Title

Emergence of binocular disparity selectivity through Hebbian learning

Authors

Tushar Chauhan* (Tushar.Chauhan@cnrs.fr)

Timothée Masquelier (Timothee.Masquelier@cnrs.fr)

Alexandre Montlibert (Alexandre.Montlibert@cnrs.fr)

Benoit R. Cottereau* (Benoit.Cottereau@cnrs.fr)

**Corresponding authors*

Affiliations

All authors are affiliated with:

1. Université de Toulouse, Centre de Recherche Cerveau et Cognition, Toulouse, France
2. Centre National de la Recherche Scientifique, Toulouse, France

Abstract

Neural responses in the early visual cortex strongly reflect the statistics of our environment. Although this has been described extensively in literature through various encoding hypotheses, an explanation as to how the cortex develops the structures to support these encoding schemes remains elusive. Here, using the more realistic example of binocular vision as opposed to monocular luminance-field images, we show that a simple Hebbian coincidence-detector is capable of accounting for the emergence of binocular, disparity selective, receptive fields. We propose a model based on spike-timing dependent plasticity (STDP) which not only converges to realistic single-cell and population characteristics, but also demonstrates how known biases in natural statistics may influence population encoding and downstream correlates of behaviour. Our modelling results suggest that Hebbian coincidence-detection is an important computational principle, and could provide a biologically plausible mechanism for the emergence of selectivity to natural statistics in the early sensory cortex.

1. Introduction

Sensory stimuli are not only the inputs to, but also shape the very process of, neural computation. This principle is a recurring theme in work investigating the computational underpinnings of early sensory processing - from the earliest works by Hebb[1] which give a conceptual framework for experience-driven plasticity in the human brain (Hebbian plasticity), to work by Horace Barlow[2,3] which provides a broader theoretical description of how dynamic stimulus-driven redundancy reduction could be an important organisational principle in early sensory processing. A modern, more rigorous extension of Barlow's ideas is the efficient coding theory[4–6] which postulates that the computational aim of early sensory processing is to use the least possible resources (neurones, energy) to code the most informative features of the stimulus (information efficiency).

A direct corollary to the efficient coding hypothesis is that if the inputs signals are coded efficiently, the statistical consistencies in the stimuli should then be reflected in the organisation and structure of the early, feedforward-dominated cortex[7]. The largest body of work on the efficient coding principle lies within the visual sensory modality. In the context of vision, a number of studies have applied efficient coding principles to natural scene images to show that information-theoretic constraints do indeed predict localised, oriented and band-pass representations, akin to those reported in the early visual cortex[8]. Over the years, several studies have shown how properties of natural images can not only explain neural selectivity[8–11], but also predict behaviour[7,12–14].

Nearly all these studies, however, rely on images of natural scenes taken using a single camera – effectively using a 2-D projection of a 3-D scene. Although this representation captures the luminance-field statistics of the scene; however, it does not reflect how visual data is actually acquired by the human brain. Humans are binocular, and use two simultaneous retinal projections which enable the sensing of disparity (distance between retinal projections of corresponding points). Disparity, in turn, can be used to make inferences about the 3-D structure of the scene such as calculations of distance, depth and surface-slant. Despite critical results in the analysis of luminance-field statistics, only a handful of studies[15–18] have attempted to analyse the relationship between binocular projections of natural scenes and the properties of binocular neurones in the early visual system. Furthermore, these studies investigate the

relationship between natural scenes and cortical selectivity in terms of a global optimisation problem which is *solved* in the adult brain, leading to cortical structures which encode relevant statistics of the stimuli. However, the question as to how these encoding schemes might actually emerge in the early sensory cortex remains, as yet, unanswered[19]. Here we show, for the very first time, how simple coincidence detection based on spike-timing dependent plasticity[20,21] (STDP) could offer a biologically plausible mechanism for arriving at neural populations close to those reported in the early visual system. We endow a neural network with a Hebbian STDP rule which reinforces synapses carrying causal spikes, and find that an unsupervised exposure of this network to natural stereoscopic stimuli leads to a converged population of neurones with binocular receptive fields which show single-cell and population characteristics close to those reported in electrophysiological studies[22,23]. Moreover, we show that our model also captures retinotopic biases in disparity statistics reported in various electrophysiological studies[24,25]. Thus, taken together, our findings suggest that known rules of synaptic plasticity are sufficient to explain the relationship between biases reported in the early visual system and the statistics of natural stimuli. Furthermore, they also provide a compelling demonstration of how coincidence-detection[26,27] could explain the emergence of selectivity in early sensory and multi-modal neural populations, both during and after development.

2. Methods

Dataset

To simulate binocular retinal input, we chose the Hunter-Hibbard database of stereo-images[16]. This database contains 139 stereo-images of natural scenes relevant to everyday binocular experience, containing objects like vegetables, fruit, stones, shells, flowers, and vegetation. The database was captured using a realistic acquisition geometry close to the configuration of the human visual system while fixating straight ahead. The lenses had a large field of view (20°), enabling them to capture binocular statistics across a sizeable part of the visual field. The distance between the two cameras was close to the human inter-ocular separation (65-70 mm), and the two cameras were focussed at realistic fixation-points (50 cm to several metres). This database is available at https://github.com/DavidWilliamHunter/Bivis) under the generous MIT license which guarantees free usage and distribution.

