



Sub-Optimality of the Early Visual System Explained Through Biologically Plausible Plasticity

Tushar Chauhan^{1,2*}, Timothée Masquelier^{1,2} and Benoit R. Cottureau^{1,2}

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

OPEN ACCESS

Edited by:

Ali Ghazizadeh,
Sharif University of Technology, Iran

Reviewed by:

Anthony N. Burkitt,
The University of Melbourne, Australia
Michael Robert DeWeese,
University of California, Berkeley,
United States

*Correspondence:

Tushar Chauhan
research@tusharchauhan.com

Specialty section:

This article was submitted to
Neural Technology,
a section of the journal
Frontiers in Neuroscience

Received: 18 June 2021

Accepted: 25 August 2021

Published: 15 September 2021

Citation:

Chauhan T, Masquelier T and
Cottureau BR (2021) Sub-Optimality
of the Early Visual System Explained
Through Biologically Plausible
Plasticity.
Front. Neurosci. 15:727448.
doi: 10.3389/fnins.2021.727448

The early visual cortex is the site of crucial pre-processing for more complex, biologically relevant computations that drive perception and, ultimately, behaviour. This pre-processing is often studied under the assumption that neural populations are optimised for the most efficient (in terms of energy, information, spikes, etc.) representation of natural statistics. Normative models such as Independent Component Analysis (ICA) and Sparse Coding (SC) consider the phenomenon as a generative, minimisation problem which they assume the early cortical populations have evolved to solve. However, measurements in monkey and cat suggest that receptive fields (RFs) in the primary visual cortex are often noisy, blobby, and symmetrical, making them sub-optimal for operations such as edge-detection. We propose that this suboptimality occurs because the RFs do not emerge through a global minimisation of generative error, but through locally operating biological mechanisms such as spike-timing dependent plasticity (STDP). Using a network endowed with an abstract, rank-based STDP rule, we show that the shape and orientation tuning of the converged units are remarkably close to single-cell measurements in the macaque primary visual cortex. We quantify this similarity using physiological parameters (frequency-normalised spread vectors), information theoretic measures [Kullback–Leibler (KL) divergence and Gini index], as well as simulations of a typical electrophysiology experiment designed to estimate orientation tuning curves. Taken together, our results suggest that compared to purely generative schemes, process-based biophysical models may offer a better description of the suboptimality observed in the early visual cortex.

Keywords: vision, cortex, plasticity, suboptimality, Independent Component Analysis, Sparse Coding, STDP, natural statistics

INTRODUCTION

The human visual system processes an enormous throughput of sensory data in successive operations to generate percepts and behaviours necessary for biological functioning (Anderson et al., 2005; Raichle, 2010). Computations in the early visual cortex are often explained through unsupervised normative models which, given an input dataset with statistics similar to our surroundings, carry out an optimisation of criteria such as energy consumption and information-theoretic efficiency (Olshausen and Field, 1996; Bell and Sejnowski, 1997;

van Hateren and van der Schaaf, 1998; Hoyer and Hyvärinen, 2000; Zhaoping, 2006; Bruce et al., 2016). While such approaches could explain why many properties of the early visual system are closely related to characteristics of natural scenes (Olshausen and Field, 1996; Bell and Sejnowski, 1997; Lee and Seung, 1999; Geisler, 2008; Hunter and Hibbard, 2015; Beyeler et al., 2019), they are not equipped to answer questions such as how cortical structures which support complex computational operations implied by such optimisation may emerge, how these structures adapt, even in adulthood (Wandell and Smirnakis, 2010; Hübener and Bonhoeffer, 2014), and why some neurones possess receptive fields (RFs) which are sub-optimal in terms of information processing (Jones and Palmer, 1987; Ringach, 2002).

It is now well established that locally driven synaptic mechanisms such as spike-timing dependent plasticity (STDP) are natural processes which play a pivotal role in shaping the computational architecture of the brain (Markram et al., 1997; Delorme et al., 2001; Caporale and Dan, 2008; Larsen et al., 2010; Masquelier, 2012; Brito and Gerstner, 2016). Indeed, locally operating implementations of generative schemes have been shown to be closer to biological measurements (see, e.g., Rozell et al., 2008). Therefore, it is only natural to hypothesise that locally operating, biologically plausible models of plasticity must offer a better description of RFs in early visual cortex. However, such line of reasoning leads to the obvious question: what exactly constitutes a “better description” of a biological system, and more specifically, the early visual cortex. Here, we use a series of criteria spanning across electrophysiology, information theory, and machine learning, to investigate how descriptions of early visual RFs provided by a local, abstract STDP model compare to biological data from the macaque. We also compare these results to two classical, and important normative schemes – Independent Component Analysis (ICA), and Sparse Coding (SC). Our results demonstrate that a local process-based model of experience-driven plasticity may be better suited to capturing the RFs of simple-cells, thus suggesting that biological preference does not always concur with forms of global, generative optimality.

More specifically, we show that STDP units are able to better capture the characteristic sub-optimality in RF shape reported in literature (Jones and Palmer, 1987; Ringach, 2002), and their orientation tuning closely matches measurements in the macaque primary visual cortex (V1) (Ringach et al., 2002). Taken together, our findings suggest that while the information carrying capacity of an STDP ensemble is not optimal when compared to generatively optimal schemes, it is precisely this sub-optimality which may make process-based, local models more suited for describing the initial stages of sensory processing.

MATERIALS AND METHODS

Dataset

The Hunter–Hibbard dataset of natural images was used (Hunter and Hibbard, 2015) for training. It is available under the MIT license at <https://github.com/DavidWilliamHunter/Bivis>, and consists of 139 stereoscopic images of natural scenes captured using a realistic acquisition geometry and a 20° field of view.

Only images from the left channel were used, and each image was resized to a resolution of 5 px/° along both horizontal and vertical directions. Inputs to all encoding schemes were 3 × 3° patches (i.e., 15 × 15 px) sampled randomly from the dataset (**Figure 1A**).

Encoding Models

Samples from the dataset were used to train and test three models corresponding to the ICA, SC, and STDP encoding schemes. Each model consisted of three successive stages (**Figure 1B**). The first stage represented retinal activations. This was followed by a pre-processing stage implementing operations which are typically associated with processing in the lateral geniculate nucleus (LGN), such as whitening and decorrelation. In the third stage, LGN output was used to drive a representative V1 layer.

During learning, 10⁵ patches (3 × 3°) were randomly sampled from the dataset to simulate input from naturalistic scenes. In this phase, the connections between the LGN and V1 layers were plastic, and modified in accordance with one of the three encoding schemes. Care was taken to ensure that the sequence of inputs during learning was the same for all three models. After training, the weights between the LGN and V1 layers were no longer allowed to change. The implementation details of the three models are described below.

Sparse Coding

Sparse Coding algorithms are based on energy-minimisation, which is typically achieved by a “sparsification” of activity in the encoding population. We used a now-classical SC scheme proposed by Olshausen and Field (1996, 1997). The pre-processing in this scheme consists of an initial whitening of the input using low pass filtering, followed by a trimming of higher frequencies. The latter was employed to counter artefacts introduced by high frequency noise, and the effects of sampling across a uniform square grid. In the frequency domain the pre-processing filter was given by a zero-phase kernel:

$$H(f) = f \cdot e^{-\left(\frac{f}{f_0}\right)^4} \quad (1)$$

Here, $f_0 = 10$ cycles/° is the cut-off frequency. The outputs of these LGN filters were then used as inputs to the V1 layer composed of 225 units (3° × 3° RF at 5 px/°). The total number of weights in the model was 50,625. Retinal projections of the converged RFs were recovered by an approximate reverse-correlation algorithm (Ringach, 2002; Ringach and Shapley, 2004) derived from a linear-stability analysis of the SC objective function about its operating point. The RFs (denoted as columns of a matrix, say ξ) were given by:

$$\xi = \mathbf{A}[\mathbf{A}^T \mathbf{A} + \lambda \mathbf{S}''(0) \mathbf{I}]^{-1} \quad (2)$$

Here, \mathbf{A} is the matrix containing converged sparse components as column vectors, λ is the regularisation parameter (for the reconstruction, it is set to 0.14 σ , where σ^2 is the variance in the input dataset), and $S(x)$ is the shape-function for the prior distribution of the sparse coefficients [this implementation uses $\log(1 + x^2)$].

