



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



Nicholas J. Hughes<sup>a</sup>, Jonathan J. Hunt<sup>a</sup>, Shaun L. Cloherty<sup>b,c</sup>, Michael R. Ibbotson<sup>b,c</sup>, Frank Sengpiel<sup>d</sup>, Geoffrey J. Goodhill<sup>a,e,\*</sup>

<sup>a</sup> Queensland Brain Institute, The University of Queensland, St Lucia, Queensland 4072, Australia

<sup>b</sup> National Vision Research Institute, Australian College of Optometry, Carlton, Victoria 3053, Australia

<sup>c</sup> Department of Optometry and Vision Science, University of Melbourne, Parkville, Victoria 3010, Australia

<sup>d</sup> School of Biosciences, Cardiff University, Cardiff, CF10 3AX Wales, UK

<sup>e</sup> School of Mathematics and Physics, The University of Queensland, St Lucia, Queensland 4072, Australia

### ARTICLE INFO

#### Article history:

Accepted 10 March 2014

Available online 20 March 2014

#### Keywords:

Optical imaging

Gaussian processes

Visual cortex

Orientation preference maps

Stripe-rearing

### ABSTRACT

An important example of brain plasticity is the change in the structure of the orientation map in mammalian primary visual cortex in response to a visual environment consisting of stripes of one orientation. In principle there are many different ways in which the structure of a normal map could change to accommodate increased preference for one orientation. However, until now these changes have been characterised only by the relative sizes of the areas of primary visual cortex 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. We first formulated a suitable prior for the stripe-reared case, and developed an efficient method for maximising the marginal likelihood of the model in order to determine the optimal parameters. We then applied this to a set of orientation maps from normal and stripe-reared cats. This analysis revealed that several parameters of overall map structure, specifically the difference between wavelength, scaling and mean of the two vector components of maps, changed in response to stripe-rearing, which together give a more nuanced assessment of the effect of rearing condition on map structure than previous measures. Overall this work expands our understanding of the effects of the environment on brain structure.

© 2014 Elsevier Inc. All rights reserved.

### Introduction

A critical question for understanding brain plasticity is to understand how the sensory environment influences brain structure. An important and common model system in this regard is the effect of visual activity during early life on the spatial arrangement of topographic maps in the primary visual cortex of mammals such as cats, ferrets and monkeys (Espinosa and Stryker, 2012). 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), has 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; Hirsch and Spinelli, 1970). This leads to an increase

in the proportion of primary visual cortical 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 is not robust to changes in the subjective spatial filtering applied to maps, which are traditionally obtained by vector averaging noisy data from single-condition optical imaging experiments. Furthermore this measure does not provide a method for determining precisely how the structure of maps changes to accommodate the larger preference for a single orientation which is 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 method 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, this method contained assumptions which are violated by maps from abnormally-reared animals.

We therefore first generalised this Gaussian process method to a much broader class of visual map data, including abnormal rearing

\* Corresponding author at: Queensland Brain Institute, The University of Queensland, St Lucia, Queensland 4072, Australia.

E-mail address: [g.goodhill@uq.edu.au](mailto:g.goodhill@uq.edu.au) (G.J. Goodhill).

conditions. We then used it to interrogate map structure in data from the stripe-reared cats of Sengpiel et al. (1999). Besides allowing for 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. These three parameters are the difference between the two vector components in terms of their wavelengths, scalings and means. 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.

## Materials and methods

### Orientation preference maps

We represent an orientation preference map as a 2D complex field  $\mathbf{m}(\mathbf{x})$ ,  $\mathbf{x} \in \mathbb{R}^2$  (mathematically equivalent to a 2D vector field), where the preferred orientation and the strength of that preference at each point are given by half the argument and the magnitude, respectively, of the complex field at that point. The argument is halved as orientation is periodic in  $\pi$  rather than  $2\pi$  radians.

