Systems/Circuits

Emergence of Binocular Disparity Selectivity through Hebbian Learning

Tushar Chauhan,[1,2] Timothée Masquelier,[1,2] Alexandre Montlibert,[1,2] and Benoit R. Cottereau[1,2]
[1]Centre de Recherche Cerveau et Cognition, Université de Toulouse, 31052 Toulouse, France, and [2]Centre National de la Recherche Scientifique, 31055 Toulouse, France

Neural selectivity in the early visual cortex strongly reflects the statistics of our environment (Barlow, 2001; Geisler, 2008). Although this has been described extensively in literature through various encoding hypotheses (Barlow and Földiák, 1989; Atick and Redlich, 1992; Olshausen and Field, 1996), an explanation as to how the cortex might develop the computational architecture 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 how 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, 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 behavior. Furthermore, we show that the receptive fields we obtain are closer in structure to electrophysiological data reported in macaques than those predicted by normative encoding schemes (Ringach, 2002). We also demonstrate the robustness of our model to the input dataset, noise at various processing stages, and internal parameter variation. Together, our modeling 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.

Key words: binocular vision; disparity selectivity; emergence; Hebbian learning; natural statistics; STDP

Significance Statement

Neural selectivity in the early visual cortex is often explained through encoding schemes that postulate that the computational aim of early sensory processing is to use the least possible resources (neurons, energy) to code the most informative features of the stimulus (information efficiency). In this article, using stereo images of natural scenes, we demonstrate how a simple Hebbian rule can lead to the emergence of a disparity-selective neural population that not only shows realistic single-cell and population tunings, but also demonstrates how known biases in natural statistics may influence population encoding and downstream correlates of behavior. Our approach allows us to view early neural selectivity, not as an optimization problem, but as an emergent property driven by biological rules of plasticity.

Introduction

Sensory stimuli are not only the inputs to, but also shape the very process of, neural computation (Hebbs, 1949; Barlow, 1961). A

modern, more rigorous extension of this idea is the efficient coding theory (Barlow and Földiák, 1989; Atick and Redlich, 1992), which postulates that the computational aim of early sensory processing is to use the least possible resources (neurons, 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 organization and structure of the early cortex (Geisler, 2008). 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 shown that information-theoretic constraints do indeed predict localized, oriented, and bandpass representations akin to those reported in the early visual cortex (Olshausen and Field, 1996). Over the years, several studies have shown how

properties of natural images can not only explain neural selectivity (Olshausen and Field, 1996, 1997; Furmanski and Engel, 2000; Karklin and Lewicki, 2009; Samonds et al., 2012; Okazawa et al., 2015), but also predict behavior (Webster and Mollon, 1997; Howe and Purves, 2002; Geisler, 2008; Geisler and Perry, 2009; Girshick et al., 2011; Cooper and Norcia, 2014; Burge and Geisler, 2015; Sebastian et al., 2017).

Nearly all these studies rely on images of natural scenes acquired 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, it does not fully reflect how visual data are acquired by the human brain. Humans are binocular and use two simultaneous retinal projections that enable the sensing of disparity (differences in retinal projections in the two eyes). 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 (Hoyer and Hyvärinen, 2000; Burge and Geisler, 2014; Hunter and Hibbard, 2015; Goncalves and Welchman, 2017) have attempted to analyze the relationship between binocular projections of natural scenes and the properties of binocular neurons in the early visual system. Furthermore, these studies investigated the relationship between natural scenes and cortical selectivity in terms of a global optimization problem that is *solved* in the adult brain, leading to cortical structures that 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 (Stanley, 2013). Here, we show how simple coincidence detection based on spike timing-dependent plasticity (STDP; Bi and Poo, 1998; Caporale and Dan, 2008) could offer a biologically plausible mechanism for arriving at neural population responses close to those reported in the early visual system. We endow a neural network with a Hebbian STDP rule and find that an unsupervised exposure of this network to natural stereoscopic stimuli leads to a converged population that shows single-cell and population characteristics close to those reported in electrophysiological studies (Anzai et al., 1999; Durand et al., 2002; Prince et al., 2002b; Sprague et al., 2015). Moreover, the emergent receptive fields differ from those obtained by optimization-based methods and are more representative of those reported in the literature (Ringach, 2002). Thus, together, our findings suggest that the 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 simple biological coincidence detection (König et al., 1996; Brette, 2012; Masquelier, 2017) could explain the emergence of selectivity in early sensory and multimodal neural populations, both during and after development.

Materials and Methods

Datasets. To simulate binocular retinal input, we chose two datasets—the Hunter–Hibbard dataset and the KITTI database (available at http://www.cvlibs.net/datasets/kitti/). The main results are reported for the Hunter–Hibbard database of stereo images (Hunter and Hibbard, 2015; available at https://github.com/DavidWilliamHunter/Bivis under the Massachusetts Institute of Technology license, which guarantees free usage and distribution). This database consists of 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

with zero elevation (Fig. 1A). The distance between the two cameras was close to the human interocular separation (65 mm), and the two cameras were focused at realistic fixation points from 50 cm to several meters away from the cameras. The lenses had a relatively large field of view (20°), enabling them to capture binocular statistics across a sizeable part of the foveal and parafoveal visual field. Figure 1A also shows a red-cyan anaglyph of one representative scene from the database, while Figure 1B shows the images captured by the left and right cameras. Due to the geometry of the cameras, there are subtle differences in the two images (disparity), which provide important information about the 3-D structure of the scene. The mimicking of realistic acquisition geometry is crucial for capturing disparity statistics, which resemble those experienced by human observers. For comparison, we also report results using the KITTI database, which uses parallel cameras, a geometry that does not correspond to the human visual system (see Fig. 7 and related discussion for more details).

Input sampling. Random locations in the scenes were sampled to provide inputs to our model. In each simulation, the sampling was restricted to a specific region of interest (ROI). In this article, we present results from the following four main ROIs: foveal (eccentricities <3°); peripheral (eccentricities >6°); upper hemifield; and lower hemifield. Figure 1B shows the foveal ROI shaded in purple. Each sample consisted of two patches corresponding to the two eyes, both centered at the same random retinal coordinates. Figure 1B shows a random sample, with the left patch outlined in red and the right patch in green. The sizes of the patches varied with the ROI and can be interpreted as the initial dendritic receptive field of the V1 neurons, which was subsequently pruned by STDP. In the simulations presented in this article, the foveal patches were 3° × 3°, while the peripheral patches were 6° × 6°. In each simulation, 100,000 input samples were used.

In Figure 3B, we present the results of our model under simulations of fixations in individuals with strabismic amblyopia. Strabismus is a misalignment of the two eyes during fixation and can lead to strabismic amblyopia if left unmanaged during childhood. While strabismic misalignment is more accurately represented by an offset of the two camera axes during dataset acquisition, in the absence of this possibility, it was approximated by using a misaligned sampling scheme. The left and right patches of each sample were no longer constrained to be centered at the same retinal coordinates (while still keeping within the ROI). This allowed us to simulate retinal activations caused by variable fixations of the amblyopic eye.

Modeling of the retinogeniculate pathway. The computational pipeline used by the model is shown in Figure 1C. The 100,000 samples were presented sequentially. The first stage of the model implemented the processing in the retinogeniculate pathway. To simulate the computations performed by the retinal ganglion cells and the lateral geniculate nucleus (LGN), the patches were convolved with ON and OFF center-surround kernels, followed by half-wave rectification. The filters were implemented using difference of Gaussian kernels, which were normalized to zero response for uniform inputs, and a response of 1 for the maximally stimulating input. Since the model was to be driven by binocular luminance images, only the magnocellular pathway with achromatic receptive fields, which feed into the dorsal stream (implicated in depth and motion perception), was modeled. The receptive field sizes were chosen to reflect the size of representative LGN center-surround magnocellular receptive fields: 0.3°/1° (center/surround) for foveal simulations; and 1°/2° (center/surround) for peripheral simulations (Solomon et al., 2002). The activity of each kernel can approximately be interpreted as a retinotopic contrast map of the sample. Four maps were calculated, corresponding to ON- and OFF-center populations in the left and right monocular pathways. Further thresholding was applied such that, on average, only a small portion (the most active 10%) of the LGN units fired. This activity was converted to relative first-spike latencies through a monotonically decreasing function (a token function, $y = 1/x$, was chosen), but all monotonically decreasing functions are equivalent (Masquelier and Thorpe, 2007), thereby ensuring that the most active units fired first, while units with lower activity fired later or not at all. Latency-based encoding of stimulus properties has been reported extensively in the early visual system (Celebrini et al., 1993; Gawne et al., 1996;
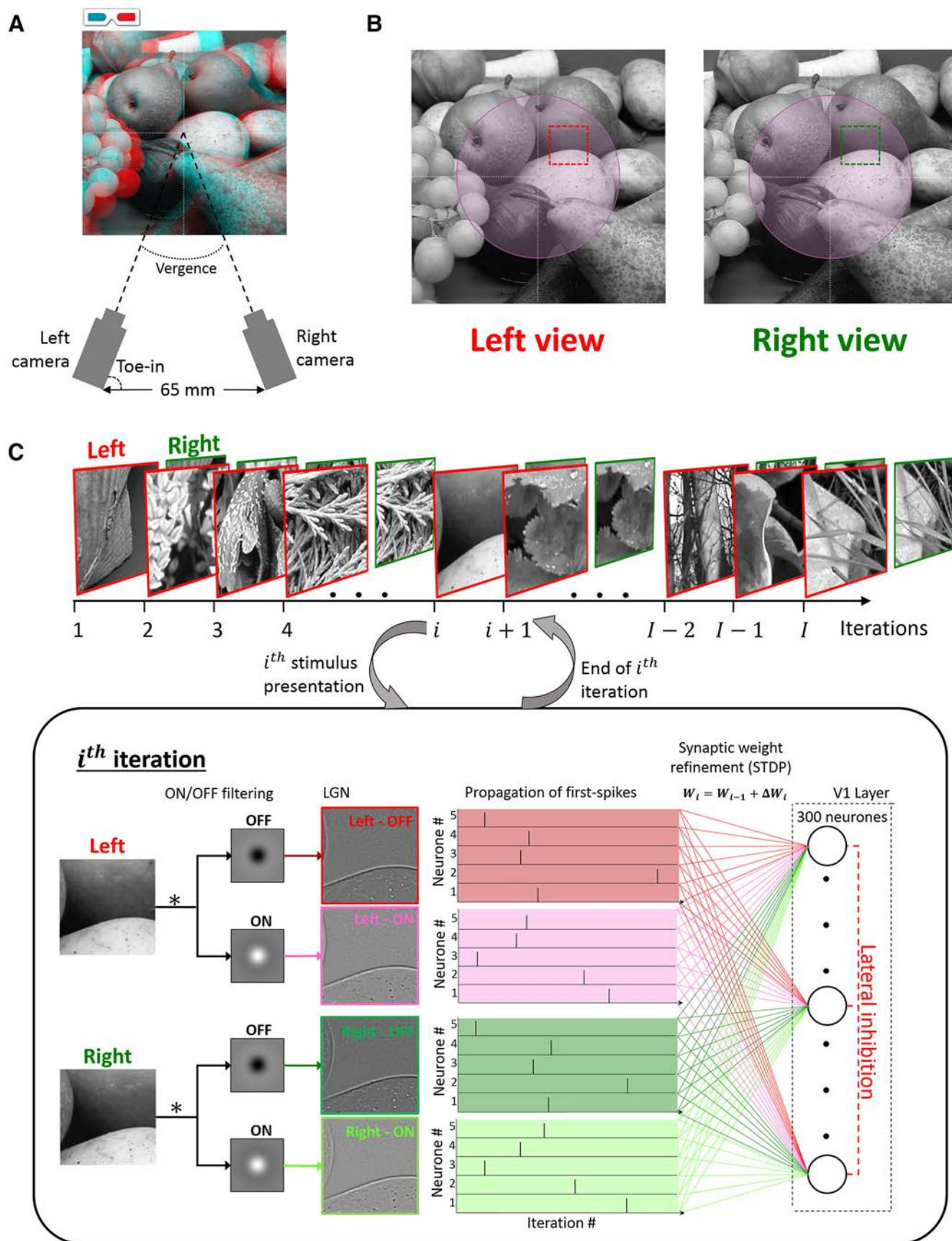
Figure 1. Processing pipeline. *A*, Acquisition geometry for the Hunter–Hibbard database. The setup mimicked the typical geometry of the human visual system (see more details in the text). This geometry is crucial for capturing realistic disparity statistics from natural scenes. A stereoscopic reconstruction (red-cyan anaglyph) of a sample from the dataset is also shown. *B*, Input images captured by the left and the right camera for the scene shown in *A*. The subtle differences in the images (disparity) provide important cues about the 3-D structure of the scene. Depending on the simulation, an ROI in the visual field was identified, and inputs were sampled only within this ROI (foveal ROI illustrated in purple). Each sample consisted of two (*Figure legend continues.*)