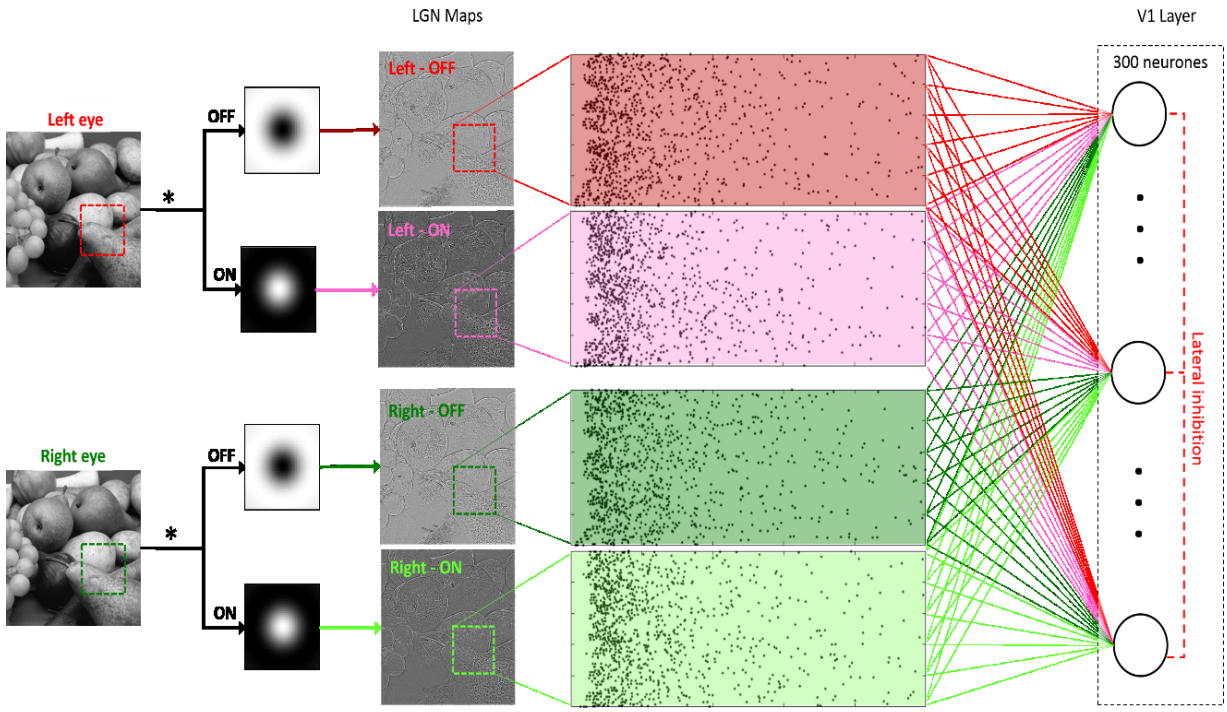

Figure 1: The proposed model. The left and right images were first convolved with ON/OFF centre difference-of-Gaussian kernels to get 4 LGN retinotopic maps (LGN Layer) corresponding to ON and OFF centre maps in the left and right eyes. Random retinal positions were then sampled, with each sample consisting of 4 patches - one from each of the 4 LGN maps. The first-spike latencies of all LGN units contained within a given sample were used to drive one update iteration of an STDP neural network (V1 layer). The network consisted of 300 integrate and fire neurones endowed with a lateral inhibition scheme to ensure sparsity. A large number of samples (results reported for $N = 100{,}000$) was used for training.

Modelling of the retinogeniculate pathway

The computational pipeline employed by the model is shown in Figure 1. The first stage of the model implemented the processing in the retinogeniculate pathway. The stereoscopic images were convolved with ON and OFF centre-surround kernels to simulate the computations performed by the retinal ganglion cells (RGCs) and the lateral geniculate nucleus (LGN). The filters were implemented using Difference of Gaussian (DoG) kernels, and their receptive field sizes were chosen to reflect the size of typical LGN centre-surround magnocellular receptive fields[28] – $0.3°/1°$ (centre/surround) for foveal simulations and $1°/2°$ (centre/surround) for peripheral simulations. The activity of each kernel over the entire image can roughly be interpreted as a retinotopic contrast map of the image. Four maps were calculated, corresponding to ON- and OFF-centre populations in the left and right monocular pathways.

Patches approximately three times the size of the LGN receptive fields were then randomly sampled from regions-of-interest (ROIs) in the retinotopic maps. This could be interpreted as

the initial dendritic receptive field of the V1 neurones, which was subsequently pruned by our model. The results presented in this paper involve two main ROIs – foveal, corresponding to patch-centre eccentricities less than 3°, and peripheral, corresponding to eccentricities between 6° and 10°. The patches sampled in these two regions measured $3° \times 3°$ and $6° \times 6°$ respectively. In addition, our investigation of upper/lower hemi-field biases used two ROIs with patches centred in the respective halves of the visual field. Each sample consisted of activations from four patches centred at the same retinal location in all maps. To simulate the fixation behaviour of typical strabismic amblyopes, a misalignment between the two eyes was introduced while keeping within the ROI (the retinal correspondence between the ON/OFF layers of the same eye was maintained).

The activity of each sample was thresholded and converted to relative first-spike latencies through a monotonically decreasing function (a token function $y = 1/x$ was chosen for the reported simulations, but all monotonically decreasing functions are equivalent[29]), thereby ensuring that the most active units fired first, while units with lower activity fired later, or not at all. The threshold was adaptive, and for each sample, ensured that only 10% of the cells fired. Latency-based encoding of stimulus properties has been reported extensively in the early visual system[30–34], and allows for fast and efficient networks describing early visual selectivity[35–37]. These trains of first- spikes (represented by their latencies) from the random samples constituted the input to the STDP network.

V1: STDP neural network

The STDP network used in the simulations consisted of 300 integrate-and-fire neurones with an initial dendritic receptive field which was three times the size of the LGN filters (foveal: $3° \times 3°$; peripheral: $6° \times 6°$). At the start, each neurone was fully connected to all LGN afferents within its receptive field through synapses with randomly assigned weights. The weights were restricted between 0 and 1 throughout the simulation. Here, the non-negative values of the weights reflect the fact that thalamic connections to V1 are predominantly excitatory in nature[38,39].

The LGN spikes from each sample (each sample being equivalent to retinotopically

corresponding patches in the four LGN maps) were propagated through the network serially. After each sample, the synaptic weights were updated using an unsupervised multiplicative STDP rule[40] which ensured a graded weight distribution in the converged units. For a synapse connecting a pair of pre- and post-synaptic neurones, this rule is described by:

$$\Delta w = \begin{cases} -\alpha^- \cdot w^{\mu^-} \cdot K(\Delta t, \tau_-), & \Delta t \leq 0 \\ \alpha^+ \cdot (1-w)^{\mu^+} \cdot K(\Delta t, \tau_+), & \Delta t > 0 \end{cases}$$

Here, a pre- and post- synaptic pair of spikes with a time difference Δt introduce a change Δw in the weight of the synapse which is given by the product of a temporal filter K and a power function of its current weight w. When a post-synaptic spike occurs shortly after a pre-synaptic spike ($\Delta t > 0$), there is a strengthening of the synapse – also called long-term potentiation (LTP). Conversely, when the post-synaptic spike occurs before the pre-synaptic spike ($\Delta t \leq 0$), or in the absence of a pre-synaptic spike, there is a weakening of the synapse or a long-term depression (LTD). The LTP and LTD are driven by their respective learning rates α^+ and α^-. The learning rates are positive ($\alpha^\pm \geq 0$) and determine the maximum amount of change in synaptic weights when $\Delta t \to \pm 0$. The parameters $\mu^\pm \in [0, 1]$ describe the degree of non-linearity in the LTP and LTD update rules. In practice, a non-linearity ensures that the final weights are graded, and prevents convergence to bimodal distributions saturated at the upper and lower limits (0 and 1 in our case).

For the network used in the present study, the learning rates were fixed with $\alpha^+ = 5 \times 10^{-3}$ and $\alpha^+/\alpha^- = (4/3)$. Furthermore, we used a high nonlinearity for the LTP process ($\mu^+ = 0.65$) to ensure that we were able to capture fading receptive fields through continuous weights, and used an almost additive LTD rule ($\mu^- = 0.05$) to ensure pruning of continuously depressed synapses. In both LTP and LTD updates we imposed $w \in [0, 1]$ by clipping the weights at the limits (hard bounds). Since our network relies on first-spike latencies, the temporal filter K was approximated as a step-function ($\tau_\pm \to \infty$), making the update rule very efficient and fast. Furthermore, a lateral inhibition scheme was implemented such that, if a neurone fired to a certain feature during learning, it simultaneously prevented other neurones in the population from firing until the next sample was processed. This scheme leads to a sparse neural population where the probability of any two neurones learning the same feature is greatly reduced.

