Stripe-rearing changes multiple aspects of the structure of primary visual cortex

Nicholas Hughes[a], Jonathan J. Hunt[a], Shaun L. Cloherty[b], Michael R. Ibbotson[b,c],
Frank Sengpiel[d], Geoffrey J. Goodhill[a,e]

[a] *Queensland Brain Institute, The University of Queensland, St Lucia, Queensland, 4072, Australia*
[b] *National Vision Research Institute, Australian College of Optometry, Carlton, Victoria, 3053, Australia*
[c] *Department of Optometry and Vision Science, University of Melbourne, Parkville, VIC, 3010, Australia*
[d] *School of Biosciences, Cardiff University, CF10 3AX, Wales, UK*
[e] *School of Mathematics and Physics, The University of Queensland, St Lucia, Queensland, 4072, Australia*
[*] *Corresponding author:* `g.goodhill@uq.edu.au`

Draft, May 24, 2013

Abstract

A paradigm example of brain plasticity is the change in the structure of the orientation map in primary visual cortex (V1) in response to an environment consisting of stripes of one orientation. However, until now this plasticity has been characterised only by the relative size of the areas of V1 representing different orientations. Here we extend to the stripe-reared case a recently-proposed Bayesian method for reconstructing orientation maps from intrinsic-signal optical imaging data, which allows a more nuanced assessment of changes in orientation map structure in response to stripe-rearing. We find that the structure of a normal map can change in several different ways to accommodate more preference for one orientation. Overall this work expands our understanding of the effects of the environment on brain structure.

1   Introduction

A critical question for understanding brain plasticity is to understand how the environment influences brain structure. A paradigmatic model system in this regard is how the form of visual activity during early life affects the spatial arrangement of topographic maps in the primary visual cortex (V1) of mammals such as cats, ferrets and monkeys. In particular the spatial arrangement of preference for orientation (Blasdel and Salama, 1986), direction (Weliky et al., 1996), spatial frequency (Hübener et al., 1997; Issa et al., 2000) and ocular dominance (Anderson et al., 1988; Bonhoeffer et al., 1995), as well as the relationships between these maps (Shmuel and Grinvald, 1996; Hübener et al., 1997), have been widely studied.

A particularly striking example of the effect of the environment on these maps is stripe-rearing, in which animals are exposed only to edges of a particular orientation during the critical period (Blakemore and Cooper, 1970). This leads to an increase in the proportion of V1 neurons preferring the reared orientation, at the expense of other orientations (Sengpiel et al., 1999). However, so far this proportion is the only quantitative measure that has been used to characterise the difference in structure between normal and stripe-reared orientation maps. This measure depends sensitively on the estimate of the map, which is traditionally obtained by vector-averaging noisy data from single-condition optical imaging experiments, followed by spatial filtering with parameters chosen subjectively. Furthermore this measure does not provide a method for determining whether there are more subtle changes in overall map structure that are induced by stripe rearing.

Recently a novel method was introduced which improves on vector averaging for determining orientation map structure, using a Bayesian approach based on Gaussian processes (Macke et al., 2011). This uses prior knowledge about map structure in a principled way rather than via subjectively-chosen smoothing parameters, and provides quantitative error estimates for the resulting map. However, Macke et al.'s original method contained assumptions which meant it could not be directly applied to estimating maps in abnormally reared animals.

We first generalised this Gaussian process method to a much broader class of visual map data, including abnormal rearing conditions. We then used it to interrogate map structure in data from the stripe-reared cats of Sengpiel et al. (1999). Besides producing more accurate estimates of the maps than in the original study, the method revealed that three parameters of map structure change with stripe-rearing, rather than just the single parameter of the proportion of neurons preferring the reared orientation. Together these three parameters provide a much richer description of both natural variability between maps from different individuals, and how maps change in response to altered environments.

2   Methods

2.1   The method of Macke et al. (2011)

The method introduced by Macke et al. (2011) for estimating orientation preference maps from optical imaging data is an application of Gaussian process (GP) regression (Rasmussen and Williams, 2006) to the problem of estimating a 2D vector field from noisy data. The important results are the choice of an appropriate prior, a method for estimating the parameters of the prior, an appropriate noise model and a method for fitting it to the data, and the use of approximation techniques for dealing with large datasets.

Briefly, the prior is defined by a zero mean function and a difference of Gaussians covariance function,

$$K(x, x', k, k') = \delta_{k,k'} \sum_{a,b=1}^{2} \frac{\alpha_a \alpha_b}{2\pi(\sigma_a^2 + \sigma_b^2)} \exp\left(-\frac{|x - x'|^2}{2(\sigma_a^2 + \sigma_b^2)}\right), \tag{1}$$

where $\delta_{i,j}$ is the Kronecker delta, x and x' are two positions in the map, and k and k' are either 1 or 2, the two vector components of the map. The assumption that the components are uncorrelated gives $K(x, x', 1, 2) \equiv K(x, x', 2, 1) \equiv 0$. The parameters of this function can be interpreted as scaling parameters, the α_i, and filter widths, the σ_i, which correspond to the wavelength of the map. Macke et al. set $\alpha := \alpha_1 = -\alpha_2$ so that K has zero mean, and $\sigma := \sigma_2 = 2\sigma_1$ to reduce the number of parameters.

The prior covariance matrix Σ_{prior}, a $n \times n$ matrix containing the covariance K between each pair of n pixels in a map, needs to be calculated, stored and inverted, which is a computationally demanding task. Macke et al. used a low-rank approximation to the prior, generated by an incomplete Cholesky decomposition. This results in a prior of the form $\Sigma_{\text{prior}} = D + GG^{\mathsf{T}}$, where D is $n \times n$ and diagonal, and G is $n \times q$, with $q \ll n$. The eigenspectrum of the chosen covariance function drops off quickly, which makes the incomplete Cholesky decomposition a good approximation. To estimate the parameters α and σ of the prior covariance function (equation 1), Macke et al. fit the covariance function to the average of the radial component of the autocorrelation function of the two vector components of the filtered, vector averaged map.

There are numerous sources of noise in images of the cortex, and therefore the most appropriate way to incorporate noise in the model is by adding a full noise covariance matrix Σ_ϵ to the prior, which models the independent noise variance at each pixel as well as the noise correlation between all pairs of pixels. However, for practical reasons in these types of experiments, there are usually many more pixels than trial images, meaning the entire matrix cannot be inferred from the data. Macke et al. overcame this problem by using a noise covariance of the form $\Sigma_\epsilon = D_\epsilon + G_\epsilon G_\epsilon^{\mathsf{T}}$, where D_ϵ is $n \times n$ and diagonal, and G_ϵ is of size $n \times q_\epsilon$, with $q_\epsilon \ll n$. D_ϵ models the independent noise in each pixel, while the low-rank term models noise correlations between a small number of pixels, determined by the chosen value of q_ϵ.

Macke et al. used an iterative approach to fit this noise model to the data. An initial estimate of the noise is obtained by calculating the variance at each pixel within each stimulus condition. This makes the assumption that the underlying activity in the cortex is identical for the presentation of an identical stimulus, so the difference between images within each stimulus condition can be attributed to noise. These estimates are then averaged over all the stimulus conditions, giving an initial estimate of the pixel-wise noise variance D_ϵ. The posterior distribution is derived using this estimate, and then the noise covariance is fit to the residuals of the mean of the posterior and the data, using a factor analysis method.

This process of deriving the posterior and fitting to the residuals is repeated several times, resulting in a good estimate of the noise covariance in the data. With the prior and noise models fit to the data as described above, the mean and covariance of the posterior distribution are derived using standard GP regression equations (Rasmussen and Williams, 2006). As both the prior and noise covariance matrices are in the form $A = D + GG^{\mathsf{T}}$ where D is diagonal, and G is of size $n \times q$, where $q \ll n$, they can be inverted using the matrix inversion lemma (Rasmussen and Williams, 2006), making calculation of the posterior distribution

computationally tractable.

2.2 Efficient inversion of a general prior covariance matrix

The most general form of prior covariance matrix for estimating OPMs using GP regression
is

$$\Sigma_{\text{prior}} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}, \tag{2}$$

where A and C are the covariance matrices of the real and imaginary components, respectively,
and B is the covariance between the two components. Macke et al. (2011) substantially
simplified the GP regression problem by assuming a prior of the form

$$\Sigma_{\text{prior}} = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix},$$

which assumes independence of the two vector components and that their covariance structure
is identical, which is a valid assumption for orientation preference maps from normally reared
animals. This simplification, combined with an incomplete Cholesky decomposition and the
matrix inversion lemma, allowed efficient calculation of the posterior distribution (see above).
To handle the more general case (equation 2), this method needs to be adjusted, but can be
performed in a very similar manner.

Firstly A, B and C are factored separately using incomplete Cholesky decompositions,
into the forms $A = D_A + G_A G_A^\mathsf{T}$, $C = D_C + G_C G_C^\mathsf{T}$ and $B = G_B G_B^\mathsf{T}$. The factorisation of B
does not have a diagonal component, as the diagonal of B corresponds to covariance between
map components, not variance, and so we do not give it special treatment, as in A and C.