Albrecht et al., 2002; Gollisch and Meister, 2008; Shriki et al., 2012) and allows for fast and efficient networks describing early visual selectivity (Thorpe et al., 2001; Masquelier and Thorpe, 2010; Masquelier, 2012). 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 samples were presented to the network sequentially (Fig. 1C). For each sample, the first spikes from the most active 10% of LGN neurons were propagated through plastic synapses to a V1 population of 300 integrate-and-fire neurons. Furthermore, a winner-take-all inhibition scheme (Maass, 2000) was implemented such that, if any V1 neuron fired during a certain iteration, it simultaneously prevented other neurons in the population from firing until the next sample was processed. After each iteration, the synaptic weights for the first V1 neuron to fire were updated using an STDP rule, and the membrane potentials of the entire population were reset to zero. This scheme leads to a sparse neural population where the probability of any two neurons learning the same feature is greatly reduced. The initial dendritic receptive fields of the neurons were three times the size of the LGN filters (foveal, 3° × 3°; peripheral, 6° × 6°). At the start, each neuron was fully connected to all LGN afferents within its receptive field through synapses with randomly assigned weights between 0 and 1. The weights were restricted between 0 and 1 throughout the simulation. The non-negative values of the weights reflect the fact that thalamic connections to V1 are excitatory in nature (Ferster and Lindström, 1983; Tanaka, 1985).

The box in Figure 1C shows the processing of the *i*th sample presented to the model. Each iteration began with the calculation of the ON and OFF LGN activity maps for the two eyes. This activity was thresholded, converted to spike latencies, and propagated through the network (for more details, see Modeling of the retinogeniculate pathway). The thresholding process allowed spikes from the fastest 10% of LGN neurons to propagate through the network. The propagated LGN spikes contributed to an increase in the membrane potential of V1 neurons (excitatory postsynaptic potentials or EPSPs) until one of the V1 membrane potentials reached threshold, resulting in a postsynaptic spike and inhibition of all other V1 neurons until the next iteration. The EPSP conducted by the synapse connecting the *m*th LGN neuron and the *n*th V1 neuron was taken as the weight of the synaptic connection itself (say w_{mn}). This allows us to write a simple expression for the membrane potential, $E_n(t)$, of the *n*th V1 neuron at time t within the following iteration:

$$E_n(t) = \begin{cases} \sum_{m \in \text{LGN}} w_{mn} \cdot H(t - t_m), & t < \min_t \{t | \max_{n \in \text{V1}} E_n(t) \geq \Theta\}, \\ \\ 0, & \text{otherwise} \end{cases}$$

where t_m is the spiking time of the *m*th LGN neuron, $H(t)$ is the Heaviside step function, and Θ is the threshold of the V1 neurons (assumed to be a constant for the entire population). The expression $\min_t \{t | \max_{n \in \text{V1}} E_n(t) \geq \Theta\}$ denotes the timing of the first spike in the V1 layer. The membrane potentials were calculated up to this time

---

←

(*Figure legend continued.*) patches with equal retinal coordinates in the left and the right eye (illustrated in red and green, respectively). All simulations presented in this article were run using $N = 100,000$ samples each. *C*, The processing pipeline. The samples were processed sequentially, one per iteration. The processing of the *i*th sample is shown in the box. First, the ON/OFF processing in the LGN was modeled using difference-of-Gaussian filters in the two eyes—leading to left-ON, left-OFF, right-ON, and right-OFF maps. The activations of the four maps were then converted to first-spike relative latencies using a simple inverse operation (first-spike latencies for only five neurons per LGN map shown in the figure). After thresholding, the earliest 10% of first spikes were allowed to propagate to the V1 layer. The V1 layer consisted of 300 neurons connected to the LGN maps through Hebbian synapses endowed with STDP. A winner-take-all inhibition scheme was also implemented to ensure no more than one neuron fired per iteration. The first spikes from the *i*th sample were propagated through the synapses, and this was followed by a change in synaptic weights dictated by the STDP algorithm (see text for details about the implemented STDP model). This concluded the processing of the *i*th sample and was followed by the $(i + 1)^{th}$ sample, which was processed in a similar fashion.

point, after which the winner-take-all inhibition scheme was triggered and all membrane potentials were reset to zero.

After the LGN spike propagation, the synaptic weights were updated using an unsupervised multiplicative STDP rule (Gütig et al., 2003). For a synapse connecting a pair of presynaptic and postsynaptic neurons, this rule is classically described as follows:

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

Here, a presynaptic and postsynaptic pair of spikes with a time difference (Δt) introduces a change in the weight (Δw) of the synapse, which is given by the product of a temporal filter K [typically, $K(\Delta t, \tau) = e^{-|\Delta t|/\tau}$] and a power function of its current weight (w). In our implementation, the efficiency and speed of calculations was greatly increased by making the windowing filter K infinitely wide (equivalent to assuming $\tau_\pm \to \infty$, or $K = 1$). This, however, does not imply that there was no temporal windowing in the model, as the thresholding at the LGN stage allowed only the fastest 10% of spikes to propagate in each iteration. When a postsynaptic spike occurs shortly after a presynaptic spike ($\Delta t > 0$), there is a strengthening of the synapse, also called long-term potentiation (LTP). Conversely, when the postsynaptic spike occurs before the presynaptic spike ($\Delta t \leq 0$), or in the absence of a presynaptic 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 non-negative ($\alpha^\pm \geq 0$) 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 nonlinearity in the LTP and LTD update rules. In practice, a nonlinearity 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). Note that, due to the winner-take-all inhibition scheme, the STDP update rule only applied to the afferent synapses of the first V1 neuron that fired. This prevented multiple neurons from learning very similar features and was instrumental in introducing sparsity in the limited population of 300 neurons.

For the network used in the present study, the learning rates were fixed with $\alpha^+ = 5 \times 10^{-3}$ and $\alpha^+/\alpha^- = (4/3)$. The rate ratio α^+/α^- is crucial to the stability of the network and was chosen based on previous work demonstrating STDP-based visual feature learning (Masquelier and Thorpe, 2007). Our simulations show that the results presented in this article are robust for a large range of this parameter (see Fig. 6). For these values of the learning rate, the threshold of the V1 neurons was fixed through trial and error at $\Theta = 18$. This value was left unmodified for all the results presented in this article. 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 we used an almost additive LTD rule ($\mu^- = 0.05$) to ensure the pruning of continuously depressed synapses. In both LTP and LTD updates, the weights were maintained in the range $w \in [0,1]$. It is to be noted that our model was designed to converge to continuous and graded weight distributions (as opposed to binary 0/1 weights), and thus the neuronal threshold should not be interpreted as the maximal number of full-strength connections allowed per neuron. Finally, the STDP and inhibition rules were active only during the learning phase, and all subsequent testing of the converged populations involved simultaneous spiking of all V1 neurons without any synaptic learning.

Analysis of converged receptive fields. The postconvergence receptive fields of the STDP neurons were approximated by a linear summation of the afferent LGN receptive fields weighted by the converged synaptic weights. The receptive field ξ_i of the *i*th V1 neurone was estimated by the following:

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

where ψ_j is the receptive field of the *j*th LGN afferent, and w_{ij} is the weight of the synapse connecting the *j*th afferent to the *i*th V1 neurone. The receptive fields calculated using this method are, in principle, similar to pointwise estimates of the receptive field calculated by electrophysiolo-

gists. To better characterize the properties of these receptive fields, they were fitted by 2-D Gabor functions. These Gabor functions are given by the following:

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

where k is a measure of the sensitivity of the Gabor function; $\boldsymbol{S}$ is a 2-D sinusoid propagating with frequency f, and phase ϕ in the direction θ; and $\boldsymbol{G}$ is a 2-D Gaussian envelope centered at $\boldsymbol{x}$, with size parameter σ, also oriented at an angle θ. A multistart global search algorithm was used to estimate optimal parameters in each eye for each neuron. 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 neuron, the goodness of the Gabor fit was characterized using the maximum of the R^2 goodness-of-fit values in the two eyes, as follows:

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

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

After the fitting procedure, the Gabor parameters were then used to characterize the structure and symmetry of the receptive field in the dominant eye (defined as the eye with the higher value of the Gabor sensitivity parameter k). This was done by calculating the symmetry-invariant coordinates proposed by Ringach (2002). These coordinates [referred to as Ringach coordinates (or RCs from here on] are derived from the Gabor parameters by the following:

$$\begin{bmatrix} n_x \\ n_y \end{bmatrix} = \begin{bmatrix} \sigma_x \\ \sigma_y \end{bmatrix} f,$$

where σ_x is the spread of the Gaussian envelope along the direction of sinusoidal propagation, σ_y is the Gaussian spread perpendicular to σ_x, and f is the frequency of the sinusoid. While n_x is a measure of the number of lobes in the receptive field, n_y reflects the elongation of the receptive field perpendicular to the sinusoidal carrier. Typical cells reported in cat and monkey have both RCs (n_x and n_y) <0.5 (Ringach, 2002).

The responses of the converged neurons to disparity [disparity tuning curves (DTCs)] were characterized using the following two methodologies: binocular correlation (BC) of the left and right receptive fields; and estimation of responses to random dot stereograms (RDSs). In the BC method, we estimated the DTCs by cross-correlating the receptive fields in the left and the right eye along a given direction. For a given neuron, these DTCs assume that all disparities will elicit responses (i.e., lead to a membrane potential greater than the membrane threshold), and thus represent the theoretical upper limit of the disparity response of the cell. In the RDS method, we estimated the DTCs by calculating the responses of the converged neurons to RDSs at various disparities. The RDS patterns used for testing consisted of an equal number of white and black dots on a gray background. The dots were perfectly correlated between the two eyes, with a dot size of ~15 min 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 responses were quantified by a binocularity interaction index (BII; Ohzawa and Freeman, 1986a; Prince et al., 2002b) given by the following:

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

where $R_{\max}$ and $R_{\min}$ are maximum and minimum responses on the RDS 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.

The phase and position disparity were estimated by methods commonly applied in electrophysiological studies—one-dimensional Gabor functions were fitted to the DTCs (for this estimation, we used the BC estimates as they are less subject to noise), and the phase of the sinusoid and the offset of the Gaussian envelope from zero disparity were taken as estimates of the phase and position disparities, respectively.

Furthermore, the symmetry of the DTCs was evaluated using the symmetry-phase (SP) metric proposed by Read and Cumming (2004). This becomes especially relevant when the shape of the DTC is close to a Gaussian curve, as the phase of the fitted Gabor becomes less reliable. The DTC [say $D(\delta)$, where δ is the disparity] was first decomposed into odd (D_-) and even (D_+) components as follows:

$$D_\pm = 0.5 \times [D(\delta) \pm D(-\delta + 2\delta_c)],$$

where $\delta_c = \dfrac{\int_{-\infty}^{\infty} |D(\delta)| \cdot \delta \cdot d\delta}{\int_{-\infty}^{\infty} |D(\delta)| \cdot d\delta}$ is the centroid of D. The SP metric was then

calculated as follows:

$$SP = \text{Tan}^{-1}(\xi_-, \xi_+),$$

where $\xi_\pm$ is the maximum deviation of $D_\pm$ from zero (including the sign), and $\text{Tan}^{-1}(y, x)$ is the four-quadrant arctangent function.