Analysis of converged receptive fields

The post-convergence receptive fields of the STDP neurones were approximated by a linear summation of the afferent LGN receptive fields weighted by the converged synaptic weights. If a weight w_{ij} connected the ith neurone to the jth LGN afferent with receptive field ψ_j, the receptive field ξ_i of the neurone was estimated by:

$$\xi_i \approx \sum_{j \in LGN} w_{ij} \psi_j$$

The receptive fields calculated using this method are, in principle, similar to point-wise estimates of the receptive field calculated by electrophysiologists. To better characterise the properties of these receptive fields, they were fitted by two dimensional Gabor functions. These Gabor functions are given by:

$$G(k, f, \theta, \phi, \boldsymbol{x}, \boldsymbol{\sigma}) = k * S(f, \theta, \phi\ ) * G(\boldsymbol{x}, \boldsymbol{\sigma}, \theta)$$

where, k is a measure of the sensitivity of the Gabor, S is a 2-D sinusoid propagating with frequency f and phase ϕ in the direction θ, and G is a 2-D Gaussian envelope centred at $\boldsymbol{x}$, with size parameter $\boldsymbol{\sigma}$, also oriented at an angle θ. A multi-start global search algorithm was used to estimate optimal parameters in each eye for each neurone. The process was expedited by using the frequency response of the cell (Fourier amplitude) to limit the search-space of the frequency parameter. For each neurone, the goodness of the Gabor fit was characterised using the maximum of the R^2 goodness-of-fit values in the two eyes:

$$R^2_{max} = \max\{R^2_{Left}, R^2_{Right}\}$$

This ensured that the goodness-of-fit calculations were independent of whether the unit was monocular or binocular.

The responses of the converged neurones to disparity (disparity tuning curves, DTCs) were evaluated by presenting them with random dot stereograms (RDSs) at various disparities. The RDS patterns used for testing were perfectly correlated, with a dot size of about $15'$ visual angles, and a dot density of 0.24 (i.e., 24% of the stimulus was covered by dots, discounting any overlap). The DTC was calculated by averaging the response over a large number of stimulus presentations ($N = 10{,}000$ reported here). The RDS responses were further quantified by a Binocularity Interaction Index[41] (BII) given by:

$$BII = \frac{R_{max} - R_{min}}{R_{max} + R_{min}}$$

where R_{max} and R_{min} are maximum and minimum responses on the disparity tuning curve. BII values close to zero indicate that the neural response is weakly modulated by disparity, while higher values indicate a sharper tuning to specific disparity values. In retinotopic simulations, the phase and position disparity were estimated by methods commonly applied in electrophysiological studies – one-dimensional Gabor functions were fitted to the DTCs and the phase of the sinusoid and the offset of the Gaussian envelope from zero-disparity were taken as the phase and position disparities respectively.

Network convergence and robustness

The evolution of the synaptic weights with time was characterised by using a Convergence Index (CI). If $\Delta w_{ij}(t)$ is the change in the weight of the synapse connecting the ith neurone to the jth LGN afferent at time t, the convergence index was defined as:

$$CI(t) = \frac{1}{N_{LGN} N_{V1}} \sum_{i \in V1} \sum_{j \in LGN} \left| \Delta w_{ij}(t) \right|$$

where N_{LGN} and N_{V1} are the number of LGN units and the number of the V1 neurones respectively. For a network driven to convergence, this index should reduce asymptotically with time. To report the convergence for a single unit, a version of the above index defined by ignoring the V1-specific terms was used.

The robustness of the model to noise was tested by adding external (image) and internal (LGN) noise to the system. Gaussian noise with an SNR varying between -3dB ($\sigma_{signal} = 0.5 \times \sigma_{noise}$) to 10dB ($\sigma_{signal} = 10 \times \sigma_{noise}$) was used, and the performance of the network at various noise levels was characterised using two metrics – the convergence index (CI), and the R^2_{max} of Gabor fitting. Taken together, the CI and R^2_{max} characterise the robustness of the network in terms of both the synaptic and the functional convergence.

Code

On acceptance of the manuscript for publication, the code will be made available on a public server. During the review process, the editor and the reviewers will be given access to the code upon request.

3. Results

Foveal population

In Figure 2a, we present a representative unit from a foveally trained population of 300 V1 neurones. The first row of Figure 2a shows the initial receptive field of the neurone. The weights were set to random values sampled from a $\sim N(0.2\,,0.07)$ Gaussian distribution, resulting in an initial receptive field with no specificity in terms of structure or excitatory/inhibitory sub-fields. The second row of the figure shows the receptive field of the same neurone after STDP learning (see also, Video). The converged receptive field is local, bandpass, oriented (vertical) and strikingly Gabor-like. Figure 2a also shows the Convergence Index (CI) of the neurone which characterises the stability of its synaptic weights over the learning period. We find that the fluctuations in the average synaptic strength decrease over time to a very small value – indicating that the weight distribution has converged.

In electrophysiology, a neurone with a binocular receptive field is often characterised in terms of its response to binocular disparity[23,42,43]. This response is measured by recording the average activity of the neurone as a function of the disparity of Random Dot Stereogram (RDS) stimuli over multiple presentations – a Disparity Tuning Curve (DTC). Figure 2a shows the DTC for the selected neurone, estimated using correlated horizontal RDS stimuli. This neurone shows a selectivity for negative, crossed disparities, indicating that it is more likely to respond to objects closer to the observer than fixation. It also has a high Binocularity Interaction Index[41,43] (BII) of 0.92 (see Methods), indicating that there is a substantial modulation of its neuronal activity by binocular disparity. Furthermore, Figure 2b shows three other examples of converged receptive fields and the corresponding DTCs (one neurone per row). The first neurone shows a tuning to zero-disparity (fixation), the second neurone to uncrossed or positive disparity (objects farther away than the fixation), and the third neurone shows an anti-symmetric tuning curve which is typically attributed to anti-phase monocular receptive fields. As seen from the examples in Figure 2b, the receptive fields have a variety of orientations and sizes, and closely