Now set $G_1 = \begin{pmatrix} G_A \\ 0 \end{pmatrix}$, $G_2 = \begin{pmatrix} 0 \\ G_C \end{pmatrix}$, $G_3 = \begin{pmatrix} G_B \\ 0 \end{pmatrix}$, $H_3 = \begin{pmatrix} 0 & G_B^\mathsf{T} \end{pmatrix}$, $G_4 = \begin{pmatrix} 0 \\ G_B \end{pmatrix}$ and $H_4 = \begin{pmatrix} G_B^\mathsf{T} & 0 \end{pmatrix}$, where $\begin{pmatrix} A & 0 \end{pmatrix}$ indicates the block matrix formed by A and an appropriately sized zero
matrix. It is simple to check that we now have $\Sigma_{\text{prior}} = G_1 G_1^\mathsf{T} + G_2 G_2^\mathsf{T} + G_3 H_3 + G_4 H_4 + \begin{pmatrix} D_A & 0 \\ 0 & D_C \end{pmatrix}$. Setting

$$G = \begin{pmatrix} G_1 | G_2 | G_3 | G_4 \end{pmatrix}, \ H = \begin{pmatrix} G_1^\mathsf{T} | G_2^\mathsf{T} | H_3 | H_4 \end{pmatrix}^\mathsf{T}$$

and

$$D = \begin{pmatrix} D_A & 0 \\ 0 & D_C \end{pmatrix},$$

gives $\Sigma_{\text{prior}} = D + GH$, and so the matrix inversion lemma,

$$\Sigma_{\text{prior}}^{-1} = D^{-1} - D^{-1} G \left(\mathbb{I} + H D^{-1} G \right)^{-1} H D^{-1},$$

where $\mathbb{I}$ is an appropriately sized identity matrix, can be used to efficiently calculate the
inverse of Σ_{prior}.

Additionally, a more general prior also has a non-zero mean. If the mean at each pixel is
represented by the vector $\mathbf{m}$, then we simply replace the data vector $\mathbf{y}$ with $\mathbf{y} - \mathbf{m}$ in all of
the calculations (Rasmussen and Williams, 2006).

2.3 Marginal likelihood maximisation

Macke et al. (2011) chose the hyperparameters of the prior by fitting its covariance function to the auto-correlation function of a filtered, vector averaged map. In this Bayesian framework, however, the marginal likelihood is a quantity which gives the probability that a given set of data came from a given prior. We maximised this quantity over the parameters of the prior to obtain a set of parameters which define the optimal prior for each dataset. The log marginal likelihood $\log p(\mathbf{y}|\Sigma_{\text{prior}})$ can be calculated as

$$\log p = -\frac{1}{2}\left(\mathbf{y}^{\mathsf{T}}\Sigma_{\text{prior}}^{-1}\mathbf{y} + \log|\Sigma_{\text{prior}}| + n\log 2\pi\right),$$

where $\mathbf{y}$ is the data vector, Σ_{prior} is the prior covariance matrix and n is the total number of data points (Rasmussen and Williams, 2006).

To maximise this function, we tried several local optimisation algorithms which did not handle the relative shallow peaks in the function well. We ultimately used the SIMPSA algorithm (Cardoso et al., 1996), a combination of simplex methods and simulated annealing, as it was tested on synthetic data sampled from priors with known parameters and performed well (data not shown).

2.4 Comparison metric

The metric used to compare maps was the absolute value of the complex correlation coefficient of two maps represented as one-dimensional complex vectors.

2.5 Simple overrepresentation quantification

To aid comparison with the original analysis of Sengpiel et al. (1999), we calculated a simple scalar value for overrepresentation, ρ. This was defined by first binning each pixel in a vector averaged map by its nearest preferred orientation out of $S = \{0°, 45°, 90°, 135°\}$. The value ρ was then defined as

$$\rho = \frac{1}{N}\left(3n_{\theta'} - \sum_{\theta \in S\setminus\theta'} n_\theta\right),$$

where N is the total number of pixels, n_θ is the number of pixels preferring θ, and θ' is either the reared orientation or the maximally represented orientation in control maps. This should yield values near zero for maps with no overrepresentation, and positive values for maps with an overrepresentation of the reared orientation.

2.6 Covariance function differences

The difference between the covariance functions of the real and imaginary components of an orientation map can be used to infer overrepresentation. Differences in both the maximum value and the distance for which the function is a minimum (i.e., the argument of the minimum) quantify the degree of overrepresentation. To formalise this notion we used the absolute relative difference between the max and argmin of the two components, written Δmax and Δargmin, respectively. Formally, Δmax is defined as

$$\Delta\text{max} = \frac{|\text{max}_{\text{real}} - \text{max}_{\text{imaginary}}|}{\frac{1}{2}(\text{max}_{\text{real}} + \text{max}_{\text{imaginary}})},$$

and Δargmin as

$$\Delta\text{argmin} = \frac{\left|\text{argmin}_{\text{real}} - \text{argmin}_{\text{imaginary}}\right|}{\frac{1}{2}(\text{argmin}_{\text{real}} + \text{argmin}_{\text{imaginary}})}.$$

2.7 Data

We used three different datasets for the analyses in this work, all consisting of optical imaging of either cat or ferret primary visual cortex: The images of ferret V1 used in Macke et al. (2011), images of a single cat V1 collected by us for use in this method development (see below for full experimental methods), and the images of cat V1 used in Sengpiel et al. (1999) (summarised in table 1), which contains full details of the experimental procedures.

For the ferret data we used region 2, as defined in Macke et al. (2011), as it contained minimal noise and was also used in that paper for the majority of the analyses. The data from Sengpiel et al. (1999) were only available in an averaged form, as due to the expense of digital storage space at the time of the experiments, not all of the data were stored permanently. Instead of individual trial images from the experiment, the data were only kept in 'block' form, the average of 8 presentations of each stimulus, divided by a blank. This resulted in, for each image set, a set of between 6 and 14 images for each stimulus condition. A schematic showing the form of the block data is given in figure 1. 13 different image sets in this form were available for 17 hemispheres of 10 cats (table 1).

ID	Cat Name	Hemisphere(s)	Treatment	Blocks/Stimulus
C1	C29_97	L	135° Stripe	10
C2	C45_97	L&R	0° Stripe	8
C3	C28_97	L	90° Stripe	6
C4		R	90° Stripe	8
C5	C16_97	L	90° Stripe	12
C6		R	90° Stripe	12
C7	C69_97	R	45° Stripe	14
C8	C27_97	L	Normal	10
C9		R	Normal	12
C10	N032207	R	Normal	14
C11	N033006	L&R	Normal	12
C12	N013103	L&R	Normal	12
C13	C842	L&R	Normal	6
Macke et al. (2011)	–	–	Normal	100
Our Data	–	–	Normal	80

Table 1: **Summary of the data used.** The hemisphere(s) column lists which hemisphere(s) were contained in each set of images. The blocks/stimulus column lists how many block images per stimulus condition were available, or images/stimulus for both our and Macke et al. (2011)'s datasets.
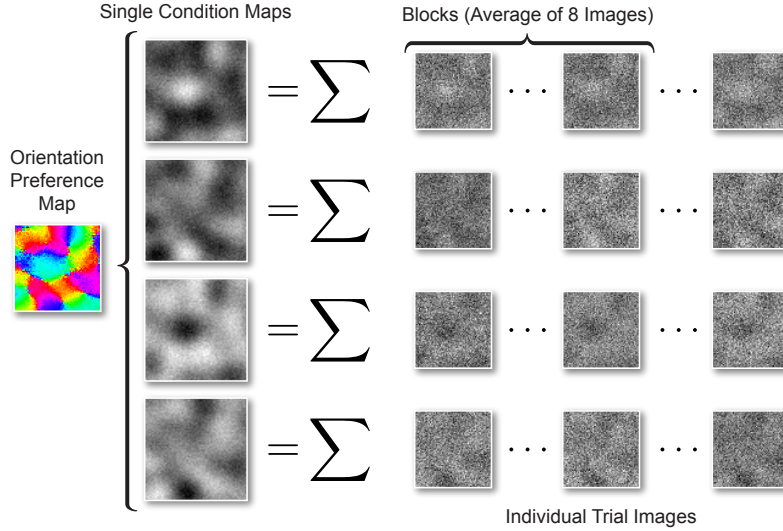
Figure 1: Schematic showing the form of the available data. Individual trial images were available for our data and the data fromMacke et al. (2011), but only block data (the average of sets of 8 images) were available for data from Sengpiel et al. (1999).

2.7.1 Synthetic maps

For the purposes of verifying the implementation of some techniques and assessing their validity, we used synthetic maps. These were generated by sampling from a prior distribution of maps, generated with specified hyperparameters, then converting the maps into the appropriate form (such as individual trial images or averaged block data), and adding noise. Noise was added to each pixel in each image individually, all drawn from the same Gaussian distribution, with a zero mean and specified variance. The process of sampling from the prior is mathematically identical to convolving white noise with the prior covariance function to generate the two vector components of a map.

The training data used to train the linear predictor consisted of 20000 maps sampled from the stripe-reared prior in the way described above, for a range of the parameters α_{real}, σ_{real}, $\alpha_{\text{imaginary}}$, $\sigma_{\text{imaginary}}$ and r, to overrepresent one of 0°, 45°, 90° or 135° to varying degrees. The values of Δmax and Δargmin for each map were then determined from the covariance functions formed by the chosen hyperparameters. The degree of overrepresentation in these maps was quantified using the measure ρ.