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

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

where N_{LGN} and N_{V1} are the number of LGN units and the number of the V1 neurons, respectively. The CI can be interpreted as a measure of the change in the synaptic weight distribution over time. Assuming a statistically stable input to a Hebbian network, the number of synapses undergoing a change in strength decreases with time. Thus, if our network is driven to convergence, CI should reduce asymptotically with time. To report the convergence for a single unit (say the ith V1 neuron), a modified form of the above index (CI_i) was defined as follows:

$$CI_i(t) = \frac{1}{N_{\text{LGN}}} \sum_{j \in \text{LGN}} |\Delta w_{ij}(t)|.$$

Robustness analysis. Biological systems show a considerable resilience to factors such as noise and internal parameter variations (Burge and Jaini, 2017). We tested the robustness of our model using the following three approaches: (1) by introducing timing jitter at both the input and the LGN level; (2) by varying a key internal parameter of our network (the ratio of the LTP rate to the LTD rate); and (3) by comparing results obtained using the Hunter–Hibbard dataset (realistic acquisition geometry) to results obtained using a dataset with a nonrealistic acquisition geometry (parallel cameras).

The robustness of the model to noise was tested by adding external (image) and internal (LGN) noise to the system. This simulated timing jitter in the firing latencies from both the retina and the LGN. Gaussian noise with a signal-to-noise ratio (SNR) varying between −3 dB ($\sigma_{\text{noise}} = 2 \times \sigma_{\text{signal}}$) and 10 dB ($\sigma_{\text{noise}} = 0.1 \times \sigma_{\text{signal}}$) was used, with σ_{signal} for a given image being defined as the variance in pixel grayscale values. The performance of the network at various noise levels was characterized using the following four metrics: the CI; the $R^2_{\max}$ of Gabor fitting; the population disparity tuning; and the BII. While the CI can be used to evaluate the stability of the converged synaptic connectivity, the $R^2_{\max}$ shows how well the receptive fields can be characterized by Gabor functions. Together, the CI and $R^2_{\max}$ characterize the robustness of the network in terms of both the synaptic and the functional convergence. Furthermore, the disparity sensitivity was characterized using the following two metrics: the population disparity tuning calculated using the BC DTC (see Materials and Methods, Analysis of converged receptive fields); and the BII. The population tuning curves give the theoretical encoding capacity of the network, while the BII is a measure of the modulation of network activity as a response to disparity.

The ratio of LTD to LTP rates ($\lambda = \alpha^-/\alpha^+$) is a critical parameter for the stability of the model because it determines the number of pruned synapses after convergence. If the probabilities of LTD and LTP events

are p^- and p^+, respectively, the learning rate γ can be approximated as follows:

$$\gamma \propto [p^+\alpha^+ + p^-\alpha^-].$$

Given our initial simulation values of the LTD and LTP rates, α_0^- and α_0^+, respectively, we defined $\alpha^- = (\alpha_0^-/\beta)$ and $\alpha^+ = \beta\alpha_0^+$, where the factor β was introduced to ensure that the two rates changed such that the overall learning rate γ remained stable. The effect of the LTP-to-LTD ratio on the convergence and functional stability of the network was then tested by running simulations for λ in logarithmic steps from 0.01 to 100 and estimating the CI, R^2_{max}, population disparity tuning, and BII parameters. This simulated models where the LTD rates were from 1/100 to 100 times the LTP rates.

After testing the robustness of our model to noise and internal parameter variations, the aim was to test the model on a second dataset. As mentioned earlier, the Hunter–Hibbard dataset used in all our simulations was chosen because it has an acquisition geometry close to the human visual system. Thus, for this analysis we chose the KITTI dataset, which was collected using a markedly different acquisition geometry. This dataset uses parallel cameras mounted 54 cm apart, leading to a highly nonecological, zero-vergence geometry. The aim of this analysis was to verify that the main results presented in this article are not specific to the dataset, and that the model is capable of capturing natural disparity statistics while being robust to acquisition geometry. Of course, in this case we did not expect the tunings of the converged neurons to match those reported in human electrophysiology (for more details, see Fig. 7 and related discussion).

Decoding of disparity. The ability of the converged network to encode disparity was tested through a decoding paradigm. RDS stimuli at 11 equally spaced disparities between $-1.5°$ and $1.5°$ were presented to the converged model, and the first-spike activity of the V1 layer was used to train, and subsequently test, linear and quadratic discriminant classifiers. The inhibition scheme used during the learning phase was turned off, allowing multiple neurons to fire in response to a single stimulus. If we denote V1 activity by the random variable X and disparity by the random variable Y, the probability that the disparity of the stimulus is y, given an observed pattern of V1 activity (say, x) can be written using the Bayes rule as follows:

$$P(Y=y|X=x) = \frac{P(X=x|Y=y)P(Y=y)}{\sum_k P(X=x|Y=k)P(Y=k)}.$$

In discriminant classifiers, the likelihood term is estimated by, for example, the following Gaussian density function:

$$P(X=x|Y=y) = \frac{1}{\sqrt{(2\pi)^{N_{V1}}|\Sigma_y|}} e^{-\frac{1}{2}(x-\mu_y)^T\Sigma_y^{-1}(x-\mu_y)},$$

where N_{V1} is the number of V1 neurons, and μ_y and Σ_y are the mean and covariance of the V1 activity for the disparity label y. In linear discriminant classifiers, only one covariance matrix (say, Σ) is assumed to describe the V1 activity for all labels (i.e., $\forall y, \Sigma_y = \Sigma$). In this case, the objective function $\delta_y(x)$ for the label y can be written as follows:

$$\delta_y(x) = \log P(Y=y) - \frac{1}{2}\mu_y^T\sum{}^{-1}\mu_y + x^T\sum{}^{-1}\mu_y.$$

This leads to discrimination boundaries that are a linear function of V1 activity (this can be verified by setting $\delta_{y1} = \delta_{y2}$ for a given value of x). On the other hand, quadratic discriminant classifiers allow each group to have a different covariance matrix. In this case, the objective function takes the following form:

$$\delta_y(x) = \log P(Y=y) - \frac{1}{2}\mu_y^T\sum_y{}^{-1}\mu_y + x^T\sum_y{}^{-1}\mu_y +$$

$$\frac{1}{2}x^T\sum_y{}^{-1}x - \frac{1}{2}\log|\sum_y|,$$
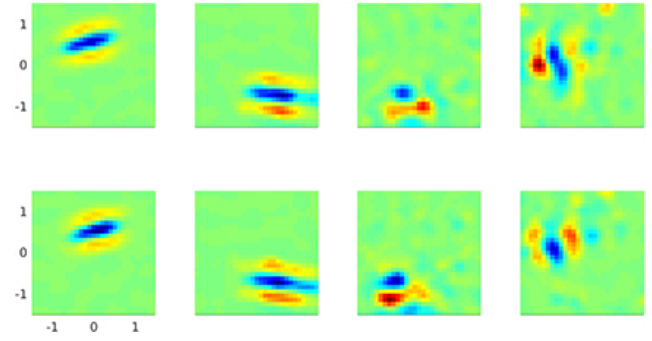
Movie 1. The emergence of receptive fields for neurones shown in Figure 2. The initial receptive fields are randomly connected to the LGN maps. As synapses are strengthened and pruned through learning, we observe the emergence of Gabor-like, binocular receptive fields.

leading to boundaries that are quadratic functions of the V1 activity.

We chose RDS stimuli because they do not contain any other information except disparity. A total of 10,000 stimuli per disparity were used, with a 70/30 training/testing split. Twenty-five-fold cross-validation testing was performed to ensure robust results. In this article, we report the detection probability (i.e., the probability of correct identification) at each disparity. It must be noted that here, decoding through classifiers is only an illustrative representation of perceptual responses as it is based on inputs from very early visual processes and ignores important interactions such as corticocortical connections and feedback.

Code accessibility. The code for the model is available on ModelDB (http://modeldb.yale.edu/245409).

Results

In this section, we present the results from simulations (10^5 iterations/simulation) where binocular stimuli from specific ROIs were presented to our model. In the model, retinal inputs were filtered through an LGN layer, which was in turn connected to a V1 layer consisting of 300 integrate-and-fire neurons through STDP-driven synapses (for details, see Materials and Methods). The main results are shown for the foveal ROI, while other ROIs (peripheral, upper hemifield, and lower hemifield) are used to illustrate retinotopic biases.

Foveal population

Starting from random receptive-fields, the foveally trained population converges to Gabor-like receptive fields found in V1 simple cells (Movie 1). This convergence is achieved by a Hebbian pruning of the initial densely connected dendritic receptive field of size 3° × 3°. The mean size of the converged binocular receptive fields when described using a Gabor fit is approximately 1° (average 1σ of the Gaussian envelope). Although in the reported foveal population we started with a 3° × 3° receptive field, simulations with an initial receptive field size of 6° × 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 approximately three times the size of the central ON/OFF region of the retinal ganglion receptive fields.

In Figure 2A, we present a representative unit from a population of 300 V1 neurons. The first row of Figure 2A shows the initial receptive field of the neuron. Since the weights were set to random values, the resulting receptive field has no specificity in terms of structure or excitatory/inhibitory subfields. The second
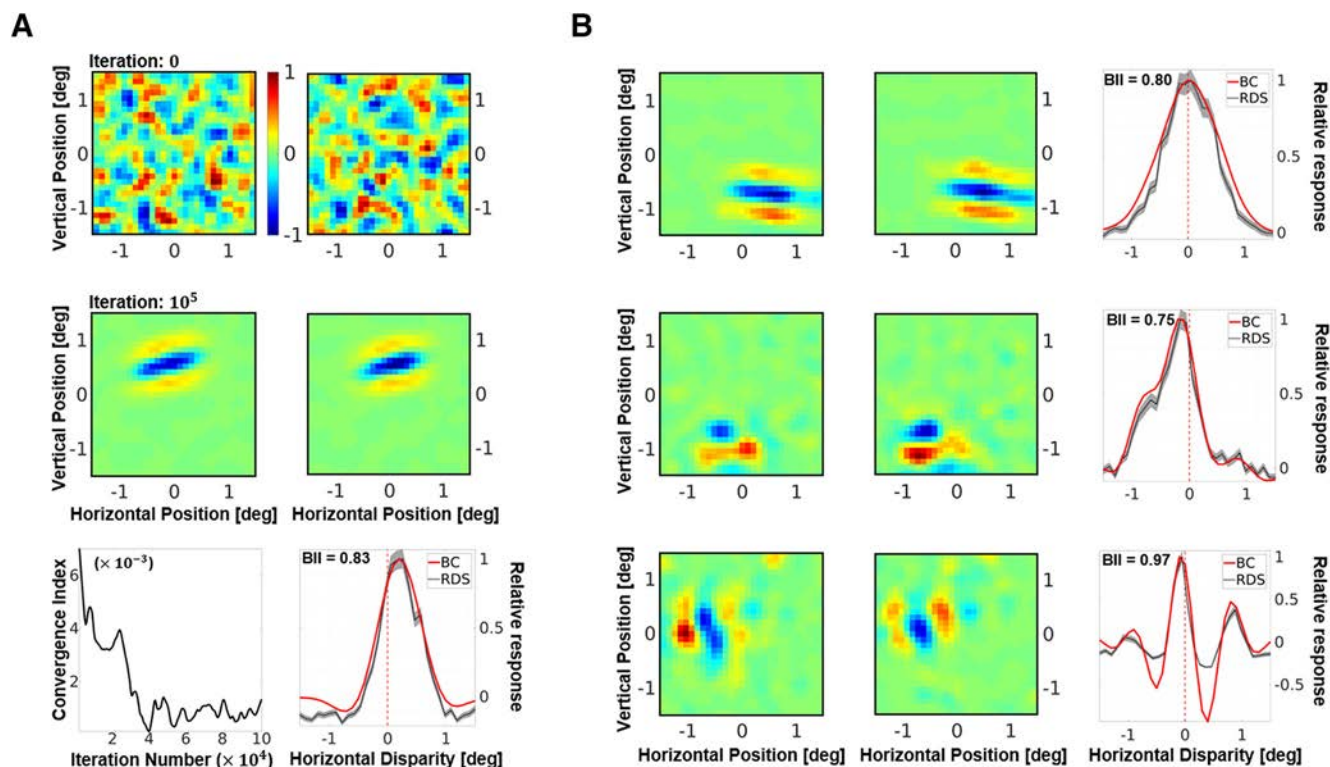
Figure 2. ***A***, A representative neuron from the simulated population. Receptive fields in the left and the right eye before (first row), and after (second row) presentation of 10^5 samples. The responses are scaled such that the maximum response is 1. The initial receptive fields are random, while the converged receptive fields show a Gabor-like structure. Bottom left, The CI, which is a measure of the mean change in synaptic strength over time. Bottom right, DTC estimates for horizontal disparity. The neuron shown here is tuned to uncrossed (or positive) disparities. The DTC estimates using two techniques are presented: binocular correlation of the left and right receptive fields (red solid line); and averaging responses to multiple ($N = 10^4$) presentations of RDS stimuli (gray dotted line). For comparison, both DTCs have been rescaled to have a peak response of 1. The SE for the RDS DTCs is shown as a gray envelope around the average responses. The BII calculated on the raw RDS DTCs is also indicated within the graph. BII is a measure of the sensitivity of the neuronal response to disparity variations, with values close to 0 indicating low sensitivity, and values close to 1 indicating high disparity sensitivity. See Materials and Methods for more details on the calculation of the receptive fields, DTCs, and BII. ***B***, Converged receptive fields and DTCs for neurons showing zero-tuned (first row), tuned near (second row), and phase-modulated (third row) responses to horizontal disparity.