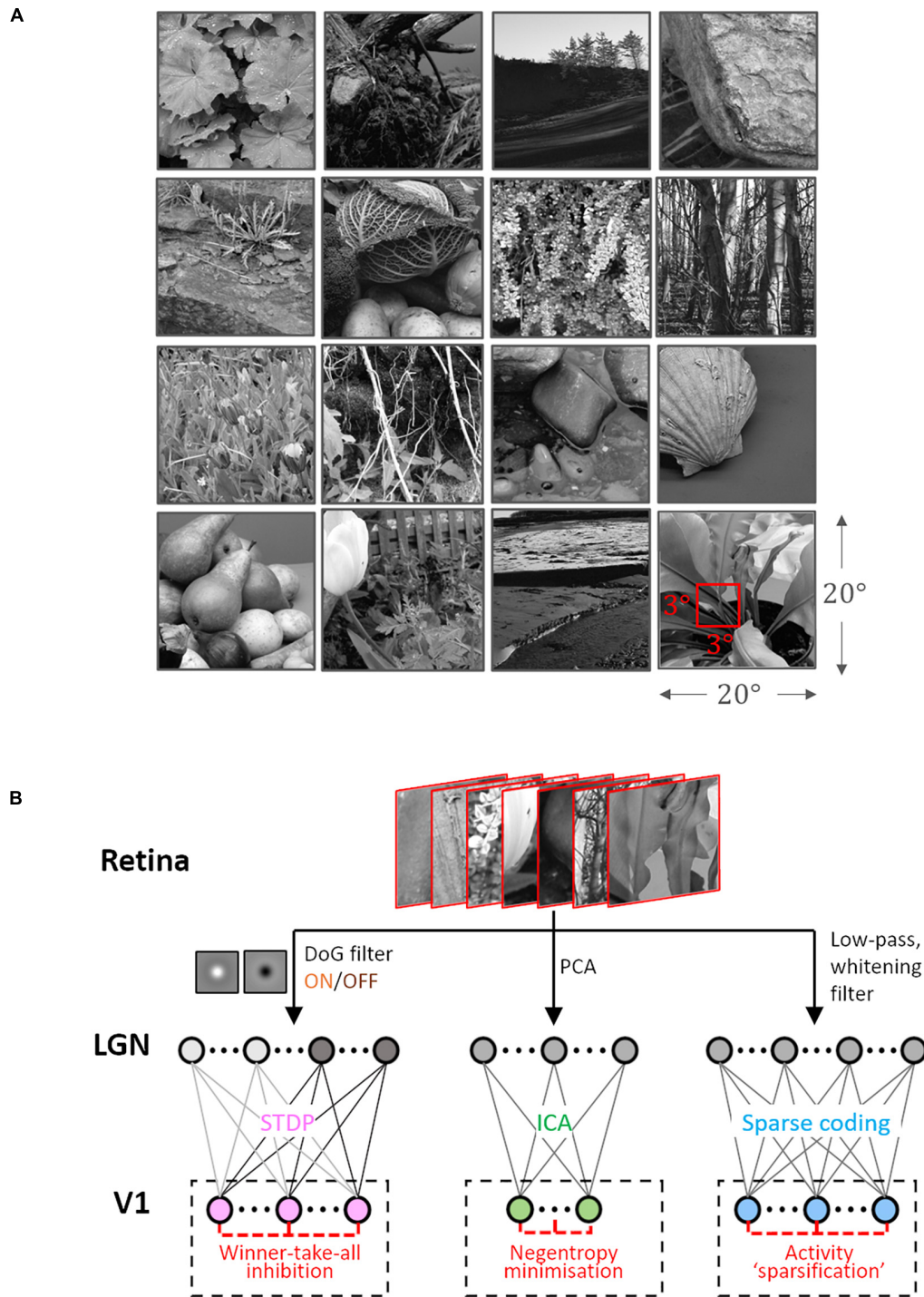


FIGURE 1 | Dataset and the computational pipeline. **(A)** Training data. The Hunter-Hibbard dataset of natural images was used. The images in the database have a $20^\circ \times 20^\circ$ field of view. Patches of size $3^\circ \times 3^\circ$ were sampled from random locations in the images (overlap allowed). The same set of 100,000 randomly sampled patches was used to train three models: Spike-timing dependent plasticity (STDP), Independent Component Analysis (ICA), and Sparse Coding (SC). **(B)** Modelling the early visual pathway. Three representative stages of early visual computation were captured by the models: retinal input, processing in the lateral geniculate nucleus (LGN), and the activity of early cortical populations in the primary visual cortex (V1). Each input patch represented a retinal input. This was followed by filtering operations generally associated with the LGN, such as decorrelation and whitening. Finally, the output from the LGN units/filters was connected to the V1 population through all-to-all (dense) plastic synapses which changed their weights during learning. Each model had a specific optimisation strategy for learning: the STDP model relied on a local rank-based Hebbian rule, ICA minimised mutual information (approximated by the negentropy), and SC enforced sparsity constraints on V1 activity. DoG, difference of Gaussian; PCA, Principal Component Analysis.

Independent Component Analysis

Independent Component Analysis algorithms are based on the idea that the activity of an encoding ensemble must be as information-rich as possible. This typically involves a maximisation of mutual information between the retinal input and the activity of the encoding ensemble. We used a classical ICA algorithm called *fastICA* (Hyvärinen and Oja, 2000) which achieves this through an iterative estimation of input negentropy. The pre-processing in this implementation was performed using a truncated Principal Component Analysis (PCA) transform ($\tilde{d} = 150$ components were used), leading to low-pass filtering and local decorrelation akin to centre-surround processing reported in the LGN. The model fit a total of 33,750 weights. If the input patches are denoted by the columns of a matrix (say \mathbf{X}), the LGN activity \mathbf{L} can be written as:

$$\mathbf{L} = \tilde{\mathbf{U}}^T \mathbf{X}_C \quad (3)$$

Here, $\mathbf{X}_C = \mathbf{X} - \langle \mathbf{X} \rangle$ and $\tilde{\mathbf{U}}$ is a matrix composed of the first \tilde{d} ($= 150$) principal components of \mathbf{X}_C . The activity of these LGN filters was then used to drive the ICA V1 layer consisting of 150 units, with its activity Σ being given by:

$$\Sigma = \mathbf{W} \mathbf{L} \quad (4)$$

Here, \mathbf{W} is the un-mixing matrix which is optimised during learning. The recovery of the RFs for ICA was relatively straight forward, as, in our implementation, they were assumed to be equivalent to the filters which must be applied to a given input to generate the corresponding V1 activity. The RFs (denoted as columns of a matrix, say ξ) were given by:

$$\xi = \tilde{\mathbf{U}} \mathbf{W}^T + \langle \mathbf{X} \rangle \quad (5)$$

Spike-Timing Dependent Plasticity

Spike-timing dependent plasticity is a biologically observed, Hebbian-like learning rule which relies on local spatiotemporal patterns in the input. We used a feedforward model based on an abstract rank-based STDP rule (Chauhan et al., 2018). The pre-processing in the model consisted of half-rectified ON/OFF filtering using difference-of-Gaussian kernels based on the properties of magno-cellular LGN cells. The outputs of these filters were converted to relative first-spike latencies using a monotonically decreasing function ($1/x$ was used), and only the earliest 10% spikes were allowed to propagate to V1 (Delorme et al., 2001; Masquelier and Thorpe, 2007). For each iteration, spikes within this 10% window were used to drive an unsupervised network of 225 integrate-and-fire neurones. The membrane potential of a V1 neurone was given by:

$$u(t) = H(\theta - u) \sum_{i \in \text{LGN}} w_i(t) \delta(t - t_i) \quad (6)$$