2.7.2 Elastic net maps

To determine general rules for the covariance functions of stripe-reared maps, maps with overrepresented orientations were generated using the elastic net model. This model has been shown to produce overrepresentation of the reared orientation (Carreira-Perpiñán et al., 2005), and therefore was used as a guide to the general structure of maps with overrepresentation caused by single orientation rearing.

The elastic net algorithm is a dimension-reducing model for determining how preferences for features of the input space (e.g., visual field position, orientation, spatial frequency) are arranged in the cortex. It folds a two-dimensional net in higher-dimensional space in an

attempt to minimise a tradeoff between coverage of the input space and continuity of the cortical representation. To increase efficiency, we only included orientation preference in our simulations.

The parameters and training regime used were as described in Carreira-Perpiñán et al. (2005), with the exception that the stimulus space consisted solely of orientation preference and selectivity. Twelve maps were generated with the orientations $0°$, $15°$, ..., $165°$ overrepresented, respectively. In each case, the overrepresented orientation was boosted to a level of 2.5, meaning that that orientation wielded 2.5 times as much influence on the elastic net as the others. This resulted in maps with a relatively large degree of overrepresentation in comparison to the real maps of Sengpiel et al. (1999), which was done to make identifying systematic changes in the maps' covariance functions easier. The simulations were performed using custom software written in MATLAB, the same as used by Carreira-Perpiñán et al. (2005), a version of which is available online at `http://faculty.ucmerced.edu/mcarreira-perpinan/research/EN.html`.

2.8   Data preprocessing

In most of the data used here the imaging covered areas outside of the primary visual cortex, and it is important to ensure that these areas do not interfere with the analyses. We therefore chose to mask the images so that only pixels inside V1 were considered. Anatomical features can be used to identify V1 with a high level of accuracy, and we used this approach to mask all of the image sets used here, except for the Macke dataset, as the chosen region contained only a small number of non-V1 cells. In images containing two hemispheres (see table 1), the two hemispheres were masked and considered separately. Additionally, the vector averaged maps which were analysed were low- and high-pass filtered to remove noise, and for some analyses the angle value at each pixel was weighted by the tuning strength, in a further effort to reduce the influence of non-orientation specific cells. We investigated the effects of weighting and masking, and decided the preceding protocol to be the best choice.

For consistency, all datasets were normalised to have a zero mean and unit variance, calculated across all pixels in each collection of images. Performing this across the entire datasets, as opposed to per image as done by Macke et al. (2011), does not affect the value distributions of the maps, which is important when analysing maps with possible overrepresentation. To account for differences in overall response between stimulus conditions, we found the transformation required to set the area outside of V1 in each stimulus response to have a zero mean and unit variance, and applied that transformation to the area within V1. This method makes the assumption that the areas outside V1 contain Gaussian noise.

In our experiment, images were captured at a rate of 5 Hz for a period of 10 s, resulting in 50 frames for each trial (see Experimental methods). To obtain a single image from each trial for analysis, the mean of the 10 pre-stimulus frames was subtracted from the 25 stimulus frames, which had the effect of removing the baseline activity from the images. Then the trial image was taken as the mean of the final 5 stimulus frames. As there are small changes in the images due to respiration and blood-flow, taking the mean of a smaller number of frames is a better approach than taking the average of all 25 stimulus frames. Mathematically stated, if r_i is frame i of a single trial, where i ranges from 1 to 50, the image used as a representation

of that trial was

$$r = \frac{1}{5} \sum_{i=31}^{35} r_i - \frac{1}{10} \sum_{i=1}^{10} r_i.$$

2.9 Experimental methods

We performed optical intrinsic signal imaging of primary visual cortex in three normally sighted juvenile cats (*felis catus*; 3.0–3.8 kg) reared from birth in outdoor enclosures subject to the prevailing night-day cycle. All surgical and experimental procedures were approved by the institutional Animal Experimentation Ethics Committee (AEEC) at the Australian National University and were performed in strict compliance with the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes from the Australian National Health and Medical Research Council (NH&MRC).

2.9.1 Anaesthesia and surgical procedures

Animals were prepared for acute physiological recordings as described previously (van Kleef et al., 2010). Animals were anaesthetised by intramuscular injection of ketamine hydrochloride (20 mg/kg) and xylazine (1 mg/kg). Once deeply anaesthetised, as confirmed by the absence of corneal and toe withdrawal reflexes, animals were placed on a heated surgery table and the cephalic vein was cannulated to permit delivery of fluids and intravenous drugs. Animals were then intubated to ensure adequate respiration and anaesthesia was then maintained for the duration of the experiment by inhalation of gaseous halothane (1.0% during surgery; 0.5–1.0% during recording) in a 60:40 mixture of N_2O and O_2. Non-invasive physiological indicators, including heart rate and end-tidal CO_2 concentration, were continuously monitored to ensure adequate levels of anaesthesia at all times. Body temperature was maintained at 37.7°C by means of an electric blanket under feedback control.

The head was held in a stereotaxic frame using ear bars, a bite bar and a head bolt positioned on the skull at the midline approximately 30 mm anterior to inter-aural zero. To prevent eye movements, animals were subject to neuro-muscular blockade by continuous intravenous infusion of gallamine triethiodide (10 mg/kg/h). During neuro-muscular blockade animals were mechanically ventilated to maintain end-tidal CO_2 between 3.5 and 4%. For fluid replacement, all animals received a constant intravenous infusion containing Hartmanns (lactated ringer) solution (25% by volume), 5% glucose 0.9% NaCl solution (25% by volume) and an amino acid solution (50% by volume) at a rate of (2.5 mL/kg/h). Animals were given daily injections to reduce salivation (atropine, 0.2 mg/kg; s.c.), cerebral oedema (dexamethasone phosphate, 1.5 mg/kg; i.m.) and the risk of infection (Clavulox, a broad spectrum antibiotic, 0.05 mL/kg; i.m.). The eyes were fitted with neutral power rigid gas-permeable contact lenses to ensure corneal perfusion, and drops (1% atropine; 10% phenylephrine) were administered daily to cause pupillary dilatation and to retract the nictitating membrane. Refractive errors were assessed by reverse ophthalmoscopy and corrected as required using spherical lenses placed in front of the eyes to focus the stimulus on the retina. For imaging, the scalp was reflected and a small trephine hole (10 mm × 8 mm; A4 to P6, L1 to L9) was made in the cranium, exposing the visual cortex (Areas 17 and 18) in the right hemisphere. A stainless steel recording chamber was affixed to the cranium with dental acrylic and the dura

mater removed. The recording chamber was then filled with silicone oil (Dow Corning 200, 50 cSt) and sealed with a glass cover slip, mechanically stabilising the cortex for imaging.

At the conclusion of the experiment animals were euthanised by intravenous injection of an overdose of barbiturate (sodium pentobarbital; Lethabarb, 150 mg/kg).

2.9.2   Intrinsic signal imaging

The exposed cortex was imaged using a Pantera 1M60P high-sensitivity 12-bit area scan CCD camera (Teledyne DALSA, Waterloo, ON Canada) fitted with a macroscope, formed using a pair of Nikkor 50 mm f/1.2 lenses arranged front-to-front (H and A, 1991), focused 400–600 μm below the cortical surface. The camera was configured to bin sensor pixels 2×2 producing images with a resolution of 512×512 pixels (24×24 μm per pixel). Image acquisition was restricted to a region of interest (ROI) comprising a subset of the full imaging frame.

During imaging, the cortex was epi-illuminated using a custom built LED light source with a peak wavelength of 520 nm (Agilent Technologies; HSMQ-C150). Increased blood flow indicative of neural activity was therefore manifest as a reduction in cortical reflectance and thereby image intensity.

Cortical responses were imaged during presentation of an ensemble of visual stimuli (see below). For each stimulus presentation images were acquired continuously at a rate of 5 Hz for a period of 10 s, the onset of which was synchronised to the phase (maximum inspiration) of the respirator. Visual stimuli were presented for 5 s beginning 2 s after the beginning of image acquisition. Each stimulus presentation was followed by a 3 s (minimum) recovery period during which the monitor displayed an isoluminant mean gray screen.

2.9.3   Visual stimuli

Visual stimuli consisted of luminance defined oriented square wave gratings presented within a circular aperture on an isoluminant gray screen matched to the mean luminance of the gratings. Stimuli were generated by a ViSaGe visual stimulus generator (Cambridge Research Systems Ltd., Cambridge, UK) and presented on a calibrated Clinton Monoray CRT monitor (modified Richardson Electronics MR2000HB-MED CRT with fast DP104 phosphor, 100 Hz refresh, 1024×768 pixels) viewed binocularly from a distance of 28 cm. Maps of orientation preference were obtained by presenting high contrast (Michelson contrast, c = 1.0), large field (60°) drifting gratings of fixed spatial and temporal frequency (0.15 cycles per degree and 2 Hz respectively) drifting in one of either 16 or 32 directions equi-spaced between 0° and 360°. Each stimulus direction, together with a blank condition (no grating), was presented at least 30 times with the order of presentation randomised across trials to avoid any systematic bias in the observed responses.