row of the figure shows the receptive field of the same neuron after STDP learning from 10^5 samples (see also, Movie 1). The converged receptive field is local, bandpass and oriented. Clear ON and OFF regions can be observed, and the receptive field closely resembles classical Gabor-like receptive fields reported for simple cells (Hubel and Wiesel, 1962; Jones and Palmer, 1987). Figure 2*A* (bottom left) also shows the CI of the neuron, which characterizes 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 neuron with a binocular receptive field is often characterized in terms of its response to binocular disparity (Freeman and Ohzawa, 1992; Prince et al., 2002a,b; i.e., its DTC). Figure 2*A* (bottom right) shows the DTC estimates for the selected neuron, estimated using both the BC (solid red line) and the RDS (dotted gray line, with SE marked as a gray envelope) methods. The DTCs obtained from BC were very similar to those obtained through RDS stimuli. As the BC DTCs are, in general, less noisy, they were used to calculate all the disparity results presented in the article. This neuron shows a selectivity for positive or uncrossed disparities, indicating that it is more likely to respond to objects farther away from the observer than fixation. It also has a high BII (Ohzawa and Freeman, 1986a; Prince et al., 2002b) of 0.83, indicating that there is a substantial modulation of its neuronal activity by binocular disparity (BII values close to 0 indicate no modulation of neuronal response by disparity,

while values close to 1 indicate high sensitivity to disparity variations). Furthermore, Figure 2*B* shows three other examples of converged receptive fields and the corresponding DTCs (one neuron per row). The first neuron shows a tuning to zero disparity (fixation), the second neuron shows a tuning to crossed or negative disparity (objects closer than the fixation), and the third neuron shows a tuning curve that is typically attributed to phase-tuned receptive fields. The receptive fields have a variety of orientations and sizes and closely resemble 2-D 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 3*A* shows the goodness-of-fit R^2 parameter for the left and the right eye for all 300 neurons in the foveal population, with the color code corresponding to the value $R^2_{max} = \max \{R^2_{left}, R^2_{right}\}$. This value is maximal of the R^2 parameter across the two eyes, and indicates how well the Gabor model fits the receptive field regardless of whether it is monocular or binocular. A majority of the neurons lie along the diagonal in the top right quadrant, showing an excellent fit in both eyes, with only a few monocular neurons lying in the top left and bottom right quadrants. It is also interesting to note that, regardless of the degree of fit, the R^2 values for the left and the right eye are tightly correlated. This binocular correlation is nontrivial and cannot be attributed simply to learning from a stereoscopic dataset, as the interocular correlations in
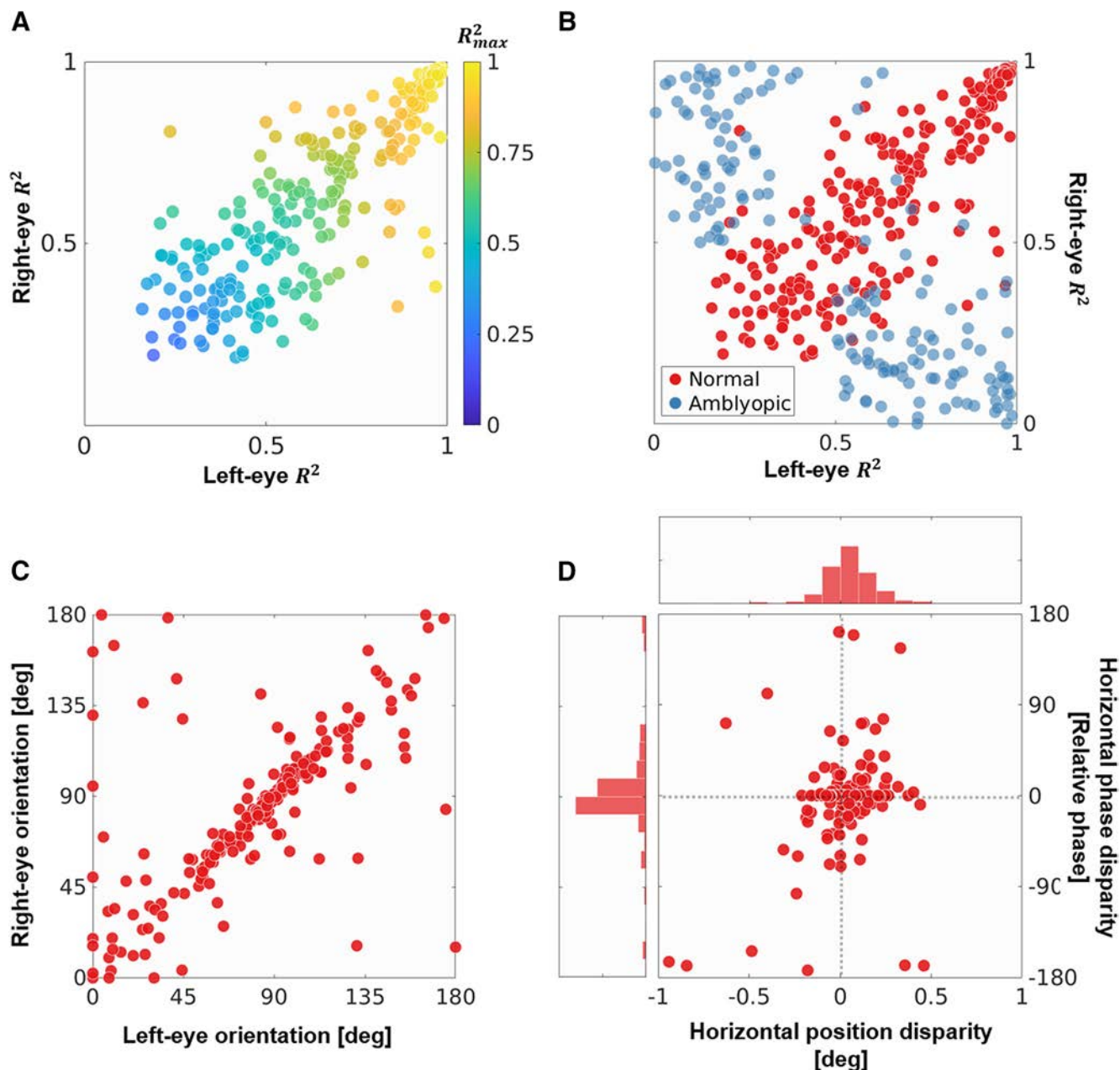
Figure 3. Characterizing a simulated population of 300 foveal V1 neurons. **A**, The goodness-of-fit R^2 criteria for Gabor fitting in the left and the right eye. Each dot represents a neuron, while the filled color indicates the quantity $R^2_{\max} = \max\{R^2_{\text{left}}, R^2_{\text{right}}\}$, a measure of how well a Gabor function describes the receptive field regardless of whether the neuron is monocular or binocular. **B**, Neurons from normal (red dots) and amblyopic (blue dots) populations trained on foveal stimuli (for more details, see Materials and Methods). **C**, The orientation parameter for Gabors fitted to the left and the right eye receptive fields. Only neurons with binocular receptive fields defined by the criterion $\min\{R^2_{\text{left}}, R^2_{\text{right}}\} \geq 0.5$ are shown. Note that the range of the angles is [0°, 180°], with 0° being the same as 180°. **D**, DTCs for the binocular neurons from **C** were estimated by the binocular cross-correlation of the receptive fields of the left and right eyes under horizontal displacement. Gabors were fitted to the resulting curves, thereby giving a measure of the position and phase disparity in the receptive fields (Prince et al., 2002a,b). This figure shows the position disparity plotted against the relative phase disparity. It also shows the histograms for the two types of disparities.

the stimuli compete with intraocular correlations. To demonstrate this, in Figure 3B we show neurons from two simulated populations: the normal foveal population from Figure 3A (red dots); and a simulated amblyopic population (blue dots). The neurons in the amblyopic population occupy the top left and bottom right quadrants (indicating that the neurons are monocular), as opposed to the binocular neurons from the normal population in the top right quadrant. In the amblyopic population, the projections in the two eyes were sampled from different foveal locations within the same image (see Materials and Methods), thereby mimicking the oculomotor behavior exhibited by strabismic amblyopes. Here, although neurons learn from left and right image sets with the same wide-field statistics, because of the lack of local intereye correlations they primarily develop monocular receptive fields, choosing 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 neurons in the foveal population. Here, we define a binocular unit by using the criterion $\min\{R^2_{\text{left}}, R^2_{\text{right}}\} \geq 0.5$ (i.e., both eyes fit the Gabor model with an R^2 value of at least 0.5). We see that there is a strong correspondence between the orientation in the two eyes (0° and 180° are
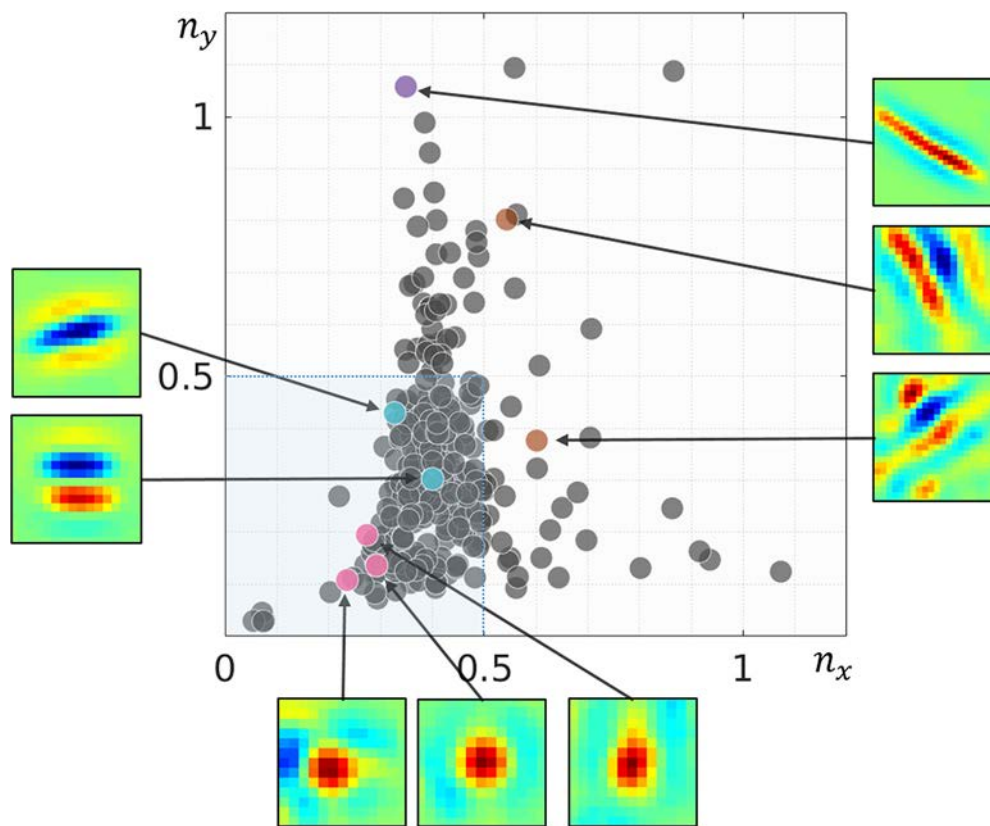
Figure 4. RC distribution for the simulated foveal population (dominant eye). RC is a symmetry-invariant size descriptor for simple-cell receptive fields (for a detailed description, see Materials and Methods). High values along the n_x-axis convey large number of lobes in the receptive field, while high values along the n_y-axis describe elongated receptive fields. Receptive fields for some representative neurons are also shown. The pink neurons are examples of blob-like receptive fields, while cyan markers illustrate two to three lobed neurons, which constitute the majority of the converged population. Brown markers show neurons with a higher number of lobes, while the purple marker shows a neuron with an elongated receptive field. Most of the converged neurons lie within the square bounded by 0 and 0.5 along both axes (shaded in blue), something that is also observed in electrophysiological recordings in monkey and cat (Ringach, 2002).