Here, t_i denotes the latency of the i -th pre-synaptic neuron, H is the Heaviside function, and θ is the spiking threshold. During learning, changes in the synaptic weights between LGN and V1 were governed by a rank-based, simplified version of the STDP rule proposed by Gütig et al. (2003). After each iteration,

the change (Δw) in the weight (w) of a given synapse was given by:

$$\Delta w = \begin{cases} -\alpha^- \cdot (w - w_{min})^{\mu^-} K(t, \tau^-), & t \leq 0 \\ \alpha^+ \cdot (w_{max} - w)^{\mu^+} K(t, \tau^+), & t > 0 \end{cases} \quad (7)$$

Here, Δt is the difference between the post- and pre-synaptic spike times, the constants α^\pm describe the learning rates for long-term potentiation (LTP) and depression (LTD), respectively, $\mu^\pm \in [0, 1]$ characterise the non-linearity of the multiplicative updates, K is a windowing function, and τ^\pm are the time-scales for LTP and LTD windows. Note that w is soft-bound such that $w \in (w_{min}, w_{max})$. The model used $w_{min} = 0$ (thalamocortical connections are known to be excitatory in nature), and $w_{max} = 1$. Since the intensity-to-latency conversion operates on an arbitrary time-scale, weight updates were based on the spike-order rather than precise spike-timing (rank-based). This meant that the window for LTP (τ^+) was variable and driven by the first 10% thalamic spikes, while the window for LTD (τ^-) was theoretically infinite. During updates, the weight was increased if a presynaptic spike occurred before the postsynaptic spike (causal firing), and decreased if it occurred after the post-synaptic spike (acausal firing). The learning rates were $\alpha^+ = 5 \times 10^{-3}$ and $\alpha^- = 0.75 \times \alpha^+$, and the nonlinearities were $\mu^+ = 0.65$ and $\mu^- = 0.05$. The model has previously been shown to be robust to both internal and external noise, and the parameter values were chosen from a range which best approximates the behaviour of the model under a realistic, V1-like regime (Chauhan et al., 2018). The neural population was homogeneous, with each neuron described by the exact same set of parameters.

During each iteration of learning, the population followed a winner-take-all inhibition rule wherein the firing of one neurone reset the membrane potentials of all other neurones. A total of 50,625 weights were fit by the model. After learning, this inhibition was no longer active and multiple units were allowed to fire for each input – allowing us to measure the behaviour of the network during testing. This also renders the model feed-forward only, making it comparable to SC and ICA. The RFs of the converged neurones were recovered using a linear approximation. If w_i denotes the weight of the synapse connecting a given neurone to the i th LGN filter with RF ψ_i , the RF ξ of the neurone was given by:

$$\xi = \sum_{i \in \text{LGN}} w_i \psi_i \quad (8)$$

Evaluation Metrics

Gabor Fitting

Linear approximations of RFs obtained by each encoding strategy were fitted using 2-D Gabor functions. This is motivated by the fact that all the encoding schemes considered here lead to linear, simple-cell-like RFs. In this case, the goodness-of-fit parameter (R^2) provides an intuitive measure of how Gabor-like a given RF is. The fitting was carried out using an adapted version of the code available at <https://uk.mathworks.com/matlabcentral/fileexchange/60700-fit2dgabor-data-options> (Ecke et al., 2021).

Frequency-Normalised Spread Vector

The shape of the RFs approximated by each encoding strategy was characterised using frequency-normalised spread vectors (FSVs) (Ringach, 2002; Chauhan et al., 2018). For a RF fitted by a Gabor-function with sinusoid carrier frequency f and envelope size $\sigma = [\sigma_x \ \sigma_y]^T$, the FSV is given by:

$$[n_x \ n_y]^T = [\sigma_x \ \sigma_y]^T f \quad (9)$$

While n_x provides an intuition of the number of cycles in the RF, n_y is a cycle-adjusted measure of the elongation of the RF perpendicular to the direction of sinusoid propagation. The FSV serves as a compact, intuitive descriptor of the RF shape-invariance to affine operations such as translation, rotation, and isotropic scaling.

Orientation Tuning

Orientation tuning curves (OTCs) were estimated by presenting each unit in each model with noisy oriented sine-wave grating (SWG) stimuli at its preferred frequency. The orientation was sampled in steps of 2° in the interval $[0^\circ, 180^\circ]$. For each orientation, the activity was averaged over phase values uniformly sampled in the interval $[0^\circ, 360^\circ]$ using a step-size of 5° . The bandwidth of an OTC was taken as its half-width at $1/\sqrt{2}$ of the peak response (Ringach et al., 2002). The whole process was repeated 100 times, and a bootstrap procedure was used to determine 95% confidence intervals.

Fisher Information

The information content in the activity of the converged units was quantified by using approximations of the Fisher information (FI, denoted here by the symbol J). If $\mathbf{x} = \{x_1, x_2, x_3, \dots, x_N\}$ is a random variable describing the activity of an ensemble of N independent units, the FI of the population with respect to a parameter θ is given by:

$$J(\theta) = \sum_{i=1}^N E \left[\left\{ \frac{\partial}{\partial x} \ln P(x_i|\theta) \right\}^2 \right]_{P(x_i|\theta)} \quad (10)$$

Here, $E[\cdot]_{P(x_i|\theta)}$ denotes expectation value with respect to the firing-state probabilities of the i th neurone in response to the stimuli corresponding to parameter value θ . In our simulations, θ was the orientation (defined as the direction of travel) of a set of SWGs with additive Gaussian noise, and was sampled at intervals of 4° in the range $[0^\circ, 180^\circ]$. The SWGs were presented at frequency of 1.25 cycles/visual degree, and the responses were calculated by averaging over 8 evenly spaced phase values in $[0^\circ, 360^\circ]$. This effectively simulated a drifting grating design within the constraints of the computational models. Each simulation was repeated 100 times and a jackknife procedure was used to estimate 95% confidence intervals. Noise was added such that the signal-to-noise ratio (SNR) varied between -6 and 6 dB in steps of 1 dB.

Decoding Using a Linear Classifier

In addition to FI approximations, we also used a linear decoder on the population responses obtained in the FI simulations. The decoder was an error-correcting output codes model composed

of binary linear-discriminant classifiers configured in a one-vs.-all scheme. Similar to the FI experiment, ground-truth values of the orientation at intervals of 4° in the range $[0^\circ, 180^\circ]$ were used as the class labels, and the activity generated by the corresponding SWG stimuli with added Gaussian noise was used as the training/testing data. The SWGs were presented at a frequency of 1.25 cycles/visual degree, and the responses were calculated by averaging over 8 evenly spaced phase values in $[0^\circ, 360^\circ]$. Each simulation was repeated 100 times, each time with five-fold validation. A jackknife procedure was used to estimate 95% confidence intervals.