3   Results

3.1   The quality of averaged data and optimal parameters for data acquisition

We developed a generalised form of the Gaussian process map estimation method of Macke et al. (2011), and applied it to three different sets of data: our data, data from Macke

et al. (2011), and data from Sengpiel et al. (1999) (see Methods). A feature of the Sengpiel et al. data was that the image sets were only available as block images, averages of sets of 8 images, rather than the individual experimental trial images. It was therefore important to ensure that the use of the averaged data would not significantly reduce the quality of the resulting map and therefore affect the results, when compared to a map generated using all individual experimental images. Our own dataset consisted of a large number of trials for many orientations, and could therefore be used to investigate this characteristic of the data.

Experimental data from the one cat we imaged, consisting of each individual trial image from 80 trials of 16 different orientation stimuli (evenly spaced in $[0°, 180°)$), was used to determine the difference in map quality induced by using averaged trials. Maps were estimated from subsets of this dataset using the original Gaussian process method as described by Macke et al. (2011). As there are many ways, for example, 80 trials can be averaged together to form 10 blocks, 40 random samples of each number of averaged trials were used to estimate maps, to obtain a distribution of map quality values. The results indicate that using averaged data does not make a substantial difference to the maps (figure 2**A**), at least in terms of overall map quality, and so the use of such data in further analyses is justified.

In addition, it is useful to investigate how best to design orientation preference map imaging experiments. In particular, given a limited number of trials, what is the trade-off between obtaining many images for just a few orientations, and thus getting a good estimate of the noise for only a sparse sampling of orientation space, versus obtaining only a few images for each of many orientations, giving a poor estimate of noise but for a dense sampling of orientation space?

We addressed this question by generating maps from sets of constant numbers of total images, split among a differing number of stimulus conditions (figure 2**B**). Again, as a subset of images can be obtained in many ways, 40 random samples of each condition were used. The results indicate that the difference between maps resulting from different numbers of stimuli is not substantial, except at very small numbers of trials. For very small numbers of images, the difference between having, for example, 1 trial for 16 stimuli, and 4 trials for 4 stimuli becomes substantial, the latter option being the better choice as an estimate of the pixel-wise noise can be obtained. Overall then, for any sufficiently large number of trials (greater than approximately 50), the choice of trial dispersal among stimuli is of little consequence to the quality of the resulting map.
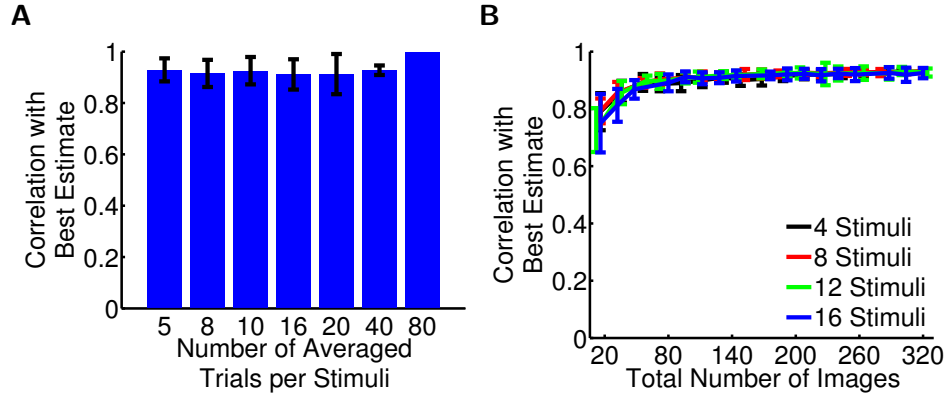
Figure 2: Using averaged data does not reduce map quality substantially, nor does the dispersal of experimental trials among stimuli. (**A**) A comparison between map quality (measured as the correlation with the map generated using all of the data, 80 trials each of 16 stimuli) and amount of data used, here as averaged blocks of trials, showing a drop in quality when any averaging is used, but with all maps being above 85% similar to the map estimated using all of the data. Data shown is mean ± std. dev. for 40 random samples. (**B**) A comparison between map quality and the total number of images used in the estimation, for different numbers of stimulus conditions. In each case, the total number of images are split evenly between each of the stimuli, where the mean ± std. dev. is shown for 40 random subsets of images.

3.2   Robust estimation of hyperparameters

Macke et al. (2011) estimated the parameters of the prior distribution by fitting the prior covariance function to the autocorrelation functions of the real and imaginary components of filtered, vector averaged maps (see Methods). For very noisy data, the autocorrelation function can also be very noisy, in which case filtering of the map is required. However, introducing this subjective step causes the process to become much less reliable. We therefore implemented a computationally expensive but more robust method for determining the hyperparameters, which involves choosing the parameter set which maximises the marginal likelihood (Rasmussen and Williams, 2006) (see Methods). This results in an inferred set of parameters which determine the prior from which the data is most likely to have come.

To validate this technique, we compared its chosen parameters to the true parameters of synthetic data, with which we find a good match (figure 3**A**). We also compared it to the autocorrelation function fit technique of Macke et al. (2011), and showed that on synthetic data (figure 3**A**), the data from Macke et al. (2011) (figure 3**B**), our data (figure 3**C**) and a representative control dataset from Sengpiel et al. (1999) (figure 3**D**), the two methods produce similar parameter estimates, but as shown in figure 3**A**, the marginal likelihood technique produces estimates closer to the true parameter values.

The peaks of the marginal likelihood (figure 3) are relatively flat, indicating that a large range of parameter values fit the data well. In most orientation preference experiments, a large number of images are captured relative to the number of data points one might have in a more typical regression problem. The large amount of data suppresses the influence of the prior, and thus the importance of accurately estimating the parameters of the prior when estimating a map diminishes. To confirm this, the correlation with the map estimated with the hyperparameters corresponding to the peak marginal likelihood was also calculated (figure 3), which has a similarly flat peak in all cases. This does not diminish the importance of this more robust technique, however, as accurate estimates are important when using the fitted parameters to analyse the map (see below).
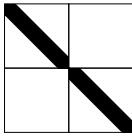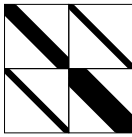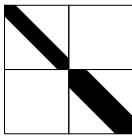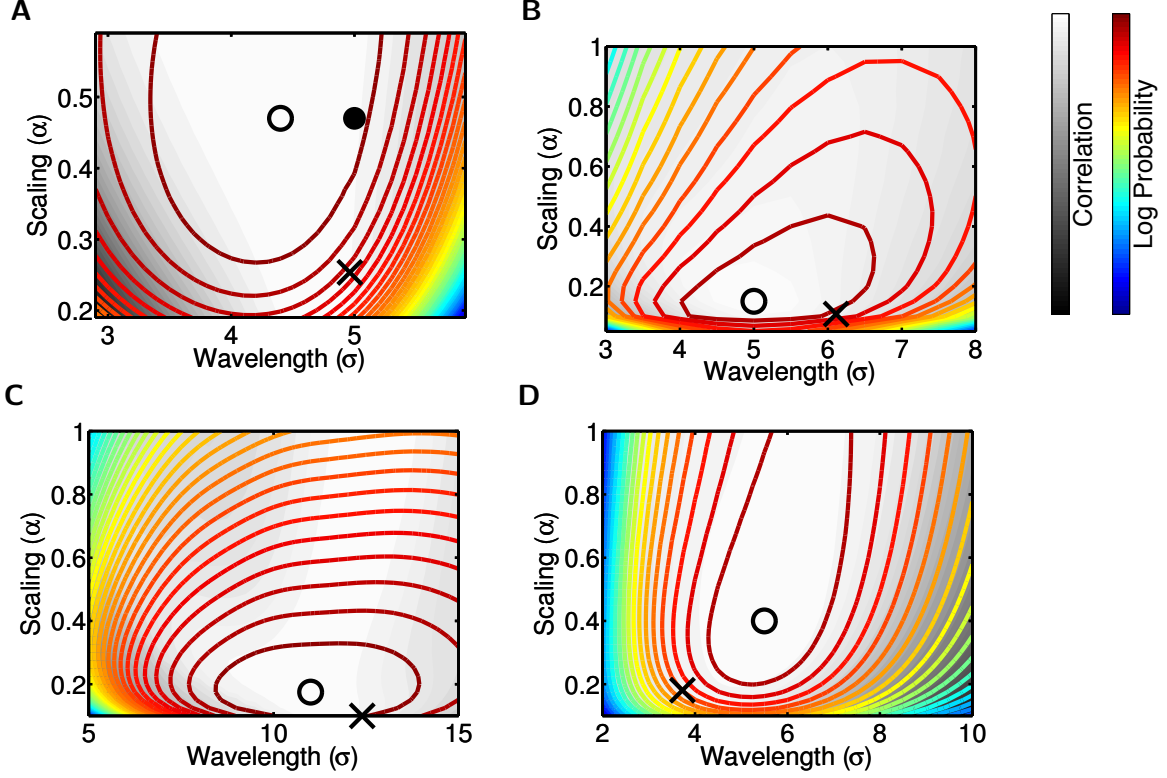
Figure 3: Hyperparameters can be reliably inferred from the data by maximising the marginal likelihood. All panels show isolines of the marginal likelihood for a range of the two hyperparameters of Macke et al. (2011)'s prior (colour scale), above a filled contour plot of the correlation between maps estimated with the specified parameters and the parameters corresponding to the peak marginal likelihood (grey scale). The peak of the marginal likelihood (empty circle) and the parameter chosen by Macke et al. (2011)'s technique (cross) are also shown. (**A**) Using synthetic test data. Here we know the true parameters of the map (filled circle) with which the maximum of the marginal likelihood function corresponds well. (**B**) Using the data from Macke et al. (2011). (**C**) Using our data. (**D**) Using a representative example of a control dataset from Sengpiel et al. (1999). Generally the parameters chosen by Macke et al. (2011)'s method agree with those fitted by maximising the marginal likelihood, with the maps resulting from each estimation being fairly similar (as shown by the greyscale correlation values). However, the latter is closer to the true parameter values, as shown in **A**.