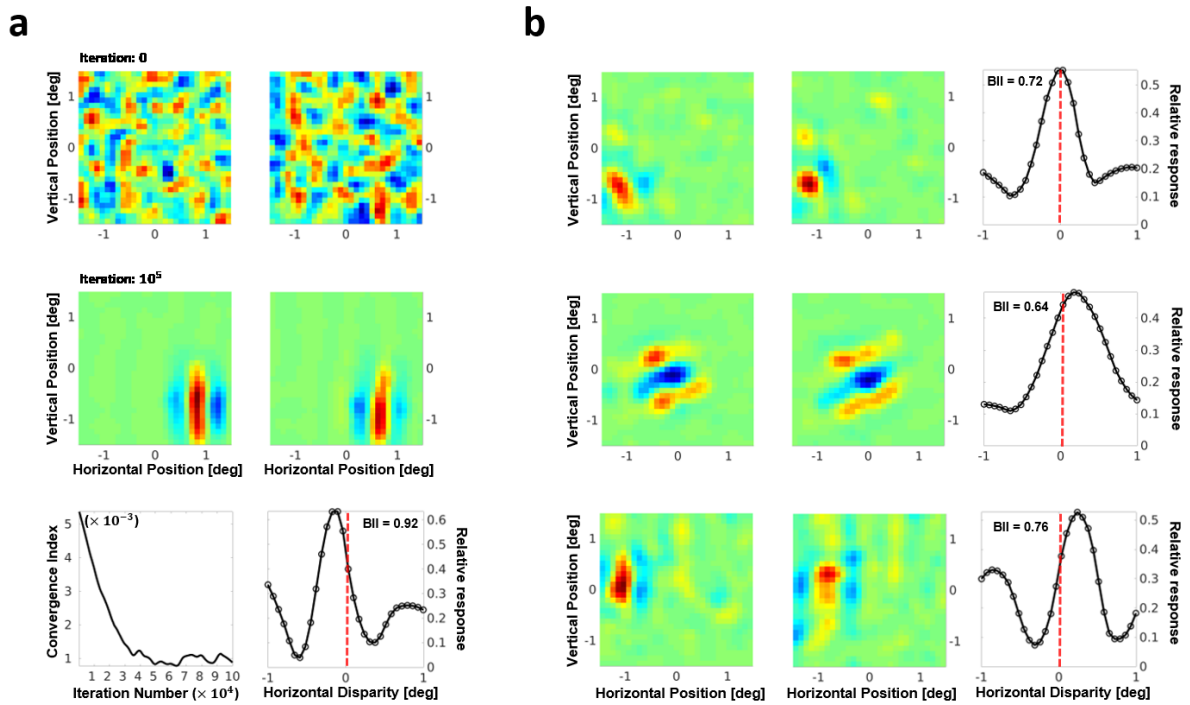

Figure 2: (a) A representative neurone from the simulated population. Receptive fields in the left and the right eye before (first row) and after (second row) a run of 10^5 iterations. The Convergence Index (bottom left), which measures the mean change in synaptic strength over time. DTC for horizontal disparity (bottom right), estimated by multiple presentations of RDS stimuli. The Binocular Integration Index (BII) is marked within the DTC graph. (b) Receptive fields and DTCs for neurones showing tuned excitatory (first row), tuned far (second row) and anti-symmetric (third row) responses to horizontal disparity.

resemble two-dimensional Gabor functions.

Receptive field properties

Since most converged receptive fields in the foveal simulation show a Gabor-like spatial structure, we investigated how well an ideal Gabor function would fit these receptive fields. Figure 3a shows the goodness-of-fit R^2 parameter for the left and the right eye for all 300 neurones in the foveal population, with the colour code corresponding to its maximum across the two eyes, which indicates how well the Gabor model fits the receptive field irrespective of whether it is monocular or binocular. A majority of the neurones lie along the diagonal in the upper right quadrant, showing an excellent fit in both eyes; with some monocular neurones lying in the upper-left and bottom-right quadrants. It is also interesting to note that irrespective

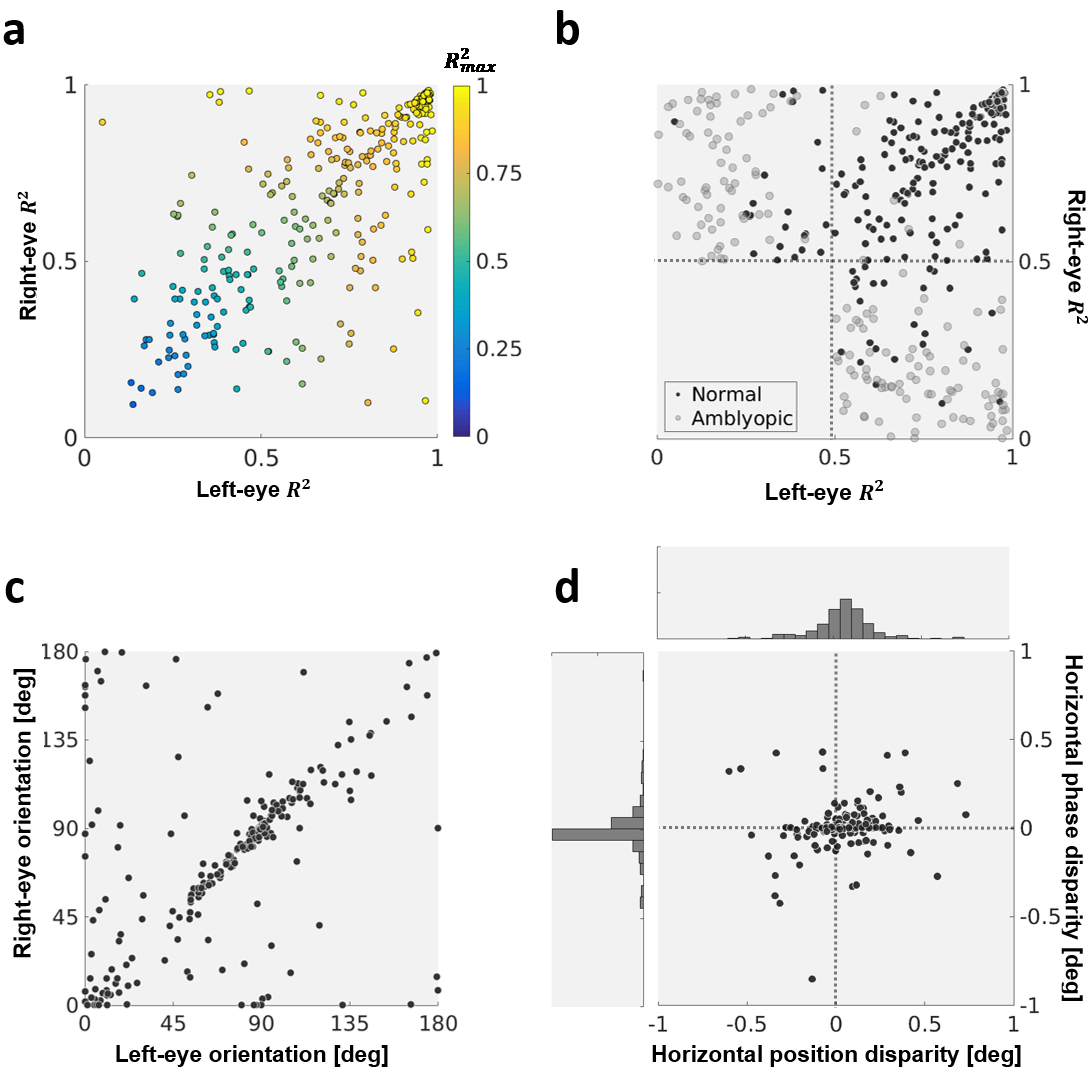

Figure 3: Characterising a simulated population of 300 foveal V1 neurones. (a) The goodness-of-fit R^2 criteria for the Gabor fitting in the left and the right eye. Each dot represents a neurone, while the filled colour indicates the quantity $R^2_{max} = \max\{R^2_{Left}, R^2_{Right}\}$, a measure of how well a Gabor function describes the receptive field. (b) Neurones ($R^2_{max} \geq 0.5$) from normal (black dots) and an amblyopic (grey dots) populations trained on foveal stimuli. The amblyopic population was trained by misaligning the two eyes randomly during each training iteration. (c) The orientation parameter for the Gabors fitted in the left and the right eye. Only neurones with binocular receptive fields defined by the criterion $\min\{R^2_{Left}, R^2_{Right}\} \geq 0.5$ are shown. Note that the range of the angles is $[0°, 180°[$ with $0°$ being the same as $180°$. (d) DTCs for the binocular neurones from (c) were estimated by multiple presentations of RDS stimuli at varying disparities. Gabors were fitted to the resulting curves, thereby giving a measure of the position and phase disparity in the receptive fields[23,42]. This figure shows the distribution of the two types of disparities.

of the degree of fit, the R^2 values for the left and the right eye are tightly correlated. This binocular correlation is non-trivial and cannot simply be attributed to learning from a stereoscopic dataset, as the inter-ocular correlations in the stimuli compete with intra-ocular

correlations.