Post-convergence Threshold Variation in STDP

To test how post-learning changes in the threshold affect the specificity of a converged network, we tested an STDP network trained using a threshold θ_{training} by increasing or decreasing its threshold (to say, θ_{testing}) and presenting it with SWGs (same stimuli as the ones used to calculate the FI). We report the results of seven simulations where the relative change in threshold was given by 25% increments/decrements, i.e.:

$$\frac{\theta_{\text{testing}} - \theta_{\text{training}}}{\theta_{\text{training}}} = \{0, \pm 0.25, \pm 0.50, \pm 0.75\} \quad (11)$$

Kullback–Leibler Divergence

For each model, we estimated probability density functions (pdfs) over parameters such as the FSVs and the population bandwidth. To quantify how close the model pdfs were to those estimated from the macaque data, we employed the Kullback–Leibler (KL) divergence. KL divergence is a directional measure of distance between two probability distributions. Given two distributions P and Q with corresponding probability densities p and q , the KL divergence (denoted D_{KL}) of P from Q is given by:

$$D_{KL}(P||Q) = \int_{\Omega} p(\mathbf{x}) \log_2 \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) d\mathbf{x} \quad (12)$$

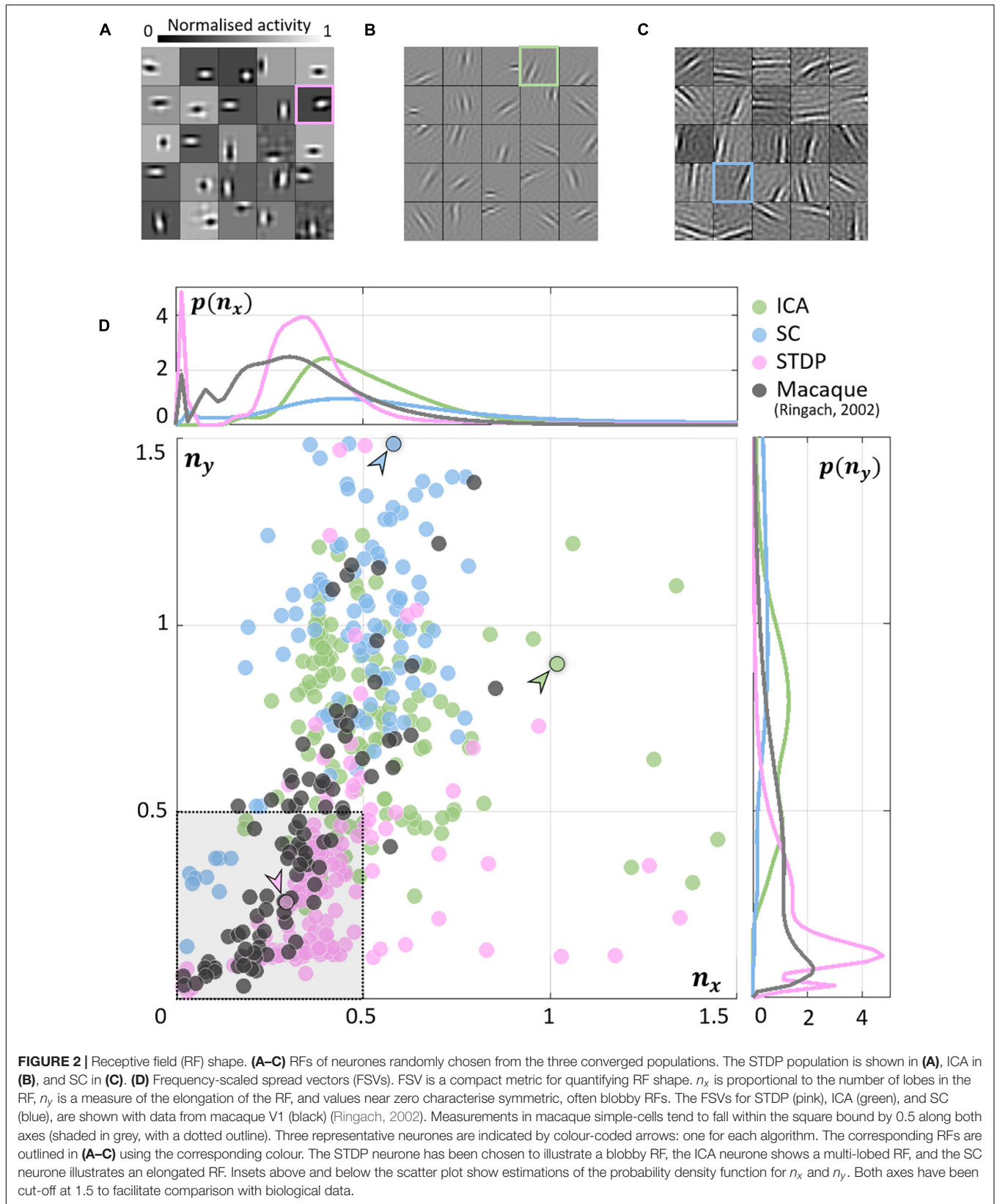
Here, Ω is the support of the distribution Q . In our analysis, we considered the reference distribution p as a pdf estimated from the macaque data, and q as the pdf (of the same variable) estimated using ICA, SC, or STDP. In this case, KL divergence lends itself to a very intuitive interpretation: it can be considered as the additional bandwidth (in bits) which would be required if the biological variable were to be encoded using one of the three computational models. Note that P and Q may be multivariate distributions.

Sparsity: Gini Index

The sparseness of the encoding was evaluated using the Gini index (GI). GI is a measure which characterises the deviation of the population-response from a uniform distribution of activity across the samples. Formally, the GI (denoted here as Λ) is given by:

$$\Lambda(\mathbf{x}) = 1 - 2 \int_0^1 L(F) dF \quad (13)$$

Here L is the Lorenz function defined on the cumulative probability distribution F of the neural activity (say, \mathbf{x}). GI is 0 if all units have the same response and tends to 1 as responses become sparser (being equal to 1 if only 1 unit responds, while others are silent). It is invariant to the range of the responses



within a given sample, and robust to variations in sample-size (Hurley and Rickard, 2009). We defined two variants of the GI which measure the spatial (Λ_s) and temporal sparsity (Λ_t) of an ensemble of encoders. Given a sequence of M inputs to an ensemble of N neurones, the spatial sparsity of the ensemble response to the m th stimulus is given by:

$$\Lambda_S(m) = \Lambda(\{x_m^1, x_m^2, \dots, x_m^N\}) \quad (14)$$

Here, x_m^n denotes the activity of the n th neurone in response to the m th input. Similarly, the temporal sparsity of the n th neurone over the entire sequence of inputs is given by:

$$\Lambda_T(n) = \Lambda(\{x_1^n, x_2^n, \dots, x_M^n\}) \quad (15)$$

Code

The code for ICA was written in python using the sklearn library which implements the classical *fastICA* algorithm. The code for SC was based on the C++ and Matlab code shared by Prof. Bruno Olshausen. The STDP code was based on a previously published binocular-STDP algorithm available at <https://senselab.med.yale.edu/ModelDB/showmodel.cshtml?model=245409#tabs-1>.

RESULTS

We used an abstract model of the early visual system with three representative stages: retinal input, LGN processing, and V1 activity (Figure 1B). To simulate retinal activity corresponding to natural inputs, patches of size $3^\circ \times 3^\circ$ (visual angles) were sampled randomly from the Hunter–Hibbard database (Hunter and Hibbard, 2015) of natural scenes (Figure 1A). 10^5 patches were used to train models corresponding to three encoding schemes: ICA, SC, and STDP. Each model used a specific procedure for implementing the LGN processing and learning the synaptic weights between the LGN and V1 (see Figure 1B and section “Materials and Methods”).

Receptive Field Symmetry

As expected, units in all models converged to oriented, edge-detector like RFs. While the RFs from ICA (Figure 2B) and SC (Figure 2C) were elongated and highly directional, STDP (Figure 2A) RFs were more compact and less sharply tuned. This is closer to what is observed from simple-cell recordings in the macaque (Ringach, 2002) where RFs show high circular symmetry, and do not seem to be optimally tuned for edge-detection (see Jones and Palmer, 1987 for similar data measured in the cat). To obtain a quantitative measure of this phenomenon, we fit Gabor functions to the RFs and considered the frequency-normalised spread vectors or FSVs of the fit (Eq. 9). The first component (n_x) of the FSV characterises the number of lobes in the RF, and the second component (n_y) is a measure of the elongation of the RF (perpendicular to carrier propagation). A considerable number of simple-cell RFs measured in macaque tend to fall within the square bounded by $n_x = 0.5$ and $n_y = 0.5$. The FSVs of a sample of neurones ($N = 93$) measured in the macaque V1 (Ringach, 2002) indicate that 59.1% of the neurones lay within this region (Figure 2D). Since they

are not elongated, and contain few lobes (typically 2–3 on/off regions), they tend to be compact – making them less effective as edge-detectors compared to more crisply tuned, elongated RFs. Amongst the three encoding schemes, while a considerable number of STDP units (82.2%) tended to fall within these realistic boundaries, ICA (10.7%) and SC (4.0%) showed a distinctive shift upwards and to the right. This trend has been observed in a number of studies using models based on ICA and SC (see, e.g., Rehn and Sommer, 2007; Puertas et al., 2010; Zylberberg et al., 2011).