3.3 Generalising the GP model

To reduce the complexity of the model, Macke et al. (2011) assumed that the real and imaginary components of the maps were independent and uncorrelated, and also that the a priori mean at each pixel was zero (shown schematically in figure 4**A**, first row). These assumptions hold for normally reared animals (dotted lines, figures 4**C** and **D**). However, in the general case, there may be correlations between the real and imaginary components, differences in their individual covariance functions, and a non-zero mean at each pixel (figure 4**A**, second row).

To investigate this empirically, we generated maps using the elastic net model with and without overrepresentation of a larger variety of orientations than available in experimental data (see Methods). The covariance functions of the overrepresented maps showed a systematic difference between the real and imaginary components of the maps, dependent on the overrepresented orientation, as well as a non-zero covariance between the map components (figure 4, panels **B**–**D**, data from two maps for each orientation, with and without overrepresentation; other maps showed similar results (data not shown)). These changes in the covariance structure lead to overrepresentation by shifting the angle distribution in a map with a uniform distribution towards the reared orientation. It is clear, then, that a more general treatment is required when investigating maps with possible overrepresentation.
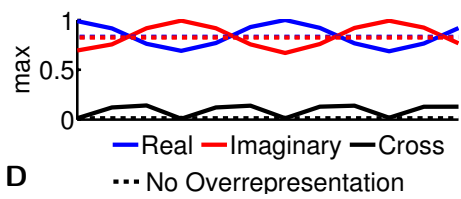
A

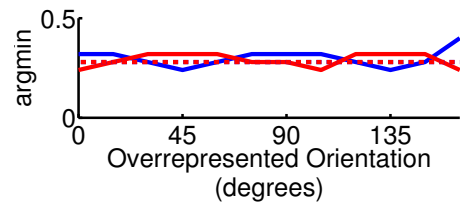|	Prior Covariance		Prior Mean
Macke et al.		$\begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \end{pmatrix}$
General		$\begin{pmatrix} A & B \\ B & C \end{pmatrix}$	$\begin{pmatrix} \mu_{\mathrm{real}}(x,y) \\ \mu_{\mathrm{imaginary}}(x,y) \end{pmatrix}$
Stripe Reared		$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$	$\begin{pmatrix} r\cos 2\theta \\ r\sin 2\theta \end{pmatrix}$

B
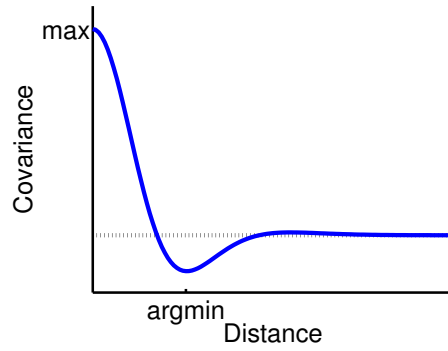
C

D

Figure 4

Figure 4: Schematic representation of the structure of the prior covariance and mean, demonstrating the different levels of generality. (**A**) The first row of the table shows the model used by Macke et al. (2011), with zero mean and matched and independent real and imaginary component covariances. The second row shows the most general model, with a non-zero, independent mean which varies with position (x, y) across the map, and independent real and imaginary covariances with correlations. The final row shows a compromise between the first two, with a non-zero mean parameterised by the overrepresented angle θ and degree r and a general covariance without correlations. (**B**) Definitions of the terms max and argmin, in relation to the covariance function of maps. (**C**) An analysis of the covariance structure of elastic net maps (data from two maps for each orientation, with and without overrepresentation; other maps showed similar results (data not shown)), showing the maximum value of the autocovariance functions of the real and imaginary components, and the maximum absolute value of the covariance between the two components, showing a dependence on the overrepresented orientation. (**D**) The distance at which the autocovariance is at a minimum for the same maps, again showing a dependence on the overrepresented orientation. This example inspired the generalised prior (panel **A**, second row). Note that the chosen overrepresented orientations are only relative, dependent upon how angles in the plane are measured, and that choosing different coordinate axes would not change the relationship between the red and blue curves, but simply shift them together.

3.4 The analysis of stripe-reared maps

3.4.1 Classical analysis

The classical method for determining overrepresentation in orientation preference maps is to bin each pixel according to which orientation of a small set (usually 0°, 45°, 90° and 135°) is nearest its preferred orientation. The difference between the number of pixels in each bin is interpreted as the level of overrepresentation in the map. This method only allows a granular analysis of the overrepresentation on a small set of predefined orientations however, and artefacts in the images may lead to large areas which are erroneously said to prefer some orientation. Taking the difference between the number of pixels preferring the reared or maximally oriented orientation and the other orientations as a scalar measure ρ (see Methods), there was significant overrepresentation of the reared orientation in the stripe-reared maps in the Sengpiel et al. (1999) dataset compared to the control animals (figure 5**A**; $p < 0.0005$, Mann-Whitney U test), confirming the results of Sengpiel et al. (1999). It is notable that this measure is not very robust to changes in the filtering parameters used to smooth the data (figure 5**B**).

3.4.2 Shifting value distributions

In an orientation preference map with overrepresentation, the value distributions of the two vector components of the map are expected, from our investigations of the elastic net model, to shift according to the degree and orientation of overrepresentation. We first investigated if this change alone would account for the overrepresentation seen in the maps. To confirm this, we analysed the value distributions of the vector averaged maps from the Sengpiel et al. (1999) dataset, and found that they do vary with rearing condition (figures 5**C** and **D**). In control animals, the two map components had near identical distributions (figure 5**C**). However, in stripe-reared maps the mean of the component corresponding to the reared orientation saw a positive shift (figure 5**D**), with the corresponding components being 0° $\leftrightarrow$ positive real, 45°t $\leftrightarrow$ positive imaginary, 90° $\leftrightarrow$ negative real, and 135° $\leftrightarrow$ negative imaginary. We assume that this pattern holds for intermediate orientations, and that the means of the value distributions follow the equations

$$\mu_{\text{real}} = r \cos 2\theta$$

$$\mu_{\text{imaginary}} = r \sin 2\theta,$$

where r and θ are the degree and orientation of overrepresentation, respectively.

To check that this is an appropriate parametrisation of the mean, we least-squares fit the means of the vector averaged map component distributions (figures 5**C** and **D**) to find a value for r, the degree of overrepresentation for each map (figure 5**E**). As expected, the mean value of r for the stripe-reared maps is significantly greater than that of the normally reared maps ($p < 0.018$, Mann-Whitney U test). Fitting the values of r and θ to the value distributions in this way provides an improved technique for estimating overrepresentation than that used in Sengpiel et al. (1999), as here we are not limited to considering only a discrete set of orientations. The distribution of values of this measure are in good agreement with the results of the original study (figure 5**A**).

Adding these two parameters, r and θ, as hyperparameters to the Gaussian process model allows us to infer them from the data by maximising the marginal likelihood (see Methods).

Performing this, with a prior otherwise the same as Macke et al.'s, provides surprising results (figure 5**F**). While the stripe-reared maps generally have larger r values than the control maps, there is a substantial degree of overlap between the two distributions. The inferred values of θ were, however, in good agreement with expectations, with all orientations being within approximately $25°$ of the reared orientation, with the exception of a single map, C4, which overrepresents the orthogonal orientation (figure 5**G**). Further investigation indicated that the region outside V1 in that particular map had a large area of constantly strong response for the reared orientation, violating our normalisation assumption of (potentially biased) Gaussian noise outside V1 (see Methods). No other method of normalisation could be found which could cope with this issue. It is worth noting that the measure ρ only indicates an overrepresentation of the reared orientation when the map is filtered appropriately. An unfiltered, vector averaged map of the C4 data has a negative ρ value, indicating that the orthogonal orientation is overrepresented. Map C3 contains data from the other hemisphere of the same cat, and overrepresents the reared orientation, suggesting that the orthogonal overrepresentation in the C4 data is erroneous and an artefact of the imaging and processing processes.