The classical method of estimating orientation preference maps from imaging data is known as vector averaging,

$$\mathbf{m}(\mathbf{x}) = \frac{1}{N} \sum_{j=1}^N (r_j(\mathbf{x}) \exp(2\theta_j i)), \quad (1)$$

where  $r_j(\mathbf{x})$  is the cortical response at location  $\mathbf{x} \in \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^2$  during the  $j^{\text{th}}$  trial out of  $N$ , where a stimulus of orientation  $\theta_j$  was presented. This map is simply the vectorially-averaged raw response to the stimuli. To account for noise in the result, vector averaged maps are usually then high- and low-pass Gaussian spatially filtered.

### Gaussian process regression map estimation

Macke et al. (2011) introduced a novel method for estimating orientation preference maps from optical imaging data by applying Gaussian process (GP) regression (Rasmussen and Williams, 2006) to the problem of estimating a 2D vector field from noisy data. The important findings of the paper 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. Here we define the model equivalently but in our own notation for the convenience of the extensions we make below.

Hereafter let  $\mathbf{r}_j = (r_j(\mathbf{x}_1), \dots, r_j(\mathbf{x}_n))^T$  refer to the vector of responses at each of the  $n$  observed points in the cortex (i.e., the experimental image reshaped as a vector). Let  $\mathbf{m}(\mathbf{x})$  be the true orientation preference map underlying the data, and let  $\mathbf{m}$  refer to the flattened and concatenated vector of the true map's components, that is,

$$\mathbf{m} = (\text{real}(\mathbf{m}(\mathbf{x}_1)), \dots, \text{real}(\mathbf{m}(\mathbf{x}_n)), \text{imag}(\mathbf{m}(\mathbf{x}_1)), \dots, \text{imag}(\mathbf{m}(\mathbf{x}_n)))^T.$$

Then we assume that the data can be written as  $\mathbf{r}_j = V_j^T \mathbf{m} + \epsilon_j$ , where  $\epsilon_j \sim \mathcal{N}(\mathbf{0}, \Sigma_\epsilon)$  and

$$V_j = (\cos 2\theta_j, \sin 2\theta_j)^T \otimes \mathbb{I}_n,$$

where  $\mathbb{I}_n$  is an  $n \times n$  identity matrix and  $\otimes$  is the Kronecker product. That is, we observe the component of the underlying map in the  $\theta_j$  direction corrupted by some correlated Gaussian noise.

A prior defining a distribution of maps is now defined as a Gaussian process, defined by an identically zero mean function and a covariance function  $K$  which specifies the covariance between points in the map both within and across the two map vector components. Macke et al. (2011) used a difference of Gaussians covariance function,

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

where  $\delta_{ij}$  is the Kronecker delta,  $\mathbf{x}$  and  $\mathbf{x}'$  are the two positions in the map, and  $k$  and  $k'$  are either 1 or 2, the real and imaginary components of the map respectively. The assumption that the components are uncorrelated gives  $K(\mathbf{x}, \mathbf{x}', 1, 2) \equiv K(\mathbf{x}, \mathbf{x}', 2, 1) \equiv 0$ . Macke et al. (2011) set  $\alpha_1 = -\alpha_2$  so that  $K$  has zero mean, and  $\sigma_2 = 2\sigma_1$  to reduce the number of parameters. These parameters can be interpreted as scaling parameters  $\alpha_i$ s and filter widths  $\sigma_i$ s, which correspond to the wavelength of the map. Macke et al. (2011) showed that this function fit the empirical autocovariance functions of their data well. We found the same result with our data, and additionally found that letting all four parameters (i.e., both  $\alpha_i$  and both  $\sigma_i$ ) vary independently did not substantially improve the fit to the empirical functions (data not shown). We therefore chose to use the same covariance function as Macke et al. (2011) but with all the multiplying constants (i.e., both  $\alpha_i$ , the sum of squared  $\sigma_i$  and  $2\pi$ ) collected into a single scaling parameter coefficient, which we call  $\alpha$ , and set  $\sigma := \sigma_2 = 2\sigma_1$  for notational simplicity:

$$K(\mathbf{x}, \mathbf{x}', k, k') = \delta(k, k') \alpha^2 \left[ \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{4\sigma^2}\right) + \frac{1}{4} \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{16\sigma^2}\right) - \frac{4}{5} \exp\left(-\frac{|\mathbf{x}-\mathbf{x}'|^2}{10\sigma^2}\right) \right]. \quad (2)$$