The inlays in Figure 2D provide estimations of the probability densities of two FSV parameters for the macaque data and the three models. An interesting insight into these distributions is given by the KL divergence (Table 1). KL divergence (Eq. 12) is a directed measure which can be interpreted as the additional number of bits required if one of the three models were used to encode data sampled from the macaque distribution. The KL divergence for the STDP model was found to be 3.0 bits indicating that, on average, it would require three extra bits to encode data sampled from the macaque distribution. In comparison, SC and ICA were found to require 8.4 and 14.6 additional bits, respectively. An examination of the KL divergence of marginal distributions of the FSV parameters showed that STDP offers excellent encoding of both the n_x (number of lobes) and the n_y (compactness) parameter. ICA does not encode either of the two parameters satisfactorily, while SC performance is closer to the STDP model (especially for parameter n_x).

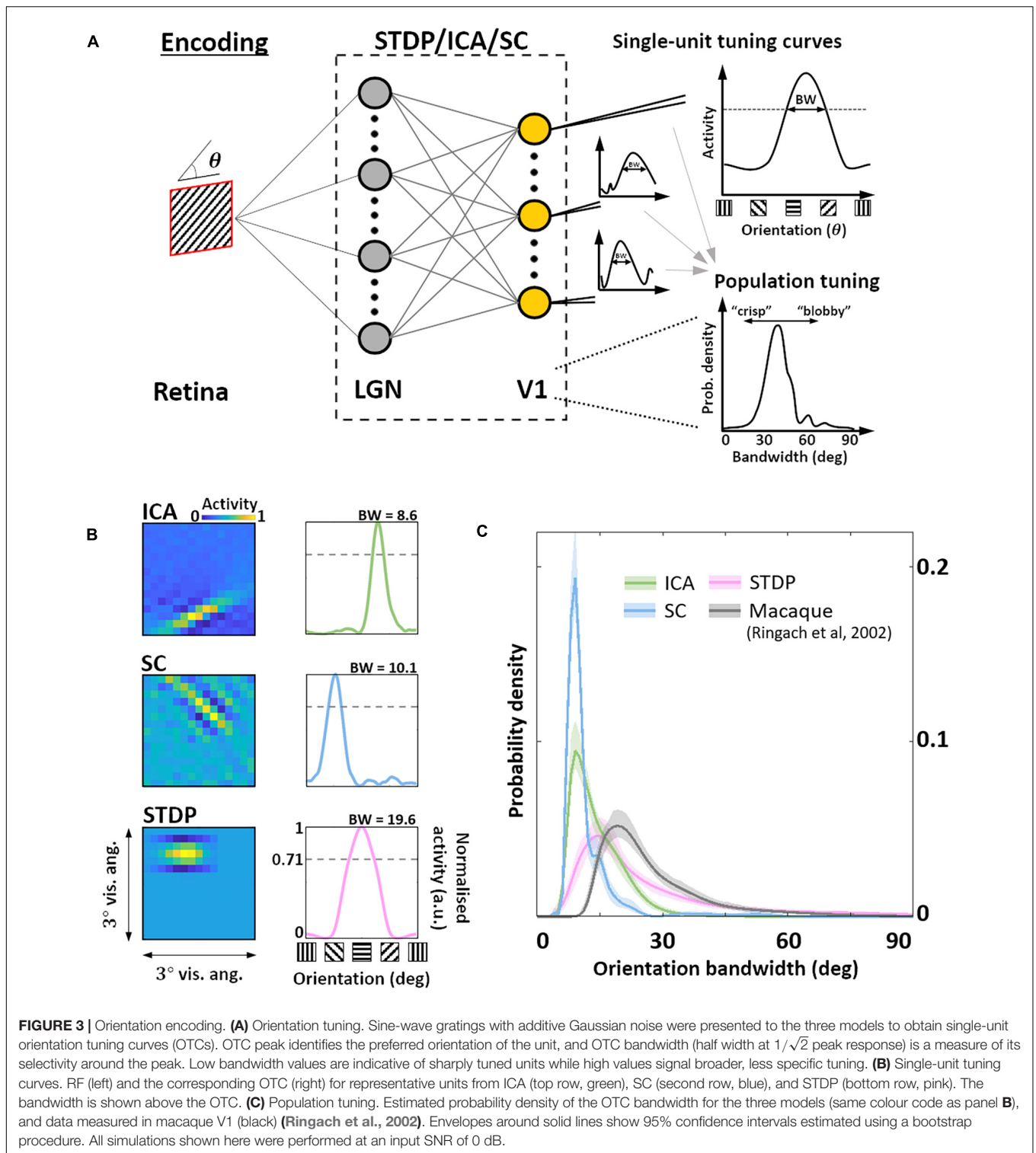
Orientation Selectivity

Given this sub-optimal, symmetric nature of STDP RF shapes, we next investigated how this affected the responses of these neurones to sharp edges. In particular, we were interested in how the orientation bandwidths of the units from the three models would compare to biological data. Given the RF shape, we hypothesised that orientation selectivity would be worse for STDP compared to the ICA and SC schemes. To test this hypothesis, we simulated a typical electrophysiological experiment for estimating orientation tuning (Figure 3A). To each unit, we presented noisy SWGs at its preferred spatial frequency and recorded its activity as a function of the stimulus

TABLE 1 | Kullback–Leibler (KL) divergence of the distribution of macaque frequency-normalised spread vectors (FSVs) from the models.

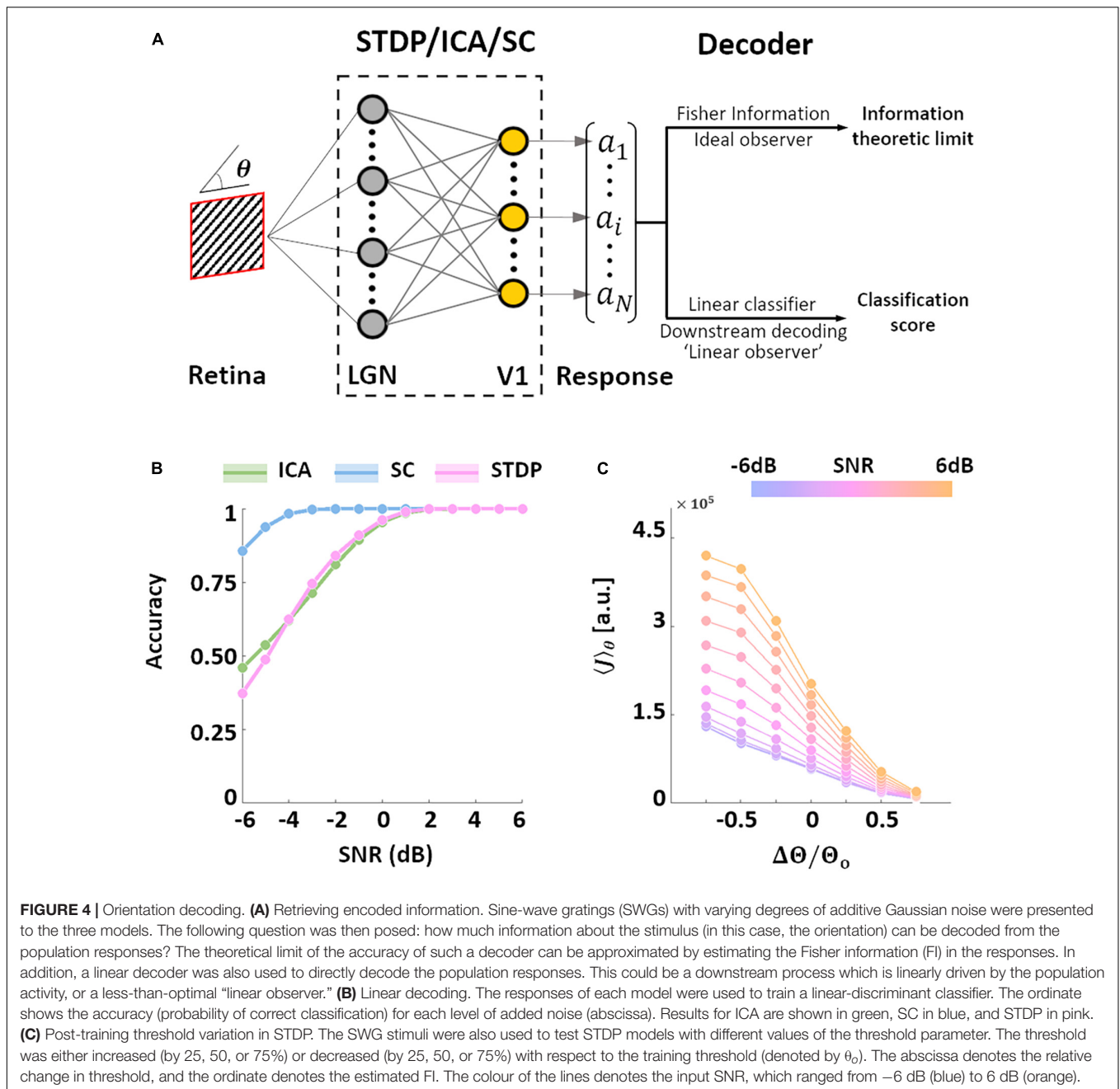
	ICA	SC	STDP
Joint distribution			
$[n_x \ n_y]^T$	14.6	8.4	3.0
Marginal distributions			
n_x	7.6	1.4	1.3
n_y	14.0	3.8	0.4