3.4.3 Shifting covariance functions

Another possible source of the overrepresentation not accounted for by the shift in value distribution means is the covariance functions. The elastic net maps predict small changes in the maximum and minimum of the argument of the covariance functions of maps with overrepresentation, and we looked for this effect in the vector averaged maps from Sengpiel et al. (1999). The covariance functions were the same for each map component in the control maps (figure 5**H**), as expected, but showed a slight difference in both magnitude and wavelength dependent on the reared orientation in the stripe-reared cases, as predicted by the elastic net maps (figures 5**I** and **J**). There was no substantial correlation between the two map components, which again is consistent with the elastic net maps, since for axis-aligned orientations (i.e., $0°$, $45°$, $90°$ and $135°$, the only orientations examined in Sengpiel et al. (1999)), the predicted correlation between components is zero (figure 4**C**).

To model these changes in the covariance functions, we added two more hyperparameters to the model, for a total of four parameters for the covariance: α_{real}, σ_{real}, $\alpha_{\text{imaginary}}$ and $\sigma_{\text{imaginary}}$. Following the results from the elastic net maps (figures 4**C** and **D**), we could then determine the level of overrepresentation by looking at the relative absolute difference between the max and argmin of the two components, Δmax and Δargmin (see Methods).

Inferring all six hyperparameters of the model from the data by maximising the marginal likelihood, and calculating the measures Δmax and Δargmin, provided a clearer picture of the differences between control and stripe-reared maps (figures 6**A** and **B**). Control maps showed generally small values of Δmax and Δargmin, and stripe-reared maps generally larger values, but interestingly this was not the case for all maps. However, all stripe-reared maps had a high value of at least one of the measures (r, Δmax and Δargmin), motivating us to investigate whether some combination of these measure may give the best indication of overrepresentation.

3.4.4 Prediction and Classification

We therefore combined the values r, Δmax and Δargmin linearly to obtain a single measure of overrepresentation. To determine the appropriate weighting coefficients, we generated a large number of maps with overrepresentation, by sampling from the appropriate priors, with a range of values of r, Δmax and Δargmin (see Methods). Calculating ρ for these maps, and then solving the linear system

$$a\,\Delta\text{max} + b\,\Delta\text{argmin} + c\,r + d = \rho$$

for the coefficients in the least-squares sense, we obtained $a = 0.53$, $b = 0.10$, $c = 0.32$, and $d = 0.04$. The coefficient of determination of the fit was $R^2 = 0.98$, validating our assumption of a linear relationship. Note that the values of the coefficients are not directly comparable, as they are of different units. We used this model to predict overrepresentation in the maps, which resulted in a significant grouping of the maps by rearing condition, as expected (figure 6**C**; $p < 0.012$, Mann-Whitney U test). These results can be compared to those from Sengpiel et al. (1999) (figure 5**A**), and to the least-squares fitting of r (figure 5**E**), with both of which there is a good match.

The fitted hyperparameters can also be used to classify maps by rearing condition without relying on the measure ρ, by training a linear support vector machine (SVM) on the Sengpiel dataset, cross-validating by leaving out a single map on each training iteration. Such an SVM could correctly classify 14/17 of the maps (2 stripe-reared and 1 control map misclassified), whereas an SVM trained on the values of ρ could only correctly classify 13/17 maps (4 stripe-reared maps misclassified). No standard non-linear kernel function performed better in either case (data not shown).

To investigate the space of hyperparameters, maps were sampled from priors spanning a realistic range of the parameters and with several overrepresented orientations. The value of ρ for these maps increases linearly, at different rates, for each of the hyperparameters, as predicted by the linear relationship investigated above (figure 6**D**). To investigate why some of the maps lie on different regions of the same iso-ρ planes in parameter space, we calculated the 2D distributions of orientation and tuning strength for all of the maps, and compared their similarity across iso-ρ planes. Only minor differences in the distributions were found, with changes in the size and shape of the peaks of orientation selectivity at the reared orientations (data not shown), which may be capturing subtle variation in the visual statistics encoded by each map.

These results lead to a model where all three metrics contribute to the overrepresentation seen in stripe-reared maps, but in slightly different ways. This can be seen visually in both the maps themselves and in the value distributions of the real and imaginary components (figure 6**E**). An increase in r shifts the distributions without affecting their shape, while changing Δargmin or Δmax has the opposite effect. We can get an overall estimate of the level of overrepresentation through our linear model, but by looking at the different components in this way we can see how the aspects combine to form the final level of overrepresentation. By looking at the distributions of orientation and selectivity in the maps we see that different combinations of the hyperparameters create maps with subtle differences in the distributions, as well as spatial structure.
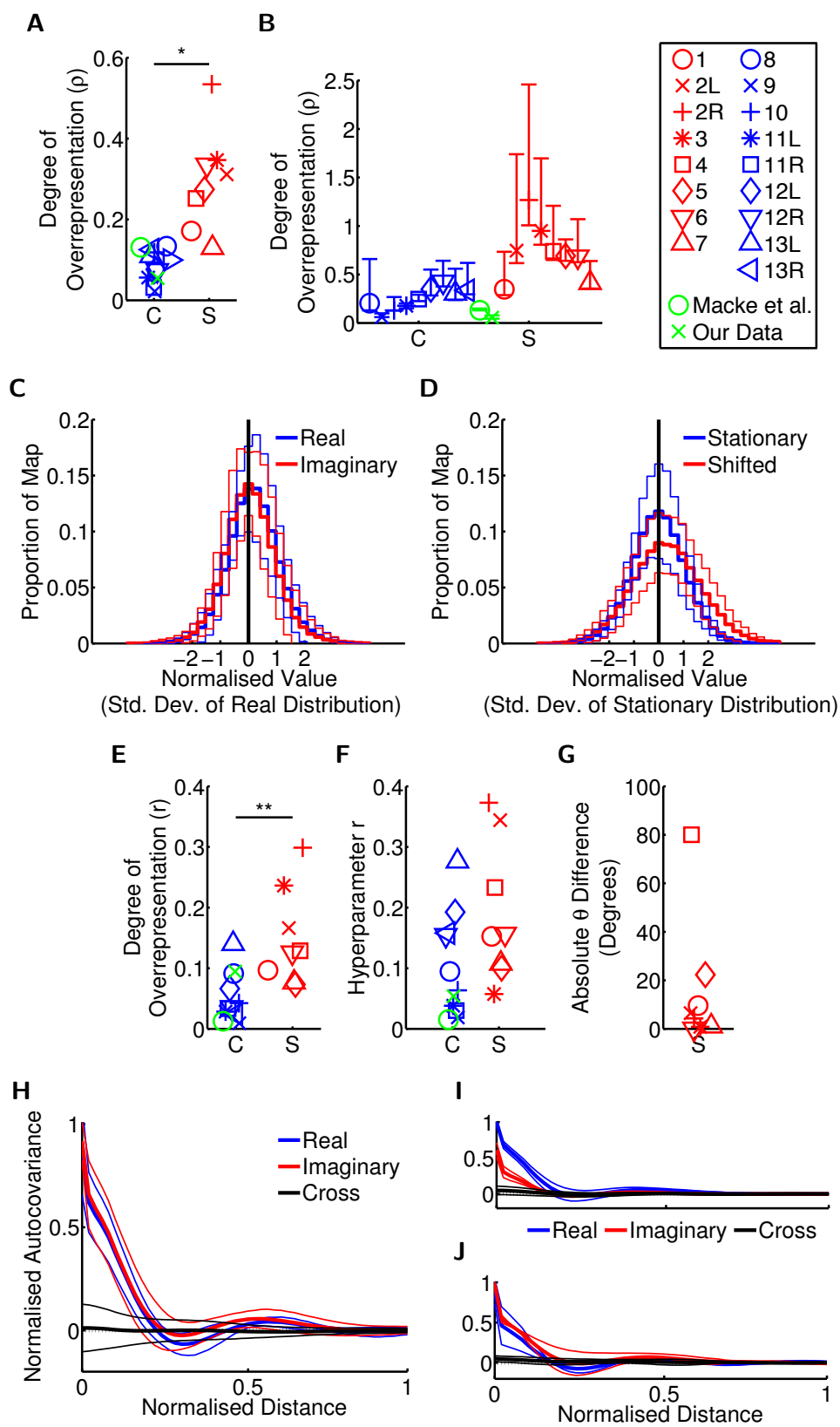
Figure 5