equivalent orientations), with a large number of neurons being oriented along the cardinal directions (i.e., either horizontally or vertically). Though not shown here, a similar correspondence is found between the Gabor frequencies in the two eyes (the frequency range is limited between 0.75 and 1.25 cycles/°, due to the use of fixed LGN receptive field sizes), with an octave bandwidth of ~1.5, a value close to those measured for simple cells with similar frequency peaks (De Valois et al., 1982). Furthermore, Figure 3D shows the position and phase encoding of the horizontal disparities as derived by fitting Gabor functions to DTCs estimated by the BC method (for details, see the legend of Fig. 3 and Materials and Methods). Both position and phase are found to code for disparities between −0.5° and 0.5° visual angles (phase disparities were converted to absolute disparities using the frequency of the fitted Gabor function), with most units coding for near-zero disparity. Here, we have presented the phase disparity estimated from the fitted Gabor functions to allow for a better comparison with previous single-cell studies (De Valois et al., 1982; Tsao et al., 2003). However, it has been shown that phases estimated using Gabor fitting do not accurately characterize the symmetry of the DTC, especially when the DTC has a Gaussian-like shape (Read and Cumming, 2004). Therefore, to complete our analysis, we also computed the SP metric (for details, see Materials and Methods). It revealed that our DTCs were exclusively even symmetric with an SP circular mean = 2.6° (95% confidence interval, −2.8° to 8.0°).

In addition to edge-detecting Gabor-like receptive fields, more compact blob-like neurons have also been reported in cat and monkey recordings (Jones and Palmer, 1987; Ringach, 2002).

These receptive fields are not satisfactorily predicted by efficient approaches such as independent component analysis (ICA) and sparse coding (Ringach, 2002). In Figure 4, we present the distribution of RCs for the dominant eye of our converged foveal population. The n_x- and n_y-axes approximately code for the number of lobes and the elongation of the receptive field perpendicular to the periodic function (for more details, see Materials and Methods). Most of the receptive fields in our converged population lie within 0.5 units on the two axes (Fig. 4, blue-shaded region). These receptive fields are not very elongated and typically contain 1 -1.5 cycles of the periodic function (Fig. 4, cyan markers show two typical receptive fields of this type). We also observe blob-like receptive fields (Fig. 4, magenta markers), which display a larger invariance to orientation. In addition, we show three receptive fields that lie further away from the main cluster, along the n_x-axis with multiple lobes (Fig. 4, brown markers), and along the n_y-axis with an elongated receptive field (Fig. 4, purple marker). This RC distribution is close to those reported in cat and monkey recordings (Jones and Palmer, 1987; Ringach, 2002), while RC distributions found by optimization-based coding schemes such as ICA and sparse coding typically tend to lie to the right of this distribution (Ringach, 2002).

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. Figure 5A
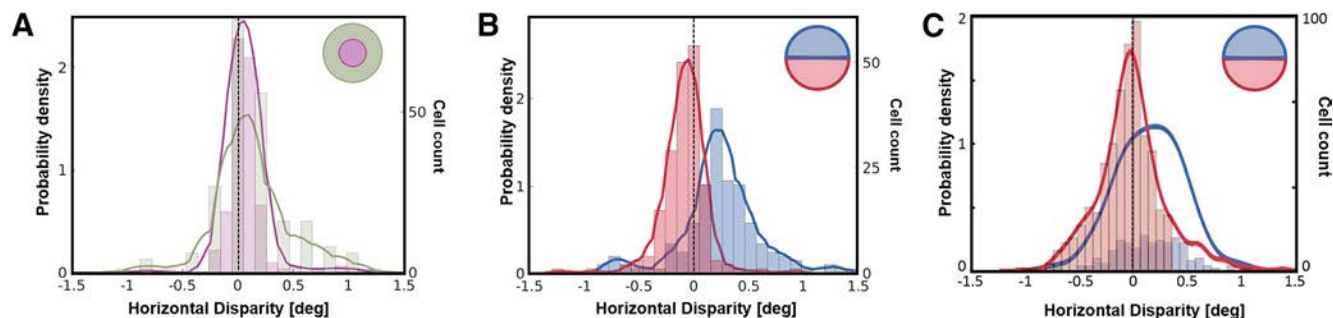
Figure 5. Retinotopic simulations and population biases in horizontal disparity. ***A***, Histograms and estimated probability density functions for horizontal disparity in two simulated populations of 300 V1 neurons each. The first population was trained on inputs sampled from the fovea (purple), while the second population was trained on inputs from the periphery (green). ***B***, Histogram and probability density estimates of two populations of 300 neurons each, trained on inputs from the upper (blue) and lower (red) hemifields. ***C***, A meta-analysis (Sprague et al., 2015) of results from electrophysiological recordings in monkey V1 neurons (820 neurons in total) by five separate groups. The neurons were split into two groups: neurons with upper hemifield receptive fields are in blue, while those with receptive fields in the lower hemifield are in red.

shows a histogram of the horizontal disparity in two simulated populations of neurons, one trained on inputs from the central field of vision (eccentricity ≤3°; Fig. 5*A*, purple), and the other on inputs from the periphery (eccentricity between 6° and 10°; Fig. 5*A*, green). The peripheral population shows a clear broadening of its horizontal disparity tuning with respect to the central population (Wilcoxon rank-pair test, $p = 0.046$). As one traverses from the fovea to the periphery, the range of encoded horizontal disparities has indeed been shown to increase, both in V1 (Durand et al., 2007) and in natural scenes (Sprague et al., 2015).

Another statistical bias reported in retinal projections of natural scenes (Sprague et al., 2015) 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 (Hibbard and Bouzit, 2005; Sprague et al., 2015). A meta-analysis (Sprague et al., 2015) of 820 V1 neurons from five separate electrophysiological studies indicates that this bias could also be reflected in the binocular neurons in the early visual system. Figure 5*C* shows the results of this meta-analysis, with the probability density for the neurons in the lower hemifield (in red) showing a bias toward crossed disparities while the upper hemifield (in blue) shows a bias toward uncrossed disparities. In comparison, Figure 5*B* presents histograms of the horizontal disparity in binocular neurons from two simulated populations—one trained on samples from the lower hemifield (red), and the other on the upper hemifield (blue). Figure 5 (compare *B*, *C*) shows that both the range and the peak tuning of the two simulated populations closely resemble those reported in electrophysiology. The peaks in Fig 5*B* are at 0.17° for the upper hemifield and −0.02° for the lower hemifield (Wilcoxon rank-pair test, $p = 2.15 \times 10^{-14}$), while the meta-analysis by Sprague et al. (2015) shown in Fig 5*D* reports the values as 0.14° and −0.09°, respectively.

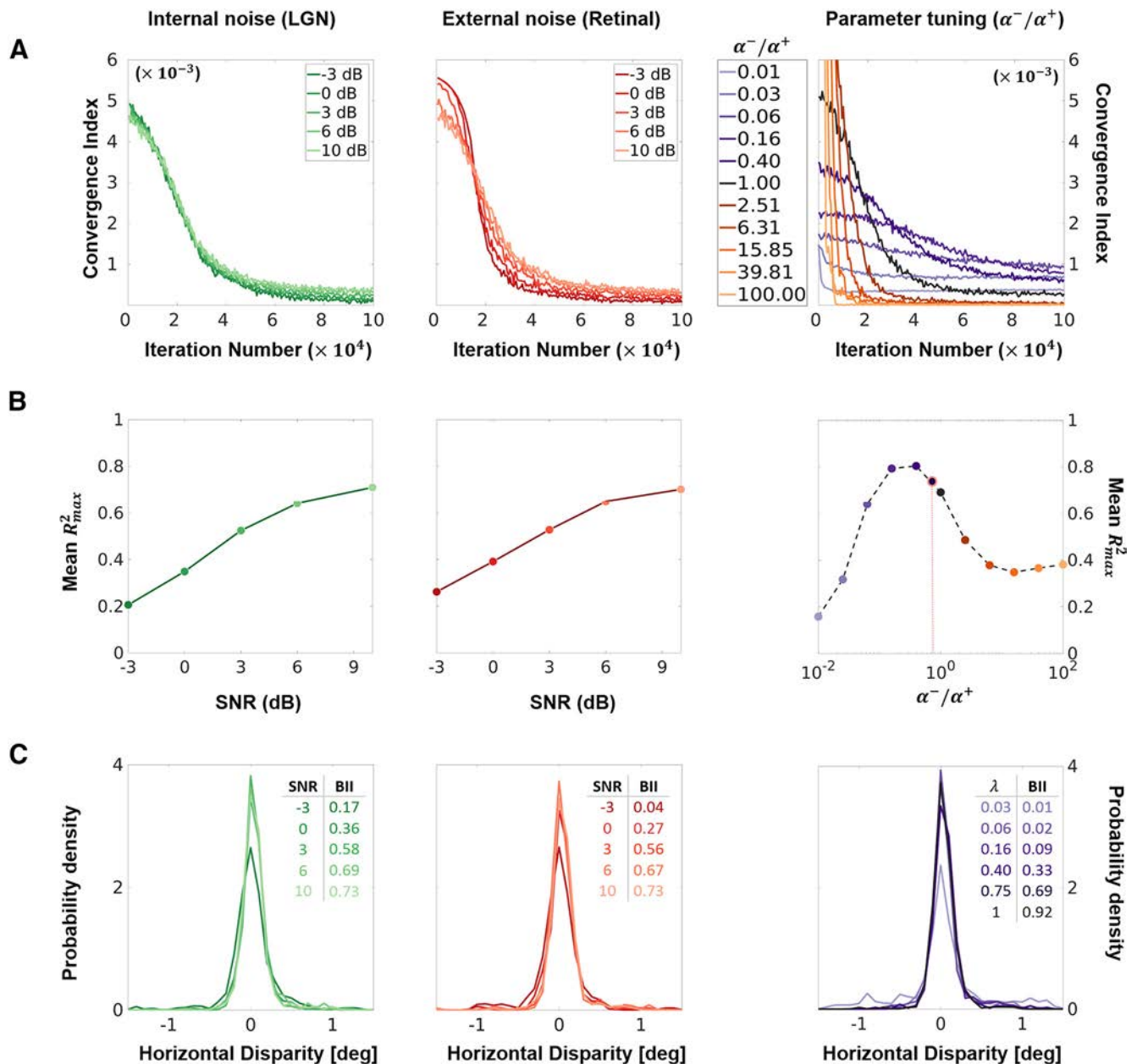**Robustness to internal and external noise**
In the results reported so far, the only sources of noise are the camera sensors used to capture the stereoscopic dataset. To further test the robustness of our model, we introduced additive white Gaussian noise $\sim N(0,\sigma_{\text{noise}})$ at two stages of the visual pipeline—external noise in the image, and internal noise in the LGN layer. The noise was manipulated such that the SNR varied in steps of ~3 dB from -3 dB ($\sigma_{\text{noise}} = 2 \times \sigma_{\text{signal}}$) to 10 dB ($\sigma_{\text{noise}} = 0.1 \times \sigma_{\text{signal}}$). Since the noise in these simulations in-