The receptive-field (RF) shape of the neurones from the models and measurements in macaque V1 (Ringach, 2002) was parametrised by estimating the frequency-normalised spread vectors (FSVs). FSVs are characterised by two parameters n_x and n_y : n_x is proportional to the number of lobes in the receptive field, and n_y is modulated by its elongation perpendicular to the direction of periodicity. The KL divergence reflects the number of additional bits required to encode the parameter(s) of interest from the macaque data using the distributions from one of the three models (ICA, SC, or STDP). All values are in bits.



orientation. This allowed us to plot its OTC (Figure 3B) and estimate the tuning bandwidth, which is a measure of the local selectivity of the unit around its peak – low values corresponding to sharply tuned neurones and higher values corresponding to broadly tuned, less selective neurones. For each of the three

models, we estimated the pdf of the OTC bandwidth, and compared it to the distribution estimated over a large set of data ($N = 308$) measured in macaque V1 (Ringach et al., 2002) (Figure 3C). We found that ICA and SC distributions peaked at a bandwidth of about 10° (ICA: 9.1° , SC: 8.5°) while the STDP



and macaque data showed much broader tunings (STDP: 15.1° , Macaque data: 19.1°). This was also reflected in the KL divergence of the macaque distribution from the three model distributions (ICA: 2.4 bits, SC: 3.5 bits, STDP: 0.29 bits). Thus, while the orientation tuning for STDP is much broader compared to ICA and SC, it is also closer to measurements in the macaque V1, indicating a better agreement with biology.

Decoding and Information Throughput

After characterising the encoding capacity of the models, we next probed the possible downstream implications of such codes. The biological goal of most neural code, in the end, is the generation of behaviour that maximises evolutionary fitness. However, due to

the complicated neural apparatus that separates behaviour from early sensory processing, it is not straightforward (or at times, even possible) to analyse the interaction between the two. Bearing these limitations in mind, we employed two separate metrics to investigate this relationship. In both cases, the models were presented with oriented SWGs, followed by a decoding analysis of the resulting neural population activity (**Figure 4A**).

We examined the performance of a decoder built on linear discriminant classifiers (these classifiers assume fixed first-order correlations in the input). Such decoders can be interpreted as linearly driven feedforward populations downstream from the thalamo-recipient layer (the “V1” populations in the three models), or a simplified, “linear” observer. Not surprisingly the

accuracy of the three models increases with the SNR. We found that SC was the most accurate of the three models under all tested noise-levels, while ICA and STDP showed very similar performances (**Figure 4B**). SC was also more robust to Gaussian noise compared to both ICA and STDP. A major difference between the three models tested in this study is that while ICA and SC are based on linear generative units, the STDP model has an intervening thresholding nonlinearity (Eq. 6). To test the effect of this thresholding on the information throughput of the STDP model, we ran simulations where, after training on natural images, the value of the threshold parameter in the STDP model was either increased or decreased (Eq. 11). The network was presented with SWGs (same stimuli as **Figure 4B**), and the average FI (Eq. 10) over the orientation parameter was estimated for each simulation condition. Note that in all simulations the model was first trained (i.e., synaptic learning using natural stimuli, see **Figure 1**) using the same “training” threshold, and the increase/decrease of the threshold parameter was imposed post-convergence. The FI increased for thresholds lower than the training threshold – possibly driven by an increase in the overall activity of the network. On the other hand, increasing the threshold led to lower FI due to the decreased bandwidth of neural activity. Thus, it is indeed possible to manipulate the information throughput of the spiking network by regulating the overall spiking activity in the network. This trend was found to occur robustly for all tested SNR values.

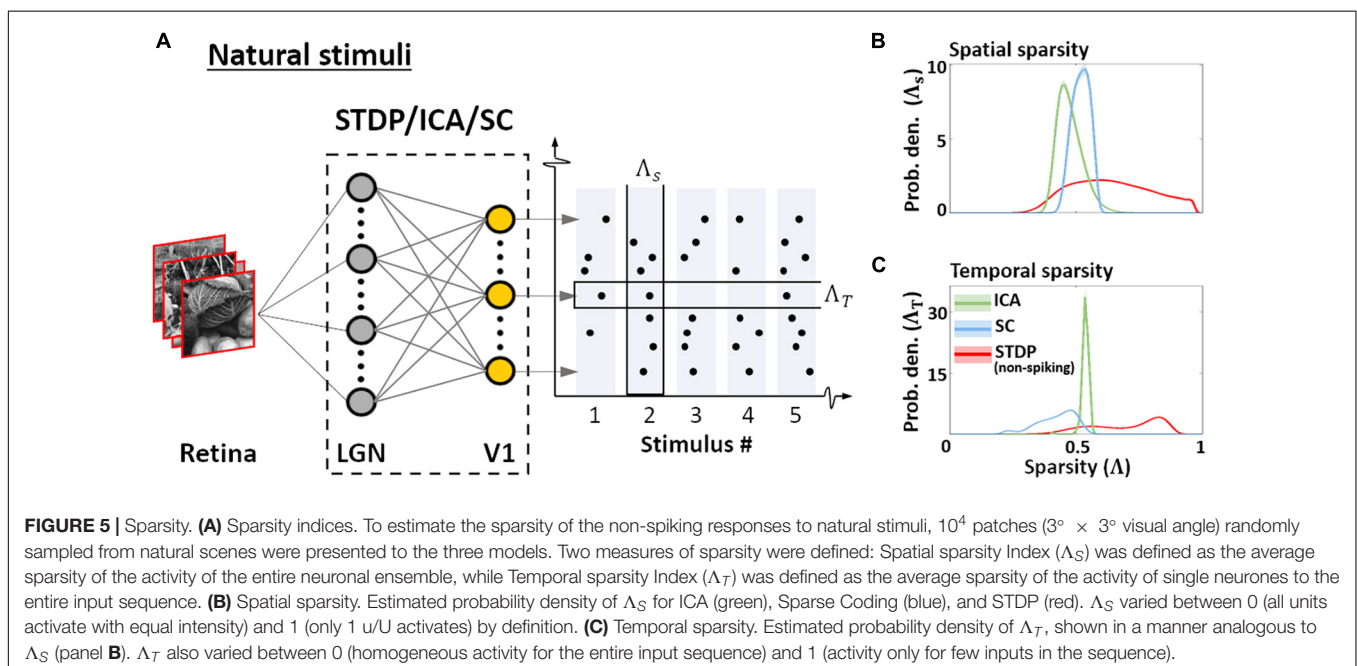
DISCUSSION

In this study, we showed that learning in a network with an abstract, rank-based STDP rule can predict biological findings at various scales. The FSVs of the converged RFs in the model show strong similarities with single-cell data measured in the

macaque primary visual cortex, while the OTCs in the model closely predict measured population tuning.