To perform Gaussian process regression, the prior distribution defined at the observed points  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  is conditioned on the data to give a posterior distribution, the mean of which is the Bayes optimal estimate of the underlying map  $\mathbf{m}$ . The prior on the set of observed points is  $p(\mathbf{m}) = \mathcal{N}(\mathbf{m}; \mathbf{0}, \Sigma_{\text{prior}})$ , where the covariance matrix is

$$\Sigma_{\text{prior}} = \begin{pmatrix} K(X, X, 1, 1) & K(X, X, 1, 2) \\ K(X, X, 2, 1) & K(X, X, 2, 2) \end{pmatrix}.$$

This matrix needs to be calculated, stored and inverted, which is a computationally demanding task. To make this tractable, Macke et al. (2011) used a low-rank approximation to the prior, generated by an incomplete Cholesky decomposition. This results in a prior of the form  $\Sigma_{\text{prior}} = D + GG^T$ , where  $D$  is  $2n \times 2n$  and diagonal, and  $G$  is  $2n \times q$ , with  $q \ll 2n$ . The eigenspectrum of the chosen covariance function goes to zero sufficiently quickly to make the incomplete Cholesky decomposition a good approximation. To estimate the parameters  $\alpha$  and  $\sigma$  of the prior covariance function (Eq. 2), Macke et al. (2011) fit the covariance function to the average of the radial component of the autocorrelation function of the two components of the filtered, vector averaged map.

As per Bayes' rule, the posterior distribution is proportional to the product of the prior and the likelihood,

$$p(\mathbf{m} | \mathbf{r}_1, \dots, \mathbf{r}_N) \propto p(\mathbf{m}) p(\mathbf{r}_1, \dots, \mathbf{r}_N | \mathbf{m}).$$

The likelihood of a single trial is  $p(\mathbf{r}_j|\mathbf{m}) = \mathcal{N}(\mathbf{r}_j; V_j^T \mathbf{m}, \Sigma_\epsilon)$ . The likelihood of all trials is the product of the individual likelihoods, assuming each is an independent observation of the true map.

$$\begin{aligned} p(\mathbf{r}_1, \dots, \mathbf{r}_N|\mathbf{m}) &= \prod_{j=1}^N \mathcal{N}(\mathbf{r}_j; V_j^T \mathbf{m}, \Sigma_\epsilon) \\ &= \prod_{j=1}^N \mathcal{N}\left(\mathbf{m}; V_j \mathbf{r}_j, (V_j \Sigma_\epsilon^{-1} V_j^T)^{-1}\right) \\ &\propto \mathcal{N}\left(\mathbf{m}; \frac{2}{N} \begin{pmatrix} \Sigma_\epsilon & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix} \sum_{j=1}^N V_j \Sigma_\epsilon^{-1} \mathbf{r}_j; \frac{2}{N} \begin{pmatrix} \Sigma_\epsilon & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix}\right) \end{aligned}$$