duces timing jitter at the retinal and LGN stages, they are especially relevant for our model, which relies on first-spike latencies. In the first and second columns of Figure 6, we show the behavior of the system under internal (in green colors) and external noise (red colors). The first two panels of Figure 6*A* show the CI of the network, plotted against the iteration number (training duration) for the internal and external noise conditions, respectively. The lightness of the color indicates the level of noise (the darker the color, the higher the noise). We see that the CI approaches zero asymptotically in all conditions, indicating that the weights converge to a stable distribution for all levels of noise. The corresponding panels in Figure 6*B* show the mean R^2_{max}, plotted against the SNR. The mean R^2_{max}, as expected, increases with the SNR (the higher the noise, the less Gabor like the receptive fields). We see that the receptive fields converge to Gabor-like functions for a large range of internal and external noise, with the R^2_{max} dropping to <0.5 only at SNR levels <3 dB ($\sigma_{\text{noise}} \geq 0.5 \times \sigma_{\text{signal}}$). In Figure 6*C*, we show the effect of the two types of noise on disparity selectivity using the population disparity tuning and the BII. While the overall disparity tuning of the converged neural population remains stable in all conditions, we see a marked decrease in the BII with noise. Since we calculate the population disparity tuning using BC RDSs (which show the theoretical limit of the tuning curve), this indicates that, while the receptive fields in the two eyes converge to similar configurations, their sensitivity decreases with an increase in noise.

Overall, these analyses demonstrate that our approach is robust to a large range of noise-induced timing jitter at the input and LGN stages. Low-noise conditions produce a synaptically converged network with Gabor-like, binocular receptive fields. Under high-noise conditions, the network still converges, albeit to receptive fields that show a decreased sensitivity to disparity and are no longer well modeled by Gabor functions.

Robustness to parameter variation
One of the key parameters of a Hebbian STDP model is the ratio (say, λ) of the LTD and LTP rates (α^- and α^+, respectively). This parameter determines the balance between the weight changes induced by potentiation and depression, and thus directly influences the number of pruned synapses after convergence. The simulations presented in this article use $\lambda = 3/4$, a value based on STDP models previously developed in the team (Masquelier and Thorpe, 2007). In this section, we test the robustness of the model to variations in this important parameter. The value of λ was varied between 0.01 and 100 in log steps (for the details, see

Figure 6. Robustness analysis. Internal (LGN) and external (Retinal) noise, and parameter (LTD/LTP rate ratio) variation were introduced in the model. The internal (left column) and external (middle column) noises were modeled as additive white Gaussian processes with SNR levels between −3 and 10 dB. The ratio of the LTD (α^-) and LTP (α^+) rates (right column) was varied in logarithmic steps from 10^{-2} to 10^2 (i.e., $\alpha^- = \alpha^+/100$ to $\alpha^- = \alpha^+ \times 100$). This ratio is a crucial parameter that determines the relative contributions of synaptic potentiation and depression to the overall learning rate. **A**, CI (see Materials and Methods) for the three types of noise plotted against the number of iterations (training period). For a stable synaptic convergence, the CI should approach zero asymptotically. Internal and external noise are plotted in green and red colors, respectively, with the lightness of the curve coding for the SNR (the darker the curve, the higher the noise). The LTD/LTP rate ratio is plotted in purple colors for $\alpha^- \leq \alpha^+$ and orange/brown colors for $\alpha^- \geq \alpha^+$. The darker the color, the closer α^- is to α^+, with black being $\alpha^- = \alpha^+$. This color coding is used throughout the figure. **B**, Mean R^2_{max}. The left and middle plots show the mean R^2_{max} plotted against the SNR of the internal and external noise respectively. The right plot shows the mean R^2_{max} plotted against the α^-/α^+ ratio. An additional dot (outlined in pink) is added to this plot to indicate the value of α^-/α^+ (= 3/4) used for all the simulations presented in the article (except this figure). **C**, Population disparity tuning and BII. Each curve in this panel shows the probability density (PD) function estimated over the preferred disparities of a neural population (corresponding to a given condition). Each PD curve is accompanied by an estimate of the mean BII of the neural population (for details, see Materials and Methods), calculated using responses to RDS stimuli. Only neural populations that show a reasonable response to RDS stimuli are shown (otherwise, the BII and PD curves are not meaningful).

Materials and Methods). The rightmost column in Figure 6 shows the results of these simulations. Figure 6A shows the CI, plotted against the iteration number (training period); the purple colors code for $\alpha^- \leq \alpha^+$, and the brown/orange colors code for $\alpha^- \geq \alpha^+$, with the intensity of the color coding for the deviation from $\lambda = 1$. Once again, the synaptic weights converge for all values of the parameter, with convergence being quicker for higher values of λ. Figure 6B shows the mean R^2_{max} plotted against

λ. In addition to the modeled values of λ, here we also show the value for $\lambda = 3/4$, which was used in all simulations in this article. The model produces Gabor-like receptive fields for a very broad range of λ values (for $R^2_{max} \geq 0.5$, this range is approximately between 0.05 and 1). Low values of λ (α^- is very low compared with α^+) lead to a decrease in performance as the network learns quickly, and without effective pruning of depressed synapses. For $\alpha^- \geq \alpha^+$, we find that the performance declines (learned fea-
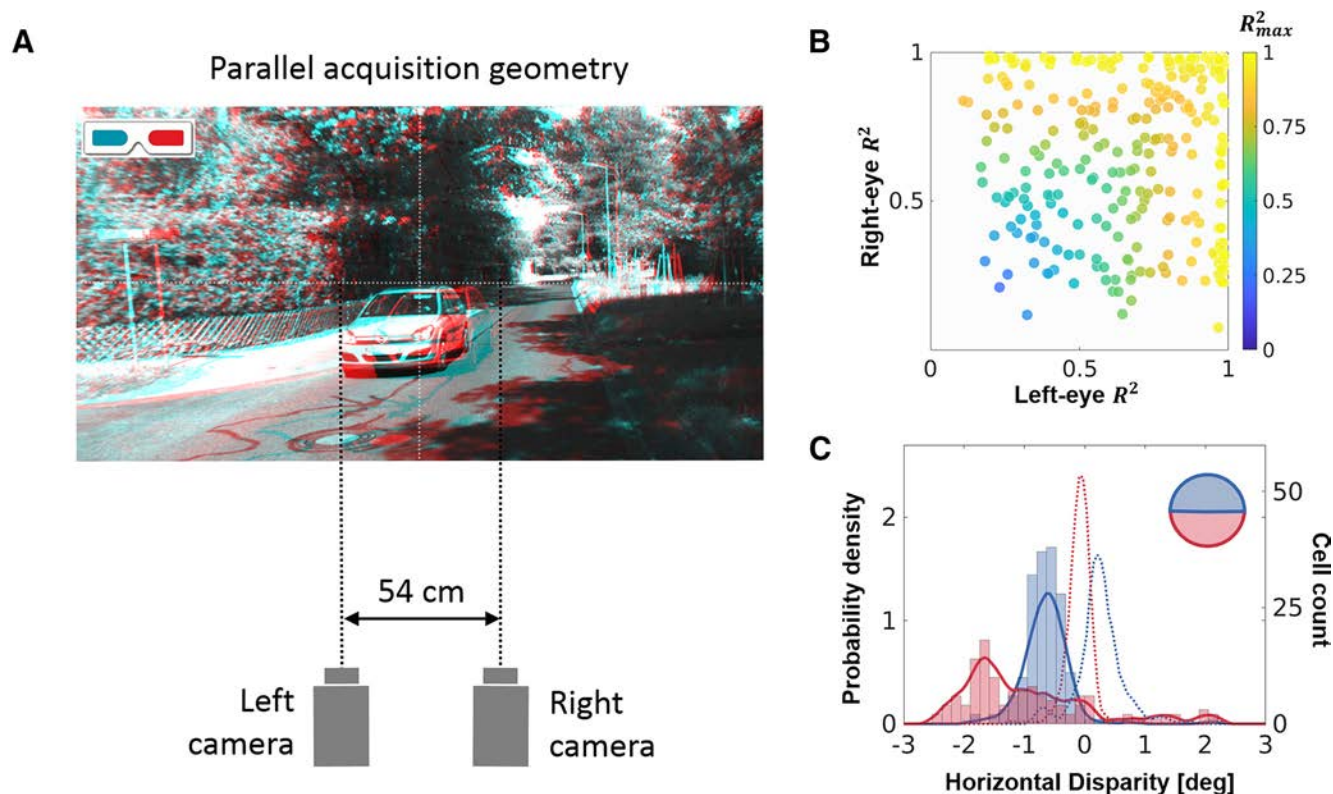
Figure 7. The KITTI dataset. The model was tested using a second dataset (KITTI; for details, see Materials and Methods), which had a different acquisition geometry. **A**, Acquisition geometry. The cameras were parallel and separated by 54 cm, a geometry that does not resemble the human visual system. A cyan-red anaglyph reconstruction of a representative scene from the KITTI dataset is also shown. **B**, Gabor fit. The R^2 value for the Gabor fit procedure on the left and right receptive fields of the converged neurons, with the color of the symbol coding for the R^2_{max} (compare to Fig. 3A for the Hunter–Hibbard dataset). **C**, Horizontal disparity tuning. Horizontal disparity tunings for two populations of neurons trained on the upper (blue) and lower (red) hemifields of the KITTI database. For comparison, the curves obtained using the Hunter–Hibbard database (Fig. 4C) are plotted in dashed lines using the same color scheme.

tures are more likely to be erased because of the strong pruning), finally saturating at $\alpha^- \geq 6\alpha^+$. Furthermore, in Figure 6C, we show how the disparity tuning of the modeled population is affected by a change in λ. The disparity tuning is stable for $0.05 < \lambda < 1$ (similar to the range that leads to $R^2_{max} \geq 0.5$), with the mean BII value increasing as we move from low values of λ toward $\lambda = 1$. For $\alpha^- \geq \alpha^+$, the sensitivity of the receptive fields is very low, and they fail to produce any response to RDS stimuli.

With this analysis, we have shown that our model shows stable synaptic convergence for all tested values of λ, and a functional convergence to Gabor-like binocular receptive fields for a large range of the parameter ($0.05 < \lambda < 1$).

Robustness to dataset acquisition geometry

All our simulations (except this section) use the Hunter–Hibbard database, which was collected using an acquisition geometry close to the human visual system. The emergent receptive fields are disparity sensitive and closely resemble those reported in the early visual system. In this analysis, we prove that this emergence is not only a property of the dataset, but also of the human-like acquisition geometry. To do this, the model was tested on another dataset (KITTI; available at http://www.cvlibs.net/datasets/kitti/), which was collected using two widely spaced (54 cm apart), parallel cameras, which is an acquisition geometry that does not resemble the human visual system at all (Fig. 7A). Furthermore, we present two results from this series of simulations. Figure 7B shows the R^2 value of Gabor fitting in the two eyes (symbols color coded by R^2_{max}) for a population trained on the foveal region of the KITTI dataset. We observe both monocular

neurons (high R^2 value in one eye, top left or bottom right quadrants) and binocular neurons (high R^2 value in both eyes, top right quadrant). In comparison, the converged receptive fields for the Hunter–Hibbard dataset are mostly binocular (Fig. 3A). This is not surprising as the large intercamera distance (54 cm compared with 6.5 cm for the Hunter–Hibbard dataset) combined with the parallel geometry results in less overlap in the left and right images of the KITTI dataset, which further leads to weaker binocular correlations.