To demonstrate this, in Figure 3b we show neurones with a reasonably good Gabor-fit ($R^2_{max} \geq 0.5$) for two simulated populations – the normal foveal population from Figure 3a (black dots) and a simulated amblyopic population (grey dots). In the amblyopic population, the projections in the two eyes were sampled from different foveal locations within the same image (see Methods), thereby mimicking the oculomotor behaviour exhibited by strabismic amblyopes. Here, although neurones learn from left and right image-sets with the same widefield statistics, due to the lack of inter-eye correlations they primarily develop monocular receptive fields – choosing local features from either the left or the right eye.

Figure 3c shows the receptive field orientations in the left and the right eye of binocular neurones in the foveal population. Here, we define a binocular unit by using the criterion: $\min\{R^2_{left}, R^2_{right}\} \geq 0.5$, i.e., both eyes fit the Gabor-model with an R^2 of at least 0.5. There is a strong correspondence between the orientation in the two eyes ($0°$ and $180°$ are equivalent orientations), with a large number of neurones being oriented along the cardinal directions, i.e., horizontally or vertically. A similar correspondence is found between the Gabor frequencies in the two eyes (see Supplementary Material). Furthermore, Figure 3d shows the position and phase encoding of the horizontal disparities as derived by fitting Gabor functions to DTCs estimated by using RDS stimuli (see legend and Methods for details). Both position and phase are found to code for disparities between $-0.5°$ and $0.5°$ visual angle, with most units coding for near-zero disparity.

Biases in natural statistics and neural populations

The disparity statistics of any natural scene show certain characteristic biases which are often reflected in the neural encoding in the early visual pathway. In this section, we investigate whether such biases are also found in our neural populations. It is established that as one traverses from the fovea to the periphery, the range of both the horizontal and the vertical disparity increases[44]. Figure 4a shows a histogram of the horizontal disparity in two simulated populations of neurones, one trained on inputs from the central field of vision (eccentricity $\leq 3°$, shown in purple), and the other on inputs from the periphery ($6° \leq$ eccentricity $\leq 10°$,

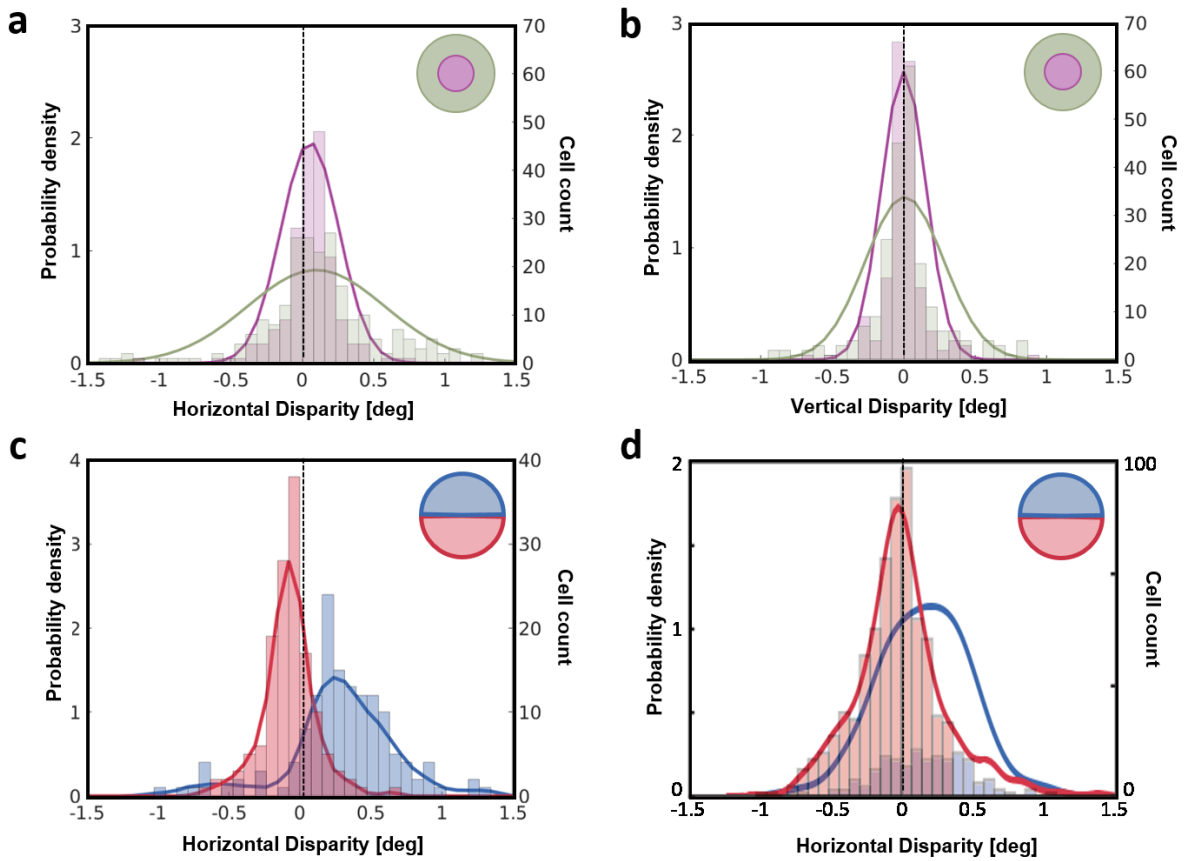

Figure 4: Retinotopic simulations and population biases in horizontal and vertical disparity. (a) Histograms and probability density functions of horizontal disparity in two simulated populations of 300 V1 neurones each. The first population was trained on inputs sampled from the fovea (purple) while the second population was trained on inputs from the periphery (green). (b) Vertical disparity for the same population as (a). (c) Histogram and population density of two populations of 300 neurones each, trained on inputs from the upper (blue) and lower(red) hemifield. (d) A meta-analysis[24] of results from electrophysiological recordings in monkey V1 neurones (820 neurones in total) by five separate groups. The neurones were split in two groups – neurones with upper-hemifield receptive fields are in blue, while those with receptive fields in the lower hemifield are in red.

shown in green). The peripheral population shows a clear broadening of its tuning with respect to the central population. The same is also found for vertical disparity in Figure 4b. Moreover, a comparison of Figure 4a and Figure 4b reveals that the tuning for horizontal disparity is much broader than vertical disparity in both foveal and peripheral populations. This, once again, is a known bias caused by the fact that the two eyes are separated along the horizontal meridian while they lie at the same height.

Another statistical bias reported in retinal projections of natural scenes[24] is that horizontal disparities in the lower hemifield are more likely to be crossed (negative) while the upper hemifield is more likely to contain uncrossed (positive) disparities. This bias is not altogether

surprising considering that objects in the lower field of vision are likely to be closer to the observer than objects in the upper field[24,45]. A meta-analysis[24] of 820 V1 neurones from five separate electrophysiological studies indicates that this bias could also be reflected in the binocular neurones in the early visual system. Figure 4d shows the results of this meta-analysis with the probability density for the neurones in the lower hemifield (in red) showing a bias towards crossed disparities while the upper hemifield (in blue) shows a bias towards uncrossed disparities. In comparison, Figure 4c presents histograms of the horizontal disparity in binocular neurones from two simulated populations – one trained on the lower hemifield (red), and the other on the upper hemifield (blue). Both the range and the peak-tuning of the two simulated populations closely resemble those reported in electrophysiology.