Figure 5: Stripe-rearing affects the value distributions of maps and their covariance functions. (**A**) Values of the overrepresentation metric ρ calculated on the data from Sengpiel et al. (1999), based on counting the number of pixels preferring each orientation (see Methods) (*$p < 0.0005$, Mann-Whitney U test). (**B**) The same data as panel **A**, but with error bars showing the range of values ρ takes over a range of realisitc filtering parameters ($\sigma_{\mathrm{lowpass}} \in \{1, 2, \ldots, 5\}$ pixels and $\sigma_{\mathrm{highpass}} \in \{15, 16, \ldots, 25\}$ pixels), showing that it is not very robust. (**C**) The normalised value distributions of the real and imaginary components of all normally reared maps ($n = 5$ cats, 10 hemispheres), with the thick and thin lines showing mean $\pm$ std. dev. The 95% confidence interval on $\mu_{\mathrm{imaginary}} - \mu_{\mathrm{real}}$ is $[-0.024, -0.021]$ (Student's t-test), showing a small difference in the means. (**D**) The normalised value distributions of all stripe-reared maps, with the components grouped so that a shift towards the reared orientation corresponds to a positive shift in the red distribution. The 95% confidence interval on $\mu_{\mathrm{shifted}} - \mu_{\mathrm{stationary}}$ is $[0.050, 0.054]$ (Student's t-test), over twice as large as the difference in the normal reared maps. (**E**) Values of r from least-squares fitting the mean parametrisation to the vector averaged maps' value distributions, showing a generally larger value of r for the stripe-reared maps than the normal maps (**$p < 0.018$, Mann-Whitney U test). (**F**) Marginal likelihood fitted values of r. (**G**) The absolute difference between the reared orientation and the fitted value of θ, showing that the most overrepresented orientations are close to the reared orientations, with the exception of a single map (see text). (**H**) The normalised autocovariance functions of the real and imaginary components, and the covariance between them, of all normally reared Sengpiel maps, showing mean $\pm$ std. dev. as in panel **C**. (**I–J**) The normalised covariance functions of (**I**) all $0°$ and $90°$ stripe-reared maps ($n = 3$ cats, 6 hemispheres) and (**J**) all $45°$ and $135°$ stripe-reared maps ($n = 2$ cats, 2 hemispheres). In the $0°$ and $90°$ cases there is a slightly smaller wavelength and a lower level of covariance at small distances in the imaginary component as predicted by the elastic net maps, and in the $45°$ and $135°$ cases the same predicted change in the real component is less clear but present. In both cases there is no substantial covariance between the two components. Note that, as in figure 4, these four orientations are not special, but rather are the result of our choice of coordinate axes, which is the most convenient for the selection of reared orientations available. Symbols show which dataset each point represents, as per the legend. C: control, S: stripe-reared.
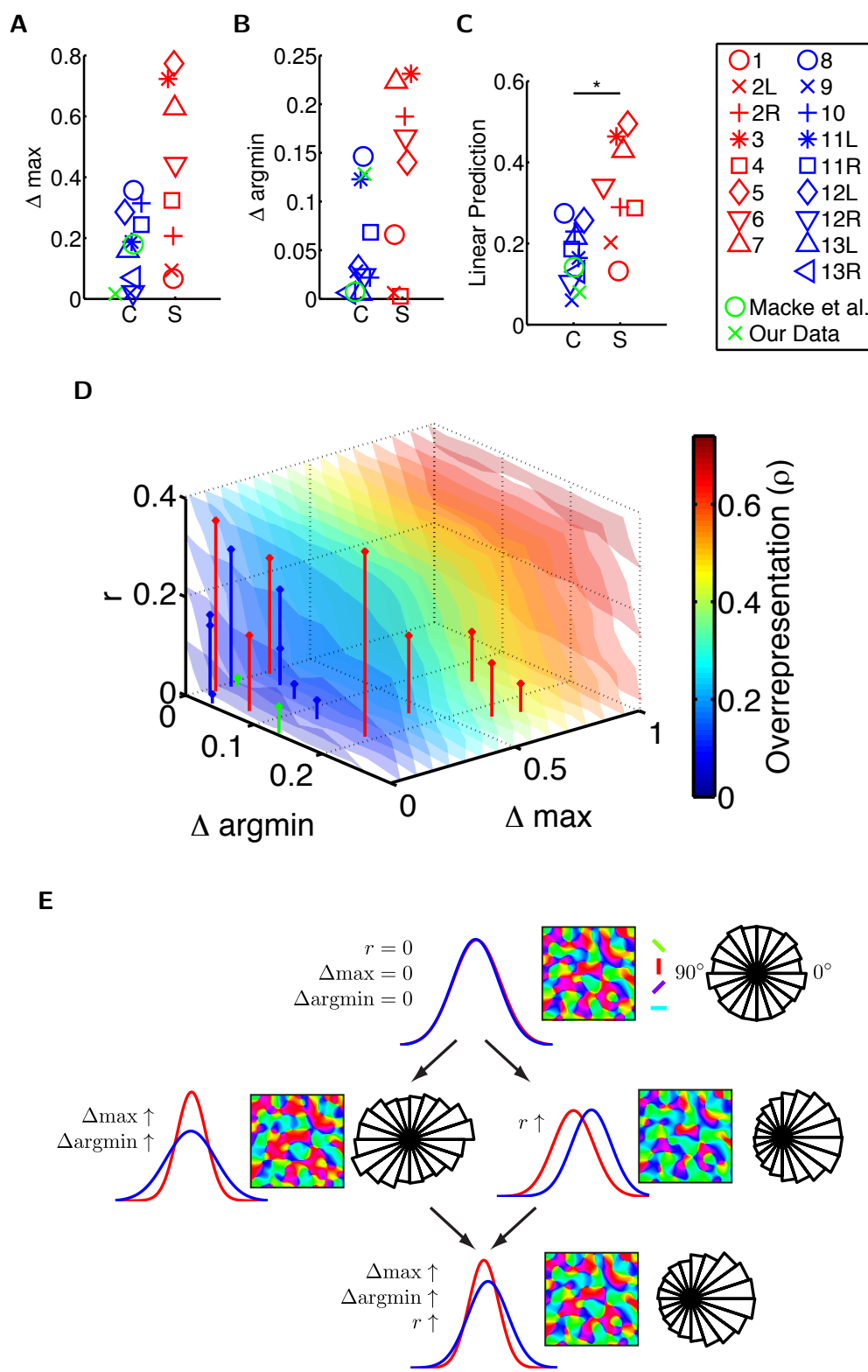
Figure 6

Figure 6: Map classification and overrepresentation prediction using fitted overrepresentation parameters.. (**A**) Relative absolute difference in maximum of marginal likelihood fitted covariance functions. (**B**) Relative absolute difference in argmin of marginal likelihood fitted covariance functions. (**C**) Results of applying the linear model to the maps, showing a significant split in the data by rearing condition (*$p < 0.012$, Mann-Whitney U test), a good match with the results in figures 5**A** and **E**. Symbols show which map each point represents, as per legend. C: control, S: stripe-reared. (**D**) Distribution of ρ in maps sampled from Gaussian process priors with varying values of the hyperparameters, with the parameter values fitted to the maps shown. The value of ρ increases linearly with each hyperparameter, as expected, but the values of each map sit at different locations on the planes on constant ρ, indicating that the same level of overrepresentation is caused by different combinations of the hyperparameters. Vertical lines indicate the location of the points in the Δmax-Δargmin plane. (**E**) Changes induced in maps by increasing different overrepresentation metrics. **(top)** A normal map, with identically distributed real and imaginary components, and an even distribution of orientations, as shown by the circular histogram. **(left)** A map with increased Δmax or Δargmin, with a change in the variance of the components, and overrepresentation of $0°$ and $90°$. **(right)** A map with increased r, showing shifted distributions, and overrepresentation of $0°$. **(bottom)** A map with all three measures increased to a lesser degree, showing both a change in shape and shift in the distributions. Compare these distributions to figure 5**D**, and note that the map has a subtle but noticeable level of overrepresentation of $0°$ (cyan pixels, see legend). The normal map was sampled from a GP prior with $\alpha = 5$ and $\sigma = 4$ and the value distributions shown are fitted Gaussians.

4 Discussion

To determine the effects of experience on brain structure it is critical to have reliable statistical methods for determining that structure based on noisy experimental data. The method for estimating orientation preference maps from optical imaging data introduced by Macke et al. (2011) was an important step forward in studying visual cortical maps, and provided substantial improvements over the classical technique of vector averaging. However, the original method was unable to address the effects of rearing in an abnormal visual environment, which is a paradigmatic method for investigating brain plasticity. We therefore generalised the method, including introducing a prior which allows some orientations to be overrepresented. This revealed that a three-dimensional measure of overrepresentation is required to capture the full nature of the induced changes in map structure following stripe rearing.

Previous work on the effect of stripe rearing characterised map structure using only a single parameter comparing pixel counts of cortical regions representing different orientations, equivalent to our ρ metric (Sengpiel et al., 1999; Tanaka et al., 2006). This suffices to indicate whether or not maps exhibit significant overrepresentation, but conveys no other information about how the structure of the maps are different to those from control animals. In contrast our three-dimensional metric directly measures changes in the spatial layout of orientation preference and tuning strength. The three parameters are derived from the standard vector field representation of orientation maps (Blasdel and Salama, 1986), and represent the relative absolute difference between the wavelengths of the real and imaginary components (Δargmin), the relative absolute difference between the scaling of these two components (Δmax), and the difference in the mean of the two components (r). How varying each of these parameters independently affects map structure is illustrated in figure 6**E**. We found that the variation of these parameters between normal and stripe-reared maps has a strikingly diverse pattern, but that, despite this, together these three values provide a more reliable prediction of rearing condition than ρ alone.