Above we have used the fact that, if the stimulus orientations are uncorrelated on average,  $\sum_{j=1}^N V_j V_j^T = \frac{N}{2} \mathbb{I}_{2n}$ , and that  $V_j^T V_j = \mathbb{I}_n$ . Now as the likelihood and the prior are Gaussian, so is the posterior,

$$p(\mathbf{m}|\mathbf{r}_1, \dots, \mathbf{r}_N) \propto \mathcal{N}(\mathbf{m}; \mu_{\text{post}}, \Sigma_{\text{post}}),$$

where

$$\mu_{\text{post}} = \Sigma_{\text{post}} \begin{pmatrix} \Sigma_\epsilon & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix}^{-1} \sum_{j=1}^N \mathbf{y}_j,$$

and

$$\Sigma_{\text{post}} = \left( \Sigma_{\text{prior}}^{-1} + \frac{N}{2} \begin{pmatrix} \Sigma_\epsilon & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix}^{-1} \right)^{-1},$$

respectively, where we have defined  $\mathbf{y}_j = \begin{pmatrix} \cos 2\theta_j \\ \sin 2\theta_j \end{pmatrix} \otimes \mathbf{r}_j$ , and  $\mu_{\text{post}}$  is the

best estimate of the underlying map  $\mathbf{m}$ .  $\Sigma_{\text{post}}$  gives an indication of error in the fit, with its diagonal giving the variance in the estimate at each pixel.

There are numerous sources of noise in images of the cortex, and therefore the noise covariance matrix  $\Sigma_\epsilon$  should in general be of full rank, modelling both 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 that the entirety of  $\Sigma_\epsilon$  cannot be inferred from the data. Macke et al. (2011) overcame this problem by using a noise covariance of the form  $\Sigma_\epsilon = D_\epsilon + G_\epsilon G_\epsilon^T$ , where  $D_\epsilon$  is  $n \times n$  and diagonal, and  $G_\epsilon$  is of size  $n \times q_\epsilon$ , with  $q_\epsilon \ll n$ .  $D_\epsilon$  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  $q_\epsilon$ . Noise correlations in cortical imaging data are primarily between nearby pixels, making this a good compromise. Note that the matching low-rank forms of the prior covariance and noise covariance matrices are coincidental, and result from separate assumptions.

Macke et al. (2011) 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. These estimates are then averaged over all the stimulus conditions, giving an initial estimate of the pixel-wise noise variance  $D_\epsilon$ . 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.

As both the prior and noise covariance matrices, as they appear in the posterior mean and covariance, are in the form  $A = D + GG^T$  where  $D$  is diagonal, and  $G$  is of size  $2n \times q$ , where  $q \ll 2n$ , they can be inverted using the matrix inversion lemma (Rasmussen and Williams, 2006),

making the calculation of the posterior distribution computationally tractable.

#### Efficient inversion of a general prior covariance matrix

The most general form of prior covariance matrix for estimating orientation preference maps using GP regression is

$$\Sigma_{\text{prior}} = \begin{pmatrix} A & B \\ B & C \end{pmatrix}, \quad (3)$$

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 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 (Eq. 3), this method must be adjusted, but can be performed in a similar manner.

Firstly,  $A$ ,  $B$  and  $C$  are factored separately using incomplete Cholesky decompositions, into the forms  $A = D_A + G_A G_A^T$ ,  $C = D_C + G_C G_C^T$  and  $B = G_B G_B^T$ . 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 a special treatment as for  $A$  and  $C$ .

Now we set  $G_1 = \begin{pmatrix} G_A \\ 0 \end{pmatrix}$ ,  $G_2 = \begin{pmatrix} 0 \\ G_C \end{pmatrix}$ ,  $G_3 = \begin{pmatrix} G_B \\ 0 \end{pmatrix}$ ,  $H_3 = (0 \ G_B^T)$ ,  $G_4 = \begin{pmatrix} 0 \\ G_B \end{pmatrix}$  and  $H_4 = (G_B^T \ 0)$ , where  $(A \ 0)$  indicates the block matrix formed by  $A$  and an appropriately sized matrix of zeros. It is simple to check that we now have  $\Sigma_{\text{prior}} = G_1 G_1^T + G_2 G_2^T + G_3 H_3 + G_4 H_4 + \begin{pmatrix} D_A & 0 \\ 0 & D_C \end{pmatrix}$ . Setting  $G = (G_1 \ G_2 \ G_3 \ G_4)$ ,  $H = (G_1^T \ G_2^T \ H_3 \ H_4)^T$ , and  $D = \begin{pmatrix} D_A & 0 \\ 0 & D_C \end{pmatrix}$  gives  $\Sigma_{\text{prior}} = D + GH$ . Therefore the matrix inversion lemma,