Robustness to internal and external noise

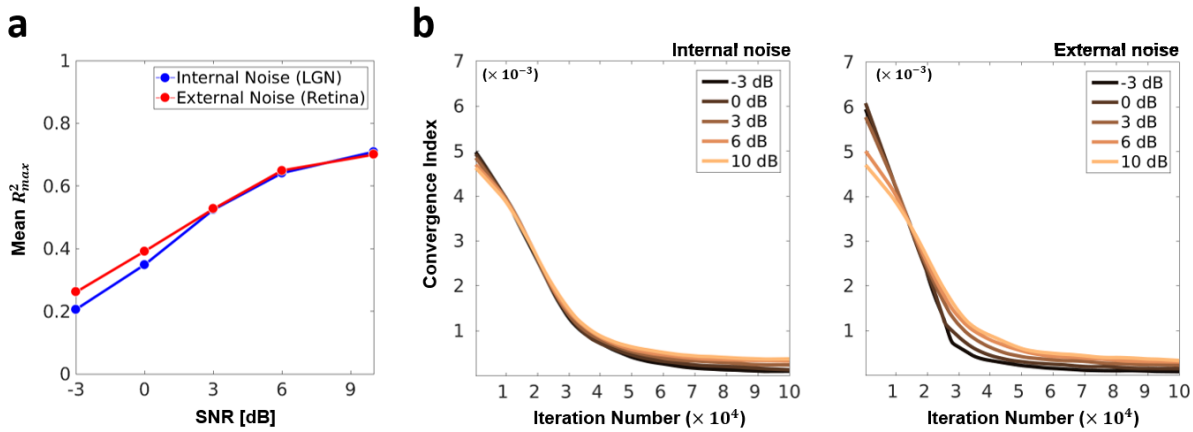

Figure 5: Robustness of the system to internal and external noise. (a) The mean R^2_{max} for simulations with additive white Gaussian noise at various SNR levels. The noise was introduced at the LGN stage (internal, in blue), and the retinal input (external, in red). The higher the value of the mean R^2_{max}, the closer the converged receptive fields are to V1 Gabor-like receptive fields. (b) The Convergence Index of each simulation in (a), plotted against the training period. The internal noise simulations are plotted to the left, while the external noise simulations are to the right. In both graphs, the lightness of the curve codes for the SNR (the darker the curve, the noisier the input). For a stable convergence of the synaptic weights, the CI should approach zero asymptotically.

In the results reported so far, the only source of noise are the camera sensors used to capture the stereoscopic dataset. To further test the robustness of our model we introduced additive white Gaussian noise $\sim N(0, \sigma_{noise})$ at two stages of the visual pipeline – external noise in the image, and internal noise in the LGN layer. Figure 5a shows the performance of the system in terms of the mean R^2_{max} across all neurones, plotted against the signal to noise ratio (SNR) of the input.

The mean R^2_{max} parameter is a measure of how closely the receptive fields of the neurones resemble Gabor-like receptive fields found in V1. Assuming $R^2 = 0.5$ as an acceptable cut-off, the system performs considerably well up to an SNR of 3dB ($\sigma_{noise} = 0.5 \times \sigma_{signal}$), beyond which the receptive fields become too noisy. Furthermore, Figure 5b shows the Convergence Index (CI, see Methods) of the network at various SNR levels for both internal (left) and external (right) noise. In each case, the network converges to an asymptotically low value of CI, indicating that the system is driven to a stable synaptic weight distribution. Overall, these analyses demonstrate that our approach is robust to a large range of internal and external noise.

4. Discussion

We have demonstrated that unsupervised Hebbian learning from natural stereoscopic images leads to a population of neurones which are capable of encoding binocular disparity statistics in natural scenes. Our findings provide evidence that known mechanisms of neuronal plasticity (such as STDP) could play a key role in the encoding of the statistics of natural scenes in the early visual pathway.

Starting from random receptive-fields, our neural population converges to Gabor-like receptive fields found in V1 simple cells (see Video and Figure 2). This convergence is achieved by a Hebbian pruning of the initial densely connected dendritic receptive field of size $3° \times 3°$. The mean size of the converged binocular receptive fields when described using a Gabor fit is about $1°$ (1σ contour of the Gaussian envelope). Although in the reported foveal population we started with a $3° \times 3°$ receptive field, our simulations with an initial receptive field size of $6° \times 6°$ yielded very similar results. Thus, the pruning process is independent of the size of the initial dendritic tree, and yields converged receptive fields which correspond to about 3 times the size of the central ON/OFF region of the retinal ganglion receptive fields.

When tested with multiple presentations of RDS stimuli, these converged neurones show selectivity to zero, near and far disparities; which is in line with electrophysiological findings[46], and gives the population the theoretical capacity to encode most realistic disparity values. This is also reflected in the range of disparities given by both the phase and position parameters extracted from the DTCs of our neural population (between $-0.5°$ and $0.5°$, see Figure 3d),

which resemble those reported in electrophysiological recordings in monkeys[42]. Interestingly, disparity tunings in this range are also reported when unsupervised[16] and supervised[17] coding schemes are applied to natural stereoscopic images. This reflects the fact that most disparities in natural scenes lie within this range[24].

We also find that the receptive fields of most converged neurones are binocular and show strong correlations between properties such as the orientation, size and frequency in the two eyes (Figure 3, Supplementary Figure S1). This suggests that the selectivity of the neurones in our model is influenced by the inter-eye correlations which primarily arise from the 3-D structure of the scene, despite competition from within-eye correlations which are mainly related to luminance-field statistics. This point is further illustrated by our simulation of the oculomotor behaviour in strabismic amblyopes where the inputs are binocular, but the inter-ocular correlations are weakened due to uncoordinated eye fixations. We find that this population of neurones also converges to Gabor-like receptive fields, but the majority of the neurones are monocular. This is not altogether surprising because STDP learning from monocular images is known to produce oriented Gabor-like receptive fields[37,47]. These results also echo findings in monkeys reared with amblyopic deficits[48,49] where the binocularity of V1 neurones is found to be severely restricted. Furthermore, we prove that our model is capable of extracting these binocular correlations under both internal and external noise up to SNR values of 3dB. Very similar results are obtained on running the model on other stereoscopic datasets such as KITTI (link).

Through retinotopic simulations, we give two further examples which show that when allowed to learn from specific regions of the visual field, our model produces population biases which correspond closely to biases reported in 3D natural statistics and electrophysiology. In the first example, we demonstrate that neurones trained on stimuli from peripheral eccentricities are selective to a larger range of horizontal and vertical disparities[25] when compared with neurones trained on foveal stimuli. We also find that all populations encode a larger range of horizontal disparity as compared to vertical disparity, once again a consequence of the simple fact that human eyes are separated along the horizontal axis while being aligned vertically. In the second example we use two retinotopic populations, one trained on the upper hemifield and the other on the lower hemifield, to confirm the intuitive bias that objects in the lower hemi-field are

more likely to be closer than fixation and objects in the upper hemifield are likely to be further away. This bias has been shown to be true in data collected from free-viewing tasks and is reflected in neural responses at both the single-unit[24] and macroscopic[50] levels. These simulations suggest that retinotopic learning based on a Hebbian approach could explain how biases in natural statistics can transfer to neural populations in the early visual system.