Optimality in Biological Systems

In neuroscience, normative schemes are typically used to relate natural stimuli to an encoding hypothesis. Most normative encoding schemes optimise a generative reconstruction of the input by minimising an error metric (e.g., the L1 or L2 losses) over a given dataset. An alternative approach to studying stimulus encoding is through the use of process-based schemes which model known biophysical mechanisms at various levels of abstraction without making explicit assumptions about optimality. Traditionally, process-based or mechanistic schemes do not employ error metrics, and have been used to study fine-grained neuronal dynamics (Kang and Sompolinsky, 2001; Moreno-Bote et al., 2014; Harnack et al., 2015). On the other hand, normative schemes are employed to describe population-level characteristics (Olshausen and Field, 1997; van Hateren and van der Schaaf, 1998; Lee and Seung, 1999; Hoyer and Hyvärinen, 2000). In this study, we show that RFs predicted by a non-generative rank-based STDP rule are closer to electrophysiological measurements in the macaque V1 when compared to generatively optimal schemes such as ICA and SC. While this study only employs the classical variations of ICA and SC, subsequent work has demonstrated that similar suboptimalities in RF shape can also be obtained by generative models when biologically plausible nonlinearities such as thresholding operations (Rehn and Sommer, 2007; Rozell et al., 2008), or pointwise maxima operations (Puertas et al., 2010) are introduced. However, the abstract rank-based STDP model used here is free from generative optimisation and offers a much more biologically plausible, normative description of “learning”



through experience in the early visual system, where there is no sensory “ground truth” to generate errors from.

Note that while process-based models can predict suboptimalities observed in biological data, they cannot account for the theoretical insights offered by generative normative schemes. Local synaptic processes such as STDP can, in fact, be viewed as neural substrates for the overall synaptic optimisation employed by these schemes. The critique that gradient descent is inherently biologically implausible is being challenged by recent studies which frame error propagation and stochastic descent in terms of local, biologically plausible rules (see, e.g., Lillicrap et al., 2016; Melchior and Wiskott, 2019; Li, 2020). It has been demonstrated that local plasticity rules can, in fact, be adapted to describe various normative hypotheses about stimulus encoding (Savin et al., 2010; Brito and Gerstner, 2016). A growing number of insightful studies now employ hybrid encoding schemes which address multiple optimisation criteria (Perrinet and Bednar, 2015; Martinez-Garcia et al., 2017; Beyeler et al., 2019), often through local biologically realistic computation (Rozell et al., 2008; Savin et al., 2010; Zylberberg et al., 2011; Isomura and Toyozumi, 2018).

Sparsity

Normative descriptions of the early visual system are grounded in the idea of efficiency – in terms of information transfer, and in terms of resource consumption. These assumptions, in turn, determine the behaviour of population responses to natural images. We quantified this behaviour by presenting the converged models with patches randomly sampled from the training dataset of natural images, and estimating the sparsity of the resulting activations using the Gini coefficient (Eq. 13; Hurley and Rickard, 2009). The sparsity was examined in two contexts (Barth and Poulet, 2012) as shown in **Figure 5A**. First, sparsity of the entire ensemble was estimated for each presented stimulus – this is a measure of how many neurones, on average, are employed by the ensemble to encode a given stimulus (Eq. 14). Second, the sparsity of individual neurones over the entire sequence of stimuli was estimated, allowing us to infer how frequently the features selected/encoded by the converged models occur in the sequence (Eq. 15). We denote the former as spatial sparsity (Λ_s), and the latter as temporal sparsity (Λ_t). For STDP, the indices were calculated for the membrane potential to facilitate comparison with ICA and SC activations. STDP membrane potential (red, **Figure 5B**) shows high variability in Λ_s , whereas ICA (green) and SC (blue) show much lower variance in comparison. This suggests that ICA and SC converge to features such that each image activates approximately equal number of units. On the other hand, the sparsity of the STDP neurones is more variable and stimulus-dependent, and likely driven by the relative probability of occurrence of specific features in the dataset – thus reflecting the Hebbian principal. ICA also exhibits a similar, small range for temporal sparsity Λ_t (**Figure 5C**) – suggesting that ICA encoding has uniform activation probability across its units. SC and STDP, however, show a much broader range of temporal sparsity across their units, with some units activating more frequently as compared to others.

Taken together, this suggests that the ICA encoding scheme converges to features such that the activation is distributed uniformly across the units, both for a given stimulus, and across multiple stimuli. This is likely to be driven by the objective of minimising reconstruction loss while maintaining minimal mutual information across the population. SC, on the other hand, equalises the probability of firing over the population for any given stimulus, but individual units may converge to features which occur more or less frequently. Once again, this behaviour is a consequence of the loss function which ensures that the network activity is sparse for each stimulus, but does not impose explicit constraints between the activity profile of individual units. As the STDP model is unsupervised and does not explicitly impose any generative loss function, we find high variability in both the spatial and temporal sparsity of STDP units. As shown in **Figure 4C**, this variability ensures that the information throughput of the network can be modulated by regulation of parameters such as the spiking threshold, even after the initial training.

Emerging Technologies and Process-Based Modelling in Neuroscience

Traditionally, detailed process-based models have suffered from constraints imposed by computational complexity, prohibitively long execution times which do not scale well for large networks, and hardware that is geared toward synchronous processing. On the other hand, most normative models can leverage faster computational libraries and architectures which have been developed over several decades, thereby leading to more efficient and scalable computation. However, with the growing availability of faster and more adaptable computing solutions such as neuromorphic hardware (event-based cameras, spike-based chips), and event-driven computational frameworks (e.g., Nengo: Bekolay et al., 2014; or Brian 2: Stimberg et al., 2019), implementations of such models are becoming increasingly accessible and scientifically tractable. These frameworks can be used not only to investigate detailed biophysical models or create biologically relevant machine and reinforcement learning pipelines, but to also investigate normative neuroscientific hypotheses which require unsupervised learning. In the future, we hope process-based modelling will be adopted more widely by cognitive and computational neuroscientists alike.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analysed in this study. This data can be found here: <http://ringachlab.net/>.

AUTHOR CONTRIBUTIONS

TC, BC, and TM conceptualised the study and interpreted the results. TC designed, programmed, and ran the simulations and

wrote the original draft. All authors reviewed and revised the manuscript.

FUNDING

This work was funded by Fondation pour la Recherche Médicale (FRM: SPF20170938752, awarded to TC) and Agence

Nationale de la Recherche (ANR-16-CE37-002-01, 3D3M, awarded to BC).

ACKNOWLEDGMENTS

We would like to thank Yves Trotter and Lionel Nowak for their critical feedback.