Figure 7C shows population disparity tunings for two separate populations trained on the upper (blue) and lower (red) hemifields of the KITTI dataset. As expected, the lower hemifield population is tuned to disparities that are more crossed (negative) than the upper hemifield population (Fig. 3C; see also Biases in natural statistics and neural populations). For comparison, the curves for the Hunter–Hibbard dataset (from Fig. 3C) are plotted as dotted lines. The parallel geometry of the cameras in the KITTI dataset results in predominantly crossed disparities (the fixation point is at infinity, and hence all disparities are, by definition, crossed). Thus, while the upper versus lower hemifield bias (see Biases in natural statistics and neural populations) still emerges from our model, the population tunings no longer correspond to those reported in electrophysiological recordings (Fig. 3D; Sprague et al., 2015).

Decoding disparity through V1 responses

Recently, the effect of natural 3-D statistics on behavior has been modeled using labeled datasets in conjunction with task-optimized filters (Burge and Geisler, 2014) and feedforward
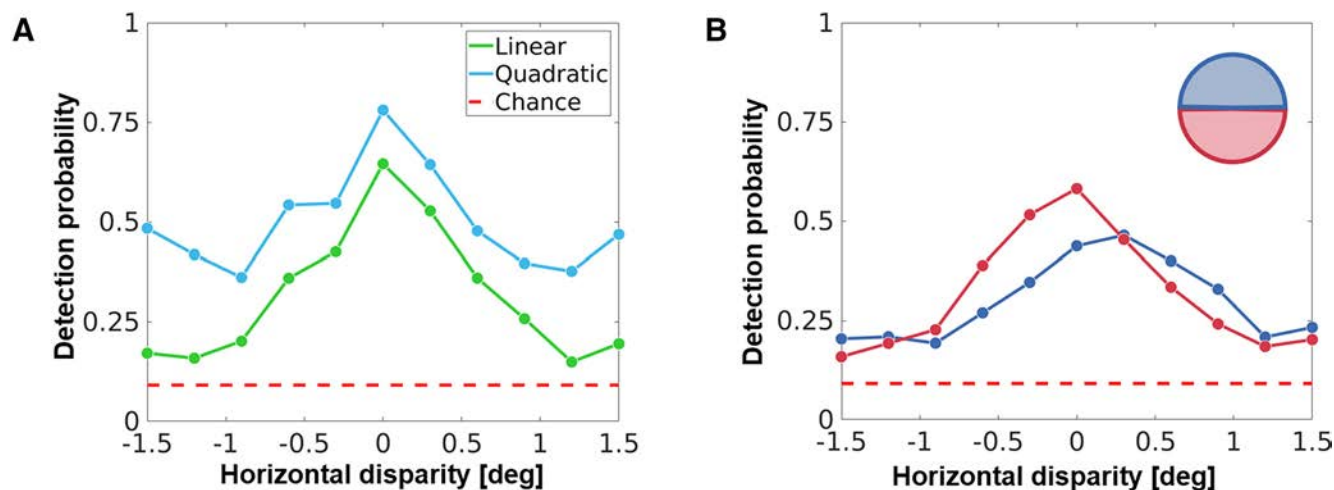
Figure 8. Decoding disparity through RDSs. **A**, Output spikes of the V1 layer (foveal simulation) to RDS stimuli at various disparities were used to train linear (green) and quadratic (blue) discriminant classifiers. The RDS stimuli were chosen as they do not contain any other information except disparity. The performance of these classifiers was then tested using different sets of labeled RDS stimuli (10,000 total RDS stimuli; 70% used for training, 30% used for testing; cross-validation using 25 folds). This plot shows the detection probability (probability of correct identification), plotted against the disparity. The quadratic classifier performs better than the linear classifier, while both classifiers perform above chance level (in dashed red). **B**, Two populations of V1 neurons trained on the upper (blue) and lower (red) hemifields were used to test a linear classifier using the same stimuli as in **A**. The detection probability of the two classifiers reflects the population bias induced by the disparity statistics of the upper and lower hemifields (Fig. 5B).

convolutional networks (Goncalves and Welchman, 2017). Typically, such models are evaluated by how well they perform at classifying/detecting previously unseen data. In Figure 8A, we show the result (detection probability) of training two very simple classifiers on the activity of the converged V1 layer from our model, using RDS stimuli at 11 equally spaced disparities between −1.5° and 1.5° (for details, see Materials and Methods). RDS stimuli were used because they contain no other meaningful information except disparity. 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 neurons are linear units similar to simple cells, and a nonlinear processing of their activity makes the decoding units conceptually closer to sharply tuned complex cells (Ohzawa et al., 1997; Cumming and DeAngelis, 2001; Henriksen et al., 2016).

We have previously seen that V1 populations trained on the upper and lower hemifields show systematic biases for uncrossed and crossed disparities, respectively (Fig. 5B). In Figure 8B, we show the performance of two linear classifiers, trained on V1 populations from the upper (blue) and lower (red) hemifields using the same method as shown in Figure 8A. The crossed/uncrossed disparity biases in the V1 layer are also 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 too, something that has indeed been reported in human subjects (Sprague et al., 2015).

Discussion

We have demonstrated how a simple Hebbian learning rule expressed through STDP leads to a population of neurons capable of encoding binocular disparity statistics in natural scenes. Our findings suggest that 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 purely efficiency-based approaches and electrophysiological data (Ringach, 2002).

The converged receptive fields show strong interocular correlations for properties such as orientation, size, and frequency (Fig. 3C shows the orientation, but similar results were found for size and frequency). This suggests that the selectivity of the neurons is driven by interocular correlations, despite competition from within-eye correlations. Although correlation-based refinement of binocular receptive fields has previously been proposed at a theoretical level (Berns et al., 1993), our results prove that natural scenes and the geometry of the human visual system are indeed capable of providing the strength of interocular correlations required for binocularity to emerge. This point is further illustrated by our simulation of the oculomotor behaviour characteristic of strabismic amblyopia. The uncoordinated eye fixations lead to weaker interocular correlations, and a majority of the converged neurones are monocular (Fig. 3B). This is not altogether surprising as STDP learning from monocular images is known to produce Gabor-like receptive fields (Delorme et al., 2001; Masquelier, 2012). These results also echo findings in monkeys reared with amblyopic deficits (Movshon et al., 1987; Kiorpes et al., 1998) where the binocularity of V1 neurons is found to be severely restricted.

The converged neurons show selectivity to disparities between −0.5° and 0.5° (Fig. 3D), which is in line with electrophysiological findings (Poggio et al., 1985; Prince et al., 2002a), and gives the population the theoretical capacity to encode a large range of realistic disparity values (Sprague et al., 2015). Interestingly, this selectivity emerges through a binocular combination of monocularly thresholded signals. This differs from classical binocular-energy models (BEMs) where a nonlinearity is applied after binocular combination of monocular signals (Ohzawa et al., 1997). Indeed, modifications of the BEM with monocular thresholding have been shown to explain electrophysiological data that cannot be explained through classical BEMs (Read et al., 2002). Through retinotopic simulations, we give two further examples that show how learning based on a Hebbian approach could explain how biases in natural statistics can transfer to neural popu-

lations in the early visual system. In the first example (Fig. 5A), we demonstrate that neurons trained on stimuli from peripheral eccentricities are selective to a larger range of horizontal disparities (Durand et al., 2002) when compared with neurons trained on foveal stimuli. In the second example (Fig. 5B), we use two neural populations to show that receptive fields in the lower/upper hemifield are biased toward crossed/uncrossed disparities. This reflects the intuitive bias that objects in the lower/upper hemifield are more likely to be closer/farther away than fixation. This bias has been demonstrated through free-viewing tasks, and is also reflected in neural responses at both single-unit (Sprague et al., 2015) and macroscopic (Nasr and Tootell, 2018) levels.

Few computational studies have previously characterized the relationship between natural statistics and binocular disparity responses. Based on their approach to learning, they can approximately be classified as unsupervised strategies (Hoyer and Hyvärinen, 2000; Hunter and Hibbard, 2015), which learn from unlabeled stimuli, and supervised approaches (Burge and Geisler, 2014; Goncalves and Welchman, 2017), which use labeled data. While results from unsupervised learning can be interpreted as possible encoding schemes for the stimuli, supervised approaches optimize model performance on specific tasks such as disparity or contrast discrimination. Both of these approaches solve an optimization problem that 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, objective function-based unsupervised approaches as it does not necessarily converge toward 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 toward an optimum is a result of the stability of natural statistics across samples.

A commonly used unsupervised approach to studying the encoding of 3-D natural statistics is the ICA (Hoyer and Hyvärinen, 2000; Hunter and Hibbard, 2015), a technique that maximizes the mutual information between the stimuli and the activity of the encoding units. Here, we mention three critical differences that separate our results from those obtained using ICA. First, Hebbian learning could potentially be suboptimal when viewed in terms of the mutual information-based objective functions used for ICA. For example, Hebbian learning leads to blobby receptive fields, which are highly feature invariant and a very inefficient choice for encoding natural images. Second, Hebbian learning is usually unable to converge to features that 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 neurons with antiphase receptive fields in the two eyes (Fig. 3D). This is in line with electrophysiological data collected in monkeys, cats, and barn owls (Prince et al., 2002a). 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 ratio of antiphase receptive fields (Hunter and Hibbard, 2015). Third, the receptive fields obtained through the Hebbian model are closer in structure to electrophysiological data than those obtained by ICA (Ringach, 2002); for instance, they have fewer lobes and show broader orientation tuning. Thus, although ICA representations offer a highly optimal encoding of the data in terms of informa-

tion transfer, they are inconsistent with electrophysiological findings, which often include neurons that are suboptimal in terms of any particular coding strategy (Ringach, 2002).

As initially suggested by Barlow (1961), although 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 behavior. We demonstrate (Fig. 8) that the activity of the converged V1 population in our model is capable of driving simple classifiers to an above-chance performance. This suggests that disparity encoding in the early visual system does not necessarily need supervised training, and an unsupervised feedforward Hebbian process can lead to neural units whose responses can be interpreted in terms of a 3-D percept through downstream processing. It must be noted that here, decoding through a classifier is only presented as a simplified illustration of perceptual responses, which may involve other processes such as corticocortical interactions and feedback.

Interestingly, our observations follow despite the lack of certain important features of the early visual system, such as parvocellular receptive fields, local V1 retinotopy, and corticocortical interactions. Recurrent V1 interactions have indeed been implicated in the refinement of feedforward disparity signals (Samonds et al., 2013). We predict that the introduction of these features in our model could lead to the emergence of other disparity-sensitive units observed in V1, such as neurons with inhibitory interocular interactions (Ohzawa and Freeman, 1986a; Read and Cumming, 2004) and complex cells (Ohzawa and Freeman, 1986b). Furthermore, using a more detailed model of the neuron (e.g., the leaky integrate-and-fire or Izhikevich models) could allow for the emergence of other biases reported in the literature, such as the correlation between preferred luminance and disparity (Samonds et al., 2012).

Receptive fields in the early visual system across species show both prenatal (Wiesel and Hubel, 1974; McLaughlin and O'Leary, 2005) and postnatal (Li et al., 2006) refinement. The exact refinement period depends on both the species and the visual feature itself (Caporale and Dan, 2008). For disparity, a preliminary form of binocular correlation has been shown to exist a few days after birth (Chino et al., 1997). However, Tao et al. (2014) showed that disparity selectivity undergoes a critical refinement through visual experience. Interestingly, Hebbian and Hebbian-like schemes have been proposed as an important component in both prenatal and postnatal refinement processes (Butts, 2002; Butts et al., 2007; Larsen et al., 2010). Although the present work focuses on the encoding properties of the model after eye opening, it nonetheless offers a more complete and interpretable picture of how computational structures in the brain emerge when compared with one-shot learning methods based on the information content of natural scenes (e.g., ICA and its derivatives). A more complete version of our model would include both prenatal and postnatal phases, and is one of the major future directions of our work.

Although this study demonstrates the relevance of Hebbian plasticity in binocular vision, this approach could easily be extended to other sensory modalities, and even multimodal representations in the early sensory cortex. For computational neuroscientists, this offers a critical advantage over conventional optimization-based approaches as it allows for models that track cortical connectivity over time, such as studies on development, critical period plasticity, and perceptual learning.