Few studies have previously characterised the relationship between natural statistics and binocular disparity responses in the visual cortex. Based on their approach to learning, they can roughly be classified under two main headings – unsupervised strategies[15,16] which learn from unlabelled stimuli, and supervised approaches[17,18] which use labelled data to learn specific properties of the scene. Whilst results from unsupervised learning can be interpreted as possible encoding schemes for the stimuli, supervised approaches optimise model performance on specific tasks such as disparity or contrast discrimination. Both these approaches solve an optimisation problem which may involve global properties like entropy and mutual information, or task-specific metrics such as classification accuracy. Although our model does not require supervision, it differs critically from traditional unsupervised approaches as it does not necessarily converge towards a static global optimum - instead, it can be interpreted as a self-regulating coincidence detector based on causal Hebbian updates. In our case, the convergence towards an optimum is a result of the stability of natural statistics across samples.

A commonly used unsupervised approach to studying the encoding of 3D natural statistics is the Independent Component Analysis (ICA)[15,16]. ICA is a form of efficient coding which maximises the mutual information between the stimuli and the activity of the encoding units. There are certain critical differences which separate our results from those obtained using ICA. First, Hebbian learning could potentially be sub-optimal when viewed in terms of global objective functions such as sparseness or mutual information. Second, Hebbian learning is usually unable to converge to features which do not recur in the input, whereas efficient coding approaches learn the optimal basis representations of the data which could be very different from any individual stimulus. As an example, the converged population in our model contains very few neurones with antiphase receptive fields in the two eyes. This is in line with electrophysiological data collected in monkeys, cats and the barn owl[42]. This result is also very intuitive because in everyday experience, the probability of a given object projecting

completely antiphase features in the two eyes is very low. In contrast, ICA models usually predict a higher number of antiphase receptive fields in the neural population[16]. Thus, although ICA representations offer a highly optimal encoding of the data in terms of information transfer, they are inconsistent with electrophysiological findings which often include neurones which are sub-optimal in terms of any particular coding strategy[51].

As initially pointed out by Barlow[2], although various encoding schemes offer explanations as to how scene statistics may be represented in early processing, the precise nature of the encoding must be capable of driving upstream processes which ultimately determine behaviour. Recently, the effect of natural 3D statistics on behaviour has been modelled using labelled datasets in conjunction with task-optimised filters[17] and feedforward convolutional networks[18]. In Supplementary Figure S2a, we show the result of training two very simple classifiers on the activity of the converged foveal population from our model using RDS stimuli at various disparities between $-1.5°$ and $1.5°$. The first, a decoder based on linear discriminant analysis (green), performs well above chance (dashed red), especially for realistic values of disparity between $-0.5°$ and $0.5°$. Interestingly, when the complexity of the decoder is increased slightly by including quadratic terms (blue), one observes a substantial increase in discrimination performance. This is not surprising as our neurones are linear units similar to simple cells, and a non-linear processing of their activity makes the decoding units conceptually closer to sharply tuned complex cells[52–54]. These results show that disparity encoding in the early visual system does not need supervised training, and an unsupervised feedforward Hebbian process can lead to neural units whose responses can be interpreted in terms of 3D percept through downstream processing. Furthermore, in Supplementary Figure S2b we demonstrate that the biases for near/far disparities shown by the lower/upper hemifield trained neural populations (see Figure 4c) are transferred to the decoding performance of the classifiers. Thus, our model predicts that biases in early binocular neural populations can drive downstream disparity discrimination biases– a property which has indeed been reported in human subjects[24].

We have shown that a simple Hebbian learning rule expressed through STDP could provide a plausible mechanism for the encoding of natural statistics in the early visual system. This form of learning through coincidence detection could present a more intuitive explanation for the emergence of representations found in the early visual system and address crucial discrepancies

between efficiency based approaches and the sub-optimality observed in electrophysiological studies[51]. Although this study demonstrates the relevance of Hebbian plasticity in binocular vision, this approach could easily be extended to other sensory modalities, and even multi-modal representations in the early sensory cortex. Finally, we would like to point out that although the present work focuses on the encoding properties of model, the coincidence based update rule also makes it possible to study the time-course of changes in the connectivity of individual units/neurones. For computational neuroscientists, this offers a critical advantage over conventional optimisation-based approaches as it allows for models which track cortical plasticity over time – such as studies on development, critical-period plasticity, and perceptual learning.

References

1. Hebbs, D. *The organization of behavior*. (Wiley and Sons, 1949).
2. Barlow, H. in *Sensory Communication* 217–234 (The MIT Press, 1961). doi:10.7551/mitpress/9780262518420.003.0013
3. Barlow, H. The exploitation of regularities in the environment by the brain. *Behav. Brain Sci.* **24,** 602–607 (2001).
4. Barlow, H. & Földiák, P. Adaptation and decorrelation in the cortex. in *The computing neuron* 54–72 (Addison-Wesley Longman Publishing Co., Inc., 1989).
5. Atick, J. J. & Redlich, A. N. What Does the Retina Know about Natural Scenes? *Neural Comput.* **4,** 196–210 (1992).
6. Zhaoping, L. Theoretical understanding of the early visual processes by data compression and data selection. *Netw. Comput. Neural Syst.* **17,** 301–334 (2006).
7. Geisler, W. Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* **59,** 167–92 (2008).
8. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381,** 607–609 (1996).
9. Olshausen, B. A. & Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.* **37,** 3311–3325 (1997).
10. Karklin, Y. & Lewicki, M. S. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature* **457,** 83–6 (2009).
11. Okazawa, G., Tajima, S. & Komatsu, H. Image statistics underlying natural texture selectivity of neurons in macaque V4. *Proc. Natl. Acad. Sci. U. S. A.* **112,** E351-60 (2015).
12. Brainard, D. Color constancy in the nearly natural image. 2. Achromatic loci. *J. Opt. Soc. Am. A* **15,** 307 (1998).

13. Thomson, M. & Foster, D. Role of second- and third-order statistics in the discriminability of natural images. *J. Opt. Soc. Am. A* **14,** 2081 (1997).

14. Webster, M. & Mollon, J. Adaptation and the color statistics of natural images. *Vision Res.* **37,** 3283–3298 (1997).

15. Hoyer, P. & Hyvärinen, A. Independent component analysis applied to feature extraction from colour and stereo images. *Netw. Comput. Neural Syst.* **11,** 191–210 (2000).

16. Hunter, D. & Hibbard, P. Distribution of independent components of binocular natural images. *J. Vis.* **15,** 6 (2015).

17. Burge, J. & Geisler, W. Optimal disparity estimation in natural stereo images. *J. Vis.* **14,** 1 (2014).

18. Goncalves, N. & Welchman, A. 'What Not' Detectors Help the Brain See in Depth. *Curr. Biol.* (2017). doi:10.1016/j.cub.2017.03.074

19. Stanley, G. Reading and writing the neural code. *Nat. Neurosci.* **16,** 259–263 (2013).

20. Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18,** 10464–72 (1998).

21. Caporale, N. & Dan, Y. Spike Timing–Dependent Plasticity: A Hebbian Learning Rule. *Annu. Rev. Neurosci.* **31,** 25–46 (2008).