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

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

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

#### Marginal likelihood maximisation

Macke et al. (2011) chose the parameters of the prior by fitting its covariance function to the autocorrelation function of a filtered, vector averaged map. This is computationally efficient, but in this Bayesian framework a more robust (though more computationally demanding) method for inferring the hyperparameters is to maximise the marginal likelihood (Rasmussen and Williams, 2006). The marginal likelihood is a quantity which gives the probability that a given set of data came from a given prior. The set of parameters which maximises this quantity defines the optimal prior for each dataset.

The marginal likelihood is defined as the integral, over the underlying function, of the product of the likelihood and the prior (Rasmussen and Williams, 2006),

$$ML = p(\mathbf{r}_1, \dots, \mathbf{r}_N | \Sigma_{\text{prior}}) = \int p(\mathbf{m}) p(\mathbf{r}_1, \dots, \mathbf{r}_N | \mathbf{m}, \Sigma_{\text{prior}}) d\mathbf{m}.$$

Let  $L$  denote the integrand, which we derive first. Using the same notation as above, let  $\Sigma'_\epsilon = \begin{pmatrix} \Sigma_\epsilon & 0 \\ 0 & \Sigma_\epsilon \end{pmatrix}$  and  $\hat{\mathbf{y}} = \sum_{j=1}^N \mathbf{y}_j$  for convenience.

$$\begin{aligned} L &= p(\mathbf{m}) p(\mathbf{r}_1, \dots, \mathbf{r}_N | \mathbf{m}, \Sigma_{\text{prior}}) \\ &= \frac{(2\pi)^{\frac{n(1-N)}{2}}}{\left| \sum_{j=1}^N \prod_{k \neq j} (V_k \Sigma_\epsilon^{-1} V_k^\top)^{-1} \right|^{\frac{1}{2}}} \exp\left(\frac{1}{N} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}}\right) \\ &\quad - \frac{1}{2} \sum_{j=1}^N \mathbf{r}_j^\top \Sigma_\epsilon^{-1} \mathbf{r}_j \mathcal{N}(\mathbf{m}; 0, \Sigma_{\text{prior}}) \mathcal{N}\left(\mathbf{m}; \frac{2}{N} \hat{\mathbf{y}}, \frac{2}{N} \Sigma'_\epsilon\right) \\ &= c_0 c_1 c_2 \exp\left(-\frac{1}{2} \mathbf{m}^\top \left(\Sigma_{\text{prior}}^{-1} + \frac{N}{2} (\Sigma'_\epsilon)^{-1}\right) \mathbf{m} + \frac{1}{2} \mathbf{m}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}}\right) \\ &\quad + \frac{1}{2} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \mathbf{m} - \frac{1}{N} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}} \end{aligned}$$

Here  $c_0 = (2\pi)^{\frac{n(1-N)}{2}} \left| \sum_{j=1}^N \prod_{k \neq j} (V_k \Sigma_\epsilon^{-1} V_k^\top)^{-1} \right|^{-\frac{1}{2}} \exp\left(\frac{1}{N} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}} - \frac{1}{2} \sum_{j=1}^N \mathbf{r}_j^\top \Sigma_\epsilon^{-1} \mathbf{r}_j\right)$ ,  $c_1 = (2\pi)^{-n/2} \left| \frac{2}{N} \Sigma'_\epsilon \right|^{-1/2}$  and  $c_2 = (2\pi)^{-n/2} |\Sigma_{\text{prior}}|^{-1/2}$ . The term  $c_0$  corrects for writing the product of the  $N$  likelihoods as a single normal distribution. For the purposes of rearranging the equation into a convenient form we use the posterior mean  $\mu_{\text{post}}$  (see above), and factor the quadratic form.