The diversity in parameter values between animals in the same rearing condition is intriguing, and suggests one or both of the following possibilities. The more subtle statistics of the visual environment during rearing in the experiments of Sengpiel et al. (1999) were not tightly controlled, and it is thus possible that some of the parameter diversity we observed is due to variations in these statistics between animals. In support of this possibility, simulations have shown that subtle differences in, for instance, the cocircularity of visual input during development have measurable effects on orientation map structure (Hunt et al., 2009). Alternatively, the parameter diversity could be due to genetic effects. In support of this possibility, recent work in rodents has shown that the orientation preferences of clonally-related neurons in V1 are more similar than those of unrelated neurons (Ohtsuki et al., 2012; Li et al., 2012), and recent work in cats has shown that molecular cues might guide the initial segregation of ocular dominance columns (Tomita et al., 2012). Some of the maps we analysed here were two hemispheres from the same animals, however the sample size was too small to determine whether these tend to have more similar statistics than maps from different animals (data not shown; see also Hunt et al. (2009); Kaschube et al. (2002)). Potentially there could also be a gene-environment interaction, such that the variance in results is not explained by treating genetic and environmental effects as independent variables (Hunter, 2005). The sample sizes required to address these questions are challenging to obtain, given that rodents do not have smooth orientation maps and so animals such as cats must be used, for which ethical concerns

usually severely limit numbers. Stripe-rearing of cats is also extremely labour-intensive.

In many situations where one is attempting to estimate anatomical or functional structures in the brain some information about the underlying form is already known. In these cases, the use of Bayesian inference (of which Gaussian process regression is a particular case) with a suitable prior encoding that knowledge allows much more robust estimations to be made (Friston et al., 2002). Bayesian methods have made an especially large impact in the analysis of functional MRI data (Woolrich, 2012). Gaussian process regression lends itself well to estimation tasks where the underlying structure is of a well known form which can be expressed mathematically. This is true for orientation preference maps, as well as some other visual cortical maps, but may apply equally well to other functional topographic maps in the cortex, such as those found in auditory (e.g., Aitkin et al. (1986)), olfactory (e.g., Astic et al. (1987)) and gustatory cortex (e.g., Chen et al. (2011)).

We demonstrated the implementation of an efficient and robust marginal likelihood maximisation technique for estimating the parameters of Gaussian process models for visual maps. Although we used this method to estimate the six parameters of our stripe-reared prior, we also showed that it performed well estimating the two parameters (α and σ) of the prior of Macke et al. (2011). This removes any subjective interpretation of the data, which can otherwise be necessary when the data is especially noisy, for example when spatially filtering an orientation preference map. The method also provides a set of parameters which are known to be optimal (under the assumption that the prior is of an appropriate form), and does not a require vector averaged map to be generated and analysed as in Macke et al. (2011). While this method is not free of assumptions, those that it does make are clearly stated, which is preferable to the subtle and subjective assumptions made by previous techniques.

The role played by visual input in early life in the formation and refinement of visual cortical maps is still an active research topic (Espinosa and Stryker, 2012). Investigations into the effects of abnormal visual input on visual cortical maps have also included ocular dominance maps (Shatz and Stryker, 1978) and the spatial relationships between these and orientation maps (Crair et al., 1997). The generalised implementation presented here allows the application of this method to maps of ocular dominance, with an appropriate chosen prior (perhaps based on Rojer and Schwartz (1990)), as well as both orientation and ocular dominance maps together, with a prior defining the relationship between them. Other visual cortical maps, such as for direction preference, spatial frequency, retinotopic location, and colour could also be more reliably estimated using this technique, given suitable priors. Using this technique to investigate what can be learnt about the effects of visual input on these maps and their combinations is a promising direction for further research.

References

L. M. Aitkin, M. M. Merzenich, D. R. Irvine, J. C. Clarey, and J. E. Nelson. Frequency representation in auditory cortex of the common marmoset (Callithrix jacchus jacchus). *J. Comp. Neurol.*, 252(2):175–85, 1986.

P. A. Anderson, J. Olavarria, and R. C. Van Sluyters. The overall pattern of ocular dominance bands in cat visual cortex. *J. Neurosci.*, 8(6):2183–200, 1988.

L. Astic, D. Saucier, and A. Holley. Topographical relationships between olfactory receptor cells and glomerular foci in the rat olfactory bulb. *Brain Res.*, 424:144–52, 1987.

C. Blakemore and G. F. Cooper. Development in the brain depends on the visual environment. *Nature*, 228:477–8, 1970.

G. Blasdel and G. Salama. Voltage-senstitive dyes reveal a modular organization in monkey striate cortex. *Nature*, 321:579–85, 1986.

T. Bonhoeffer, D.-S. Kim, D. Malonek, D. Shoham, and A. Grinvald. Optical Imaging of the Layout of Functional Domains in Area 17 and Across the Area 17/18 Border in Cat Visual Cortex. *Eur. J. Neurosci.*, 7, 1995.

M. F. Cardoso, R. L. Salcedo, and S. Feyo de Azevedo. The simplex-simulated annealing approach to continuous non-linear optimization. *Comput. Chem. Eng.*, 20(9):1065–1080, 1996.

M. A. Carreira-Perpiñán, R. J. Lister, and G. J. Goodhill. A computational model for the development of multiple maps in primary visual cortex. *Cerebral Cortex*, 15(8):1222–1233, 2005.

X. Chen, M. Gabitto, Y. Peng, N. J. P. Ryba, and C. S. Zuker. A gustotopic map of taste qualities in the mammalian brain. *Science*, 333(6047):1262–6, 2011.

M. C. Crair, E. S. Ruthazer, D. C. Gillespie, and M. P. Stryker. Relationship between the ocular dominance and orientation maps in visual cortex of monocularly deprived cats. *Neuron*, 19(2):307–18, 1997.

J. S. Espinosa and M. P. Stryker. Development and plasticity of the primary visual cortex. *Neuron*, 75(2):230–49, 2012.

K. J. Friston, W. Penny, C. Phillips, S. Kiebel, G. Hinton, and J. Ashburner. Classical and Bayesian inference in neuroimaging: theory. *NeuroImage*, 16(2):465–83, 2002.

R. E. H and G. A. A tandem-lens epifluorescence macroscope – hundred-fold brightness advantage for wide-field imaging. *J. Neurosci. Meth.*, 36:127–37, 1991.

M. Hübener, D. Shoham, A. Grinvald, and T. Bonhoeffer. Spatial relationships among three columnar systems in cat area 17. *J. Neurosci.*, 17(23):9270–84, 1997.

J. J. Hunt, C. E. Giacomantonio, H. Tang, D. Mortimer, S. Jaffer, V. Vorobyov, G. Ericksson, F. Sengpiel, and G. J. Goodhill. Natural scene statistics and the structure of orientation maps in the visual cortex. *NeuroImage*, 47(1):157–72, 2009.

D. J. Hunter. Gene-environment interactions in human diseases. *Nat. Rev. Genet.*, 6(4): 287–98, 2005.

N. P. Issa, C. Trepel, and M. P. Stryker. Spatial frequency maps in cat visual cortex. *J. Neurosci.*, 20(22):8504–14, 2000.

M. Kaschube, F. Wolf, T. Geisel, and S. Löwel. Genetic influence on quantitative features of neocortical architecture. *J. Neurosci.*, 22(16):7206–17, 2002.

Y. Li, H. Lu, P. Cheng, S. Ge, H. Xu, S. H. Shi, and Y. Dan. Clonally Related Visual Cortical Neurons Show Similar Stimulus Feature Selectivity. *Nature*, 486(7401):118–21, 2012.

J. H. Macke, S. Gerwinn, L. E. White, M. Kaschube, and M. Bethge. Gaussian process methods for estimating cortical maps. *NeuroImage*, 56(2):570–81, 2011.

G. Ohtsuki, M. Nishiyama, T. Yoshida, T. Murakami, M. Histed, C. Lois, and K. Ohki. Similarity of visual selectivity among clonally related neurons in visual cortex. *Neuron*, 75 (1):65–72, 2012.

C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, 2006.

A. S. Rojer and E. L. Schwartz. Cat and monkey cortical columnar patterns modeled by bandpass-filtered 2D white noise. *Biol. Cybern.*, 62:381–91, 1990.

F. Sengpiel, P. Stawinski, and T. Bonhoeffer. Influence of experience on orientation maps in cat visual cortex. *Nature Neurosci.*, 2(8):727–32, 1999.

C. J. Shatz and M. P. Stryker. Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *J. Physiol.*, 281:267–83, 1978.

A. Shmuel and A. Grinvald. Functional organization for direction of motion and its relationship to orientation maps in cat area 18. *J. Neurosci.*, 16(21):6945–64, 1996.

S. Tanaka, J. Ribot, K. Imamura, and T. Tani. Orientation-restricted continuous visual exposure induces marked reorganization of orientation maps in early life. *NeuroImage*, 30 (2):462–77, 2006.

K. Tomita, M. Sperling, S. B. Cambridge, T. Bonhoeffer, and M. Hübener. A Molecular Correlate of Ocular Dominance Columns in the Developing Mammalian Visual Cortex. *Cereb. Cortex*, pages 1–11, 2012.

J. P. van Kleef, S. L. Cloherty, and M. R. Ibbotson. Complex cell receptive fields: evidence for a hierarchical mechanism. *J. Physiol.*, 588:3457–70, 2010.

M. Weliky, W. H. Bosking, and D. Fitzpatrick. A systematic map of direction preference in primary visual cortex. *Nature*, 379(6567):725–8, 1996.

M. W. Woolrich. Bayesian inference in fMRI. *NeuroImage*, 62(2):801–10, 2012.