 2533–2537. https://doi.org/10.1016/S0042-6989(98)00118-7.
Qiu, F. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: v2 combines stereoscopic cues with gestalt rules. *Neuron, 47*(1), 155–166. https://doi.org/10.1016/j.neuron.2005.05.028.
Read, J., & Cumming, B. (2004). Understanding the cortical specialization for horizontal disparity. *Neural Computation, 16*(10), 1983–2020. https://doi.org/10.1162/0899766041732440.
Read, J., & Cumming, B. (2007). Sensors for impossible stimuli may solve the stereo correspondence problem. *Nature Neuroscience, 10*(10), 1322–1328. https://doi.org/10.1038/nn1951.
Read, J., & Cumming, B. (2017). Visual perception: Neural networks for stereopsis. *Current Biology, 27*(12), R594–R596. https://doi.org/10.1016/j.cub.2017.05.013.
Read, J., Parker, A., & Cumming, B. (2003). A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. *Visual Neuroscience, 19*(06), 735–753. https://doi.org/10.1017/S0952523802196052.
Ringach, D. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88*(1)http://jn.physiology.org/content/88/1/455.short.
Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception, 8*(2), 125–134. https://doi.org/10.1068/p080125.
Rosenberg, A., & Angelaki, D. E. (2014). Reliability-dependent contributions of visual orientation cues in parietal cortex. *Proceedings of the National Academy of Sciences, 111*(50), 18043–18048. https://doi.org/10.1073/pnas.1421131111.
Rosenberg, A., Cowan, N. J., & Angelaki, D. E. (2013). The visual representation of 3D object orientation in parietal cortex. *Journal of Neuroscience, 33*(49), 19352–19361. https://doi.org/10.1523/JNEUROSCI.3174-13.2013.
Samonds, J. M., Potetz, B. R., & Lee, T. S. (2012). Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. *Proceedings of the National Academy of Sciences, 109*(16), 6313–6318. https://doi.org/10.1073/pnas.1200125109.
Samonds, J., Potetz, B., Tyler, C., & Lee, T.-S. (2013). Recurrent connectivity can account for the dynamics of disparity processing in V1. *Journal of Neuroscience, 33*(7), 2934–2946. https://doi.org/10.1523/JNEUROSCI.2952-12.2013.
Sanada, T. M., Nguyenkim, J. D., & DeAngelis, G. C. (2012). Representation of 3-D surface orientation by velocity and disparity gradient cues in area MT. *Journal of Neurophysiology, 107*(8), 2109–2122. https://doi.org/10.1152/jn.00578.2011.
Saxe, A., Bhand, M., Mudur, R., Suresh, B., & Ng, A. Y. (2011). Unsupervised learning models of primary cortical receptive fields and receptive field plasticity. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 24 (pp. 1971–1979). Curran Associates, Inc. http://papers.nips.cc/paper/4331-unsupervised-learning-models-of-primary-cortical-receptive-fields-and-receptive-field-plasticity.pdf.

Seemiller, E., Cumming, B., & Candy, T. R. (2018). Human infants can generate vergence responses to retinal disparity by 5 to 10 weeks of age. Journal of Vision, 18(6), 17–17. https://doi.org/10.1167/18.6.17.

Servos, P., & Goodale, M. A. (1994). Binocular vision and the on-line control of human prehension. *Experimental Brain Research, 98*(1), 119–127. https://doi.org/10.1007/BF00229116.

Séverac Cauquil, A., Delaux, S., Lestringant, R., Taylor, M. J., & Trotter, Y. (2009). Neural correlates of chromostereopsis: An evoked potential study. *Neuropsychologia, 47*(12), 2677–2681. https://doi.org/10.1016/j.neuropsychologia.2009.05.002.

Simmons, D., & Kingdom, F. (1994). Contrast thresholds for stereoscopic depth identification with isoluminant and isochromatic stimuli. *Vision Research, 34*(22), 2971–2982. https://doi.org/10.1016/0042-6989(94)90269-0.

Simoncelli, E., & Olshausen, B. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*(1), 1193–1216.