References

Albrecht DG, Geisler WS, Frazor RA, Crane AM (2002) Visual cortex neurons of monkeys and cats: temporal dynamics of the contrast response function. J Neurophysiol 88:888–913. CrossRef Medline

Anzai A, Ohzawa I, Freeman RD (1999) Neural mechanisms for processing binocular information I. Simple cells. J Neurophysiol 82:891–908. CrossRef Medline

Atick JJ, Redlich AN (1992) What does the retina know about natural scenes? Neural Comput 4:196–210. CrossRef

Barlow H (1961) Possible principles underlying the transformations of sensory messages. In: Sensory communication, pp 217–234. Cambridge, MA: MIT.

Barlow H (2001) The exploitation of regularities in the environment by the brain. Behav Brain Sci 24:602–607. Medline

Barlow H, Földiák P (1989) Adaptation and decorrelation in the cortex. In: The computing neuron (Durbin R, Miall C, Mitchison G, eds), pp 54–72. Chicago: Addison-Wesley Longman.

Berns GS, Dayan P, Sejnowski TJ (1993) A correlational model for the development of disparity selectivity in visual cortex that depends on prenatal and postnatal phases. Proc Natl Acad Sci U S A 90:8277–8291. CrossRef Medline

Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci 18:10464–10472. CrossRef Medline

Brette R (2012) Computing with Neural Synchrony Sporns O, ed. PLoS Comput Biol 8:e1002561. CrossRef Medline

Burge J, Geisler WS (2014) Optimal disparity estimation in natural stereo images. J Vis 14:1, 1–18. CrossRef Medline

Burge J, Geisler WS (2015) Optimal speed estimation in natural image movies predicts human performance. Nat Commun 6:7900. CrossRef Medline

Burge J, Jaini P (2017) Accuracy maximization analysis for sensory-perceptual tasks: computational improvements, filter robustness, and coding advantages for scaled additive noise. PLoS Comput Biol 13: e1005281. CrossRef Medline

Butts D (2002) Retinal waves: implications for synaptic learning rules during development. Neuroscientist 8:243–253. CrossRef Medline

Butts DA, Kanold PO, Shatz CJ (2007) A burst-based "Hebbian" learning rule at retinogeniculate synapses links retinal waves to activity-dependent refinement. PLoS Biol 5:e61. CrossRef Medline

Caporale N, Dan Y (2008) Spike timing-dependent plasticity: a Hebbian learning rule. Annu Rev Neurosci 31:25–46. CrossRef Medline

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

Chino Y, Smith E, Hatta S, Cheng H (1997) Postnatal development of binocular disparity sensitivity in neurons of the primate visual cortex. J Neurosci 17:296–307. CrossRef Medline

Cooper EA, Norcia AM (2014) Perceived depth in natural images reflects encoding of low-level luminance statistics. J Neurosci 34:11761–11768. CrossRef Medline

Cumming BG, DeAngelis GC (2001) The physiology of stereopsis. Annu Rev Neurosci 24:203–238. CrossRef Medline

Delorme A, Perrinet L, Thorpe SJ (2001) 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. CrossRef

De Valois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res 22:545–559. CrossRef Medline

Durand JB, Zhu S, Celebrini S, Trotter Y (2002) Neurons in parafoveal areas V1 and V2 encode vertical and horizontal disparities. J Neurophysiol 88:2874–2879. CrossRef Medline

Durand JB, Celebrini S, Trotter Y (2007) Neural bases of stereopsis across visual field of the alert macaque monkey. Cereb Cortex 17:1260–1273. CrossRef Medline

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

Freeman RD, Ohzawa I (1992) Development of binocular vision in the kitten's striate cortex. J Neurosci 12:4721–4736. CrossRef Medline

Furmanski CS, Engel S (2000) An oblique effect in human primary visual cortex. Nat Neurosci 3:535–536. CrossRef Medline

Gawne TJ, Kjaer TW, Richmond BJ (1996) Latency: another potential code for feature binding in striate cortex. J Neurophysiol 76:1356–1360. CrossRef Medline

Geisler WS (2008) Visual perception and the statistical properties of natural scenes. Annu Rev Psychol 59:167–192. CrossRef Medline

Geisler WS, Perry JS (2009) Contour statistics in natural images: grouping across occlusions. Vis Neurosci 26:109–121. CrossRef Medline

Girshick AR, Landy MS, Simoncelli EP (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. Nat Neurosci 14:926–932. CrossRef Medline

Gollisch T, Meister M (2008) Rapid neural coding in the retina with relative spike latencies. Science 319:1108–1111. CrossRef Medline

Goncalves NR, Welchman AE (2017) "What not" detectors help the brain see in depth. Curr Biol 27:1403–1412.e8. CrossRef Medline

Gütig R, Aharonov R, Rotter S, Sompolinsky H (2003) Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. J Neurosci 23:3697–3714. CrossRef Medline

Hebbs D (1949) The organization of behavior. New York: Wiley.

Henriksen S, Cumming BG, Read JC (2016) A single mechanism can account for human perception of depth in mixed correlation random dot stereograms. PLoS Comput Biol 12:e1004906. CrossRef Medline

Hibbard PB, Bouzit S (2005) Stereoscopic correspondence for ambiguous targets is affected by elevation and fixation distance. Spat Vis 18:399–411. CrossRef Medline

Howe CQ, Purves D (2002) Range image statistics can explain the anomalous perception of length. Proc Natl Acad Sci U S A 99:13184–13188. CrossRef Medline

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

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154. CrossRef Medline

Hunter DW, Hibbard PB (2015) Distribution of independent components of binocular natural images. J Vis 15:6, 1–31. CrossRef Medline

Jones JP, Palmer LA (1987) An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. J Neurophysiol 58:1233–1258. CrossRef Medline

Karklin Y, Lewicki MS (2009) Emergence of complex cell properties by learning to generalize in natural scenes. Nature 457:83–86. CrossRef Medline

Kiorpes L, Kiper DC, O'Keefe LP, Cavanaugh JR, Movshon JA (1998) Neuronal correlates of amblyopia in the visual cortex of macaque monkeys with experimental strabismus and anisometropia. J Neurosci 18:6411–6424. CrossRef Medline

König P, Engel AK, Singer W (1996) Integrator or coincidence detector? The role of the cortical neuron revisited. Trends Neurosci 19:130–137. CrossRef Medline

Larsen RS, Rao D, Manis PB, Philpot BD (2010) STDP in the developing sensory neocortex. Front Synaptic Neurosci 2:9. CrossRef Medline

Li Y, Fitzpatrick D, White LE (2006) The development of direction selectivity in ferret visual cortex requires early visual experience. Nat Neurosci 9:676–681. CrossRef Medline

Maass W (2000) On the computational power of winner-take-all. Neural Comput 12:2519–2535. CrossRef Medline

Masquelier T (2012) 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. CrossRef Medline

Masquelier T (2017) STDP allows close-to-optimal spatiotemporal spike pattern detection by single coincidence detector neurons. Neuroscience. Advance online publication. Retrieved September 14, 2018. doi:10.1016/j.neuroscience.2017.06.032. CrossRef Medline

Masquelier T, Thorpe SJ (2007) Unsupervised learning of visual features through spike timing dependent plasticity. PLoS Comput Biol 3:e31. CrossRef Medline

Masquelier T, Thorpe SJ (2010) Learning to recognize objects using waves of spikes and spike timing-dependent plasticity. Paper presented at the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, July.

McLaughlin T, O'Leary DD (2005) Molecular Gradients and development of retinotopic maps. Annu Rev Neurosci 28:327–355. CrossRef Medline

Movshon JA, Eggers HM, Gizzi MS, Hendrickson AE, Kiorpes L, Boothe RG (1987) Effects of early unilateral blur on the macaque's visual system. III. Physiological observations. J Neurosci 7:1340–1351. CrossRef Medline

Nasr S, Tootell RBH (2018) Visual field biases for near and far stimuli in disparity selective columns in human visual cortex. Neuroimage 168: 358–365. CrossRef Medline

Ohzawa I, Freeman RD (1986a) The binocular organization of simple cells in the cat's visual cortex. J Neurophysiol 56:221–242. CrossRef Medline

Ohzawa I, Freeman RD (1986b) The binocular organization of complex cells in the cat's visual cortex. J Neurophysiol 56:243–259. CrossRef Medline

Ohzawa I, Deangelis GC, Freeman RD (1997) Encoding of binocular disparity by complex cells in the cat's visual cortex. J Neurophysiol 77:2879–2909. CrossRef Medline

Okazawa G, Tajima S, Komatsu H (2015) Image statistics underlying natural texture selectivity of neurons in macaque V4. Proc Natl Acad Sci U S A 112:E351–E360. CrossRef Medline

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609. CrossRef Medline

Olshausen BA, Field DJ (1997) Sparse coding with an overcomplete basis set: a strategy employed by V1? Vision Res 37:3311–3325. CrossRef Medline

Poggio GF, Motter BC, Squatrito S, Trotter Y (1985) Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. Vision Res 25:397–406. CrossRef Medline

Prince SJ, Cumming BG, Parker AJ (2002a) Range and mechanism of encoding of horizontal disparity in macaque V1. J Neurophysiol 87:209–221. CrossRef Medline

Prince SJ, Pointon AD, Cumming BG, Parker AJ (2002b) Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. J Neurophysiol 87:191–208. CrossRef Medline

Read JC, Cumming BG (2004) Ocular dominance predicts neither strength nor class of disparity selectivity with random-dot stimuli in primate V1. J Neurophysiol 91:1271–1281. CrossRef Medline

Read JC, Parker AJ, Cumming BG (2002) A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. Vis Neurosci 19:735–753. Medline

Ringach DL (2002) Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. J Neurophysiol 88:455–463. CrossRef Medline

Samonds JM, Potetz BR, Lee TS (2012) Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. Proc Natl Acad Sci U S A 109:6313–6318. CrossRef Medline

Samonds JM, Potetz BR, Tyler CW, Lee TS (2013) Recurrent connectivity can account for the dynamics of disparity processing in V1. J Neurosci 33:2934–2946. CrossRef Medline

Sebastian S, Abrams J, Geisler WS (2017) Constrained sampling experiments reveal principles of detection in natural scenes. Proc Natl Acad Sci U S A 114:E5731–E5740. CrossRef Medline

Shriki O, Kohn A, Shamir M (2012) Fast coding of orientation in primary visual cortex. PLoS Comput Biol 8:e1002536. CrossRef Medline

Solomon SG, White AJ, Martin PR (2002) Extraclassical receptive field properties of parvocellular, magnocellular, and koniocellular cells in the primate lateral geniculate nucleus. J Neurosci 22:338–349. CrossRef Medline

Sprague WW, Cooper EA, Tošić I, Banks M (2015) Stereopsis is adaptive for the natural environment. Sci Adv 1:e1400254–e1400254. CrossRef Medline

Stanley GB (2013) Reading and writing the neural code. Nat Neurosci 16: 259–263. CrossRef Medline

Tanaka K (1985) Organization of geniculate inputs to visual cortical cells in the cat. Vision Res 25:357–364. CrossRef Medline

Tao X, Zhang B, Shen G, Wensveen J, Smith EL 3rd, Nishimoto S, Ohzawa I, Chino YM (2014) Early monocular defocus disrupts the normal development of receptive-field structure in V2 neurons of macaque monkeys. J Neurosci 34:13840–13854. CrossRef Medline

Thorpe S, Delorme A, Van Rullen R (2001) Spike-based strategies for rapid processing. Neural Netw 14:715–725. CrossRef Medline

Tsao DY, Conway BR, Livingstone MS (2003) Receptive fields of disparity-tuned simple cells in macaque V1. Neuron 38:103–114. CrossRef Medline

Webster MA, Mollon JD (1997) Adaptation and the color statistics of natural images. Vision Res 37:3283–3298. CrossRef Medline

Wiesel TN, Hubel DH (1974) Ordered arrangement of orientation columns in monkeys lacking visual experience. J Comp Neurol 158:307–318. CrossRef Medline