REFERENCES

- Anderson, C., Van Essen, D., and Olshausen, B. (2005). "CHAPTER 3 - Directed visual attention and the dynamic control of information flow," in *Neurobiology of Attention*, eds L. Itti, G. Rees, and J. K. Tsotsos (Burlington, VT: Academic Press), 11–17. doi: 10.1016/B978-012375731-9/50007-0
- Barth, A. L., and Poulet, J. F. A. (2012). Experimental evidence for sparse firing in the neocortex. *Trends Neurosci.* 35, 345–355. doi: 10.1016/j.tins.2012.03.008
- Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T., Rasmussen, D., et al. (2014). Nengo: a Python tool for building large-scale functional brain models. *Front. Neuroinformatics* 7:48. doi: 10.3389/fninf.2013.00048
- Bell, A., and Sejnowski, T. (1997). The "independent components" of natural scenes are edge filters. *Vision Res.* 37, 3327–3338. doi: 10.1016/S0042-6989(97)00121-1
- Beyeler, M., Rounds, E., Carlson, K., Dutt, N., and Krichmar, J. (2019). Neural correlates of sparse coding and dimensionality reduction. *PLoS Comput. Biol.* 15:e1006908. doi: 10.1371/journal.pcbi.1006908
- Brito, C. S. N., and Gerstner, W. (2016). Nonlinear Hebbian learning as a unifying principle in receptive field formation. *PLoS Comput. Biol.* 12:e1005070. doi: 10.1371/journal.pcbi.1005070
- Bruce, N., Rahman, S., and Carrier, D. (2016). Sparse coding in early visual representation: from specific properties to general principles. *Neurocomputing* 171, 1085–1098. doi: 10.1016/j.neucom.2015.07.070
- Caporale, N., and Dan, Y. (2008). Spike timing-dependent plasticity: a Hebbian learning rule. *Annu. Rev. Neurosci.* 31, 25–46. doi: 10.1146/annurev.neuro.31.060407.125639
- Chauhan, T., Masquelier, T., Montlibert, A., and Cottureau, B. (2018). Emergence of binocular disparity selectivity through Hebbian learning. *J. Neurosci.* 38, 9563–9578. doi: 10.1523/JNEUROSCI.1259-18.2018
- Delorme, A., Perrinet, L., and Thorpe, S. J. (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. doi: 10.1016/S0925-2312(01)00403-9
- Ecke, G. A., Papp, H. M., and Mallot, H. A. (2021). Exploitation of image statistics with sparse coding in the case of stereo vision. *Neural Netw.* 135, 158–176. doi: 10.1016/j.neunet.2020.12.016
- Geisler, W. (2008). Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59, 167–192. doi: 10.1146/annurev.psych.58.110405.085632
- Gütig, R., Aharonov, R., Rotter, S., and Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *J. Neurosci.* 23, 3697–3714.
- Harnack, D., Pelko, M., Chaillet, A., Chitour, Y., and van Rossum, M. C. W. (2015). Stability of neuronal networks with homeostatic regulation. *PLoS Comput. Biol.* 11:e1004357. doi: 10.1371/journal.pcbi.1004357
- Hoyer, P., and Hyvärinen, A. (2000). Independent component analysis applied to feature extraction from colour and stereo images. *Netw. Comput. Neural Syst.* 11, 191–210. doi: 10.1088/0954-898X_11_3_302
- Hübener, M., and Bonhoeffer, T. (2014). Neuronal plasticity: beyond the critical period. *Cell* 159, 727–737. doi: 10.1016/j.cell.2014.10.035
- Hunter, D., and Hibbard, P. (2015). Distribution of independent components of binocular natural images. *J. Vis.* 15:6. doi: 10.1167/15.13.6
- Hurley, N., and Rickard, S. (2009). Comparing measures of sparsity. *IEEE Trans. Inf. Theory* 55, 4723–4741. doi: 10.1109/TIT.2009.2027527
- Hyvärinen, A., and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural Netw.* 13, 411–430. doi: 10.1016/S0893-6080(00)00026-5
- Isomura, T., and Toyozumi, T. (2018). Error-gated Hebbian rule: a local learning rule for principal and independent component analysis. *Sci. Rep.* 8:1835. doi: 10.1038/s41598-018-20082-0
- Jones, J., and Palmer, L. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. Neurophysiol.* 58, 1233–1258.
- Kang, K., and Sompolinsky, H. (2001). Mutual Information of population codes and distance measures in probability space. *Phys. Rev. Lett.* 86, 4958–4961. doi: 10.1103/PhysRevLett.86.4958
- Larsen, R. S., Rao, D., Manis, P. B., and Philpot, B. D. (2010). STDP in the developing sensory neocortex. *Front. Synaptic Neurosci.* 2:9. doi: 10.3389/fnsyn.2010.00009
- Lee, D., and Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788–791. doi: 10.1038/44565
- Li, H. L. (2020). Faster biological gradient descent learning. *ArXiv[Preprint] ArXiv 200912745 Cs*,
- Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nat. Commun.* 7:13276. doi: 10.1038/ncomms13276
- Markram, H., Lübke, J., Frotscher, M., Sakmann, B., Hebb, D. O., Bliss, T. V. P., et al. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275, 213–215. doi: 10.1126/science.275.5297.213
- Martinez-Garcia, M., Martinez, L. M., and Malo, J. (2017). Topographic Independent Component Analysis reveals random scrambling of orientation in visual space. *PLoS One* 12:e0178345. doi: 10.1371/journal.pone.0178345
- 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. doi: 10.1007/s10827-011-0361-9
- Masquelier, T., and Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput. Biol.* 3:e31. doi: 10.1371/journal.pcbi.0030031
- Melchior, J., and Wiskott, L. (2019). Hebbian-Descent. *ArXiv[Preprint] ArXiv190510585 Cs* Stat. Available online at: <http://arxiv.org/abs/1905.10585> [Accessed August 20, 2021],
- Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci.* 17, 1410–1417. doi: 10.1038/nn.3807
- Olshausen, B., and Field, D. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381, 607–609. doi: 10.1038/381607a0
- Olshausen, B., and Field, D. (1997). Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Res.* 37, 3311–3325. doi: 10.1016/S0042-6989(97)00169-7
- Perrinet, L. U., and Bednar, J. A. (2015). Edge co-occurrences can account for rapid categorization of natural versus animal images. *Sci. Rep.* 5:11400. doi: 10.1038/srep11400
- Puertas, J., Bornschein, J., and Lücke, J. (2010). "The maximal causes of natural scenes are edge filters," in *Advances in Neural Information Processing Systems*, eds J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta (Red Hook, NY: Curran Associates, Inc), 1939–1947.
- Raichle, M. (2010). Two views of brain function. *Trends Cogn. Sci.* 14, 180–190. doi: 10.1016/j.tics.2010.01.008
- Rehn, M., and Sommer, F. (2007). A network that uses few active neurones to code visual input predicts the diverse shapes of cortical receptive fields. *J. Comput. Neurosci.* 22, 135–146. doi: 10.1007/s10827-006-0003-9

- Ringach, D. (2002). Spatial structure and symmetry of simple-cell receptive fields in Macaque primary visual cortex. *J. Neurophysiol.* 88, 455–463.
- Ringach, D., and Shapley, R. (2004). Reverse correlation in neurophysiology. *Cogn. Sci.* 28, 147–166. doi: 10.1207/s15516709cog2802_2
- Ringach, D., Shapley, R., and Hawken, M. (2002). Orientation selectivity in Macaque V1: diversity and laminar dependence. *J. Neurosci.* 22, 5639–5651. doi: 10.1523/JNEUROSCI.22-13-05639.2002
- Rozell, C., Johnson, D., Baraniuk, R., and Olshausen, B. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Comput.* 20, 2526–2563. doi: 10.1162/neco.2008.03-07-486
- Savin, C., Joshi, P., and Triesch, J. (2010). Independent component analysis in spiking neurons. *PLoS Comput. Biol.* 6:e1000757. doi: 10.1371/journal.pcbi.1000757
- Stimberg, M., Brette, R., and Goodman, D. F. (2019). Brian 2, an intuitive and efficient neural simulator. *eLife* 8:e47314. doi: 10.7554/eLife.47314
- van Hateren, J., and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Biol. Sci.* 265, 359–366. doi: 10.1098/rspb.1998.0303
- Wandell, B., and Smirnakis, S. (2010). Plasticity and stability of visual field maps in adult primary visual cortex. *Nat. Rev. Neurosci.* 10:873. doi: 10.1038/nrn2741
- Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Netw. Comput. Neural Syst.* 17, 301–334. doi: 10.1080/09548980600931995
- Zylberberg, J., Murphy, J. T., and DeWeese, M. R. (2011). A Sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of V1 simple cell receptive fields. *PLoS Comput. Biol.* 7:e1002250. doi: 10.1371/journal.pcbi.1002250

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Chauhan, Masquelier and Cottureau. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.