Sprague, W., Cooper, E., Tosic, I., & Banks, M. (2015). Stereopsis is adaptive for the natural environment. e1400254 e1400254 *Science Advances, 1*(4), https://doi.org/10.1126/sciadv.1400254.

Stevenson, S., Cormack, L., Schor, C., & Tyler, C. (1992). Disparity tuning in mechanisms of human stereopsis. *Vision Research, 32*(9), 1685–1694. https://doi.org/10.1016/0042-6989(92)90161-B.

Su, C.-C., Cormack, L. K., & Bovik, A. C. (2013). Color and depth priors in natural images. *IEEE Transactions on Image Processing, 22*(6), 2259–2274. https://doi.org/10.1109/TIP.2013.2249075.

Tanaka, K. (1985). Organization of geniculate inputs to visual cortical cells in the cat. *Vision Research, 25*(3), 357–364. https://doi.org/10.1016/0042-6989(85)90060-4.

Tao, L., Shelley, M., McLaughlin, D., & Shapley, R. (2004). An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proceedings of the National Academy of Sciences, 101*(1), 366–371. https://doi.org/10.1073/pnas.2036460100.

Tao, Xiaofeng, Zhang, Bin, Shen, Guofu, Wensveen, Janice, Smith, Earl L., 3rd, Nishimoto, Shinji, Ohzawa, Izumi, & Chino, Yuzo M. (2014). Early monocular defocus disrupts the normal development of receptive-field structure in V2 neurons of macaque monkeys. *J Neurosci, 34*(41), 13840–13854. https://doi.org/10.1523/JNEUROSCI.1992-14.2014.

Tyler, C. (1973). Stereoscopic vision: Cortical limitations and a disparity scaling effect. *Science, 181*(4096), 276–278. https://doi.org/10.1126/science.181.4096.276.

Tyler, C. (1991). The horopter and binocular fusion. In Vision and Visual Disorders. Vol. 9, Binocular Vision (pp. 19–37). Macmillan Publishers.

van Hateren, J., & Ruderman, D. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London Series B: Biological Sciences, 265*(1412), 2315–2320. https://doi.org/10.1098/rspb.1998.0577.

van Hateren, J., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings Biological Sciences/The Royal Society, 265*(1394), 359–366. https://doi.org/10.1098/rspb.1998.0303.

Verhoef, B.-E., Bohon, K. S., & Conway, B. R. (2015). Functional Architecture for disparity in macaque inferior temporal cortex and its relationship to the architecture for faces, color, scenes, and visual field. *The Journal of Neuroscience, 35*(17), 6952–6968. https://doi.org/10.1523/JNEUROSCI.5079-14.2015.

Watt, S. J., & Bradshaw, M. F. (2003). The visual control of reaching and grasping: Binocular disparity and motion parallax. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 404–415. https://doi.org/10.1037/0096-1523.29.2.404.

Westheimer, G., & McKee, S. P. (1980). Stereoscopic acuity with defocused and spatially filtered retinal images. *JOSA, 70*(7), 772–778. https://doi.org/10.1364/JOSA.70.000772.

Wiesel, T., & Hubel, D. (1963). Single-cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology, 26*(6), 1003–1017. https://doi.org/10.1152/jn.1963.26.6.1003.

Wiesel, T., & Hubel, D. (1974). Ordered arrangement of orientation columns in monkeys lacking visual experience. *The Journal of Comparative Neurology, 158*(3), 307–318. https://doi.org/10.1002/cne.901580306.

Yang, Z., & Purves, D. (2003). A statistical explanation of visual space. *Nature Neuroscience, 6*(6), 632–640. https://doi.org/10.1038/nn1059.

Zhao, Y., Rothkopf, C. A., Triesch, J., & Shi, B. E. (2012). A unified model of the joint development of disparity selectivity and vergence control. *IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL), 2012*, 1–6. https://doi.org/10.1109/DevLrn.2012.6400876.

Zhaoping, L. (2000). Pre-attentive segmentation in the primary visual cortex. *Spatial Vision, 13*, 25–50.

Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Network: Computation in Neural Systems, 17*(4), 301–334. https://doi.org/10.1080/09548980600931995.