22. Anzai, A., Ohzawa, I. & Freeman, R. D. Neural Mechanisms for Processing Binocular Information I. Simple Cells. *J Neurophysiol* **82,** 891–908 (1999).

23. Prince, S. J. D., Pointon, A. D., Cumming, B. G. & Parker, A. J. Quantitative Analysis of the Responses of V1 Neurons to Horizontal Disparity in Dynamic Random-Dot Stereograms. *J Neurophysiol* **87,** 191–208 (2002).

24. Sprague, W., Cooper, E., Tosic, I. & Banks, M. Stereopsis is adaptive for the natural environment. *Sci. Adv.* **1,** e1400254–e1400254 (2015).

25. Durand, J.-B., Zhu, S., Celebrini, S. & Trotter, Y. Neurons in Parafoveal Areas V1 and V2 Encode Vertical and Horizontal Disparities. *J. Neurophysiol.* **88,** 2874 LP-2879 (2002).

26. König, P., Engel, A. & Singer, W. Integrator or coincidence detector? The role of the cortical neuron revisited. *Trends Neurosci.* **19,** 130–137 (1996).

27. Brette, R. Computing with Neural Synchrony. *PLoS Comput. Biol.* **8,** e1002561 (2012).

28. Solomon, S. G., White, A. J. R. & Martin, P. R. Extraclassical Receptive Field Properties of Parvocellular, Magnocellular, and Koniocellular Cells in the Primate Lateral Geniculate Nucleus. *J. Neurosci.* **22,** 338 LP-349 (2002).

29. Masquelier, T. & Thorpe, S. J. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* **3,** e31 (2007).

30. Albrecht, D., Geisler, W., Frazor, R. & Crane, A. Visual Cortex Neurons of Monkeys and Cats: Temporal Dynamics of the Contrast Response Function. *J. Neurophysiol.* **88,** (2002).

31. Gawne, T., Kjaer, T. & Richmond, B. Latency: another potential code for feature binding

in striate cortex. *J. Neurophysiol.* **76,** (1996).

32. Gollisch, T. & Meister, M. Rapid Neural Coding in the Retina with Relative Spike Latencies. *Science (80-.).* **319,** (2008).

33. Celebrini, S., Thorpe, S., Trotter, Y. & Imbert, M. Dynamics of orientation coding in area V1 of the awake primate. *Vis. Neurosci.* **10,** 811–825 (1993).

34. Shriki, O., Kohn, A. & Shamir, M. Fast Coding of Orientation in Primary Visual Cortex. *PLoS Comput. Biol.* **8,** e1002536 (2012).

35. Thorpe, S., Delorme, A. & Van Rullen, R. Spike-based strategies for rapid processing. *Neural Networks* **14,** 715–725 (2001).

36. Masquelier, T. & Thorpe, S. J. Learning to recognize objects using waves of spikes and Spike Timing-Dependent Plasticity. in *The 2010 International Joint Conference on Neural Networks (IJCNN)* 1–8 (IEEE, 2010). doi:10.1109/IJCNN.2010.5596934

37. Masquelier, T. Relative spike time coding and STDP-based orientation selectivity in the early visual system in natural continuous and saccadic vision: a computational model. *J. Comput. Neurosci.* **32,** 425–441 (2012).

38. Tanaka, K. Organization of geniculate inputs to visual cortical cells in the cat. *Vision Res.* **25,** 357–364 (1985).

39. Ferster, D. & Lindström, S. An intracellular analysis of geniculo-cortical connectivity in area 17 of the cat. *J. Physiol.* **342,** 181–215 (1983).

40. Gütig, R., Aharonov, R., Rotter, S. & Sompolinsky, H. Learning Input Correlations through Nonlinear Temporally Asymmetric Hebbian Plasticity. *J. Neurosci.* **23,** 3697–3714 (2003).

41. Ohzawa, I. & Freeman, R. The binocular organization of complex cells in the cat's visual cortex. *J. Neurophysiol.* **56,** 243–59 (1986).

42. Prince, S. J. D., Cumming, B. G. & Parker, A. J. Range and Mechanism of Encoding of Horizontal Disparity in Macaque V1. *J. Neurophysiol.* **87,** 209 LP-221 (2002).

43. Freeman, R. & Ohzawa, I. Development of binocular vision in the kitten's striate cortex. *J. Neurosci.* **12,** (1992).

44. Durand, J.-B., Celebrini, S. & Trotter, Y. Neural Bases of Stereopsis across Visual Field of the Alert Macaque Monkey. *Cereb. Cortex* **17,** 1260–1273 (2007).

45. Hibbard, P. & Bouzit, S. Stereoscopic correspondence for ambiguous targets is affected by elevation and fixation distance. *Spat. Vis.* **18,** 399–411 (2005).

46. Poggio, G. F., Motter, B. C., Squatrito, S. & Trotter, Y. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Res.* **25,** 397–406 (1985).

47. Delorme, A., Perrinet, L. & Thorpe, S. J. Networks of integrate-and-fire neurons using Rank Order Coding B: Spike timing dependent plasticity and emergence of orientation selectivity. *Neurocomputing* **38–40,** 539–545 (2001).

48. Kiorpes, L., Kiper, D. C., O'Keefe, L. P., Cavanaugh, J. R. & Movshon, J. A. Neuronal Correlates of Amblyopia in the Visual Cortex of Macaque Monkeys with Experimental

Strabismus and Anisometropia. *J. Neurosci.* **18,** 6411 LP-6424 (1998).

49. Movshon, J. *et al.* Effects of early unilateral blur on the macaque's visual system. III. Physiological observations. *J. Neurosci.* **7,** 1340–51 (1987).

50. Nasr, S. & Tootell, R. Visual field biases for near and far stimuli in disparity selective columns in human visual cortex. *Neuroimage* (2016). doi:https://doi.org/10.1016/j.neuroimage.2016.09.012

51. Ringach, D. Spatial Structure and Symmetry of Simple-Cell Receptive Fields in Macaque Primary Visual Cortex. *J. Neurophysiol.* **88,** (2002).

52. Ohzawa, I., Deangelis, G. & Freeman, R. Encoding of Binocular Disparity by Complex Cells in the Cat's Visual Cortex. *J. Neurophysiol.* **77,** 2879–2909 (1997).

53. Cumming, B. G. & DeAngelis, G. C. The physiology of stereopsis. *Annu. Rev. Neurosci.* **24,** 203–38 (2001).

54. Henriksen, S., Cumming, B. G. & Read, J. C. A. A Single Mechanism Can Account for Human Perception of Depth in Mixed Correlation Random Dot Stereograms. *PLoS Comput. Biol.* **12,** e1004906 (2016).

Acknowledgments

We thank Simon J. Thorpe and Lionel Nowak for their comments on the study. This research was supported by an IDEX Emergence (ANR-11-IDEX-0002-02, BIS-VISION) and an ANR JCJC (ANR-16-CE37-0002-01, 3D3M) grants to B.R.C.

Contributions

T.C., B.R.C. and T.M. conceived the studies and discussed the theoretical implications of the results. T.C. performed the modelling, analysed the results and wrote the paper in discussion with B.R.C. and T.M. A.M. contributed to the coding of the model.

Competing financial interests

None.

Materials and Correspondence

Request for code and other correspondence should be addressed to T.C. or B.R.C.