$$\begin{aligned} L &= c_0 c_1 c_2 \exp\left(-\frac{1}{2} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}} - \frac{1}{2} (\mathbf{m} - \mu_{\text{post}})^\top \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right) (\mathbf{m} - \mu_{\text{post}})\right) \\ &\quad \times \left(\mathbf{m} - \mu_{\text{post}}\right) + \frac{1}{2} \mu_{\text{post}}^\top \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right) \mu_{\text{post}} \end{aligned}$$

The marginal likelihood is then the integral of the above with respect to  $\mathbf{m}$ .

$$\begin{aligned} ML &= c_0 c_1 c_2 \exp\left(-\frac{1}{2} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}} + \frac{1}{2} \mu_{\text{post}}^\top \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right) \mu_{\text{post}}\right) \int \exp\left(-\frac{1}{2} (\mathbf{m} - \mu_{\text{post}})^\top \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right) (\mathbf{m} - \mu_{\text{post}})\right) d\mathbf{m} \\ &= c_0 c_1 c_2 \exp\left(-\frac{1}{2} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}} + \frac{1}{2} \mu_{\text{post}}^\top \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right) \mu_{\text{post}}\right) \frac{(2\pi)^{\frac{n}{2}}}{\left| \frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1} \right|^{\frac{1}{2}}} \\ &= (2\pi)^{-\frac{np}{2}} \left| \sum_{j=1}^N \prod_{k \neq j} (V_k \Sigma_\epsilon^{-1} V_k^\top)^{-1} \right|^{\frac{1}{2}} \frac{2}{N} \Sigma'_\epsilon \Sigma_{\text{prior}} \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right)^{-\frac{1}{2}} \\ &\quad \exp\left(\frac{2-N}{2N} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}} - \frac{1}{2} \sum_{j=1}^N \mathbf{r}_j^\top \Sigma_\epsilon^{-1} \mathbf{r}_j + \frac{1}{2} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right)^{-1} (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}}\right) \end{aligned}$$

For numerical reasons, we usually calculate the log marginal likelihood, which is

$$\begin{aligned} \log ML &= -\frac{Nn}{2} \log 2\pi - \frac{1}{2} \left| \sum_{j=1}^N \prod_{k \neq j} (V_k \Sigma_\epsilon^{-1} V_k^\top)^{-1} \right| - \frac{1}{2} \log \left| \frac{2}{N} \Sigma'_\epsilon \right| - \frac{1}{2} \log |\Sigma_{\text{prior}}| \\ &\quad - \frac{1}{2} \log \left| \frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1} \right| - \frac{2-N}{2N} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}} - \frac{1}{2} \sum_{j=1}^N \mathbf{r}_j^\top \Sigma_\epsilon^{-1} \mathbf{r}_j \\ &\quad + \frac{1}{2} \hat{\mathbf{y}}^\top (\Sigma'_\epsilon)^{-1} \left(\frac{N}{2} (\Sigma'_\epsilon)^{-1} + \Sigma_{\text{prior}}^{-1}\right)^{-1} (\Sigma'_\epsilon)^{-1} \hat{\mathbf{y}}. \end{aligned}$$

We have shown above how to calculate the inverse in the last term of this equation efficiently. For the determinant in the fourth term, we can use the matrix determinant lemma after rearranging (Rasmussen and Williams, 2006),

$$\det(D + GH) = \det(D) \det(\mathbb{I} + HD^{-1}G).$$

We only modelled independent pixel-wise noise when calculating the marginal likelihood (i.e.,  $\Sigma_\epsilon$  is diagonal), both to reduce computational time, and because the results were not substantially different when correlated noise was included (data not shown). As this makes  $\Sigma_\epsilon$  a constant relative to the hyperparameters, to save computation time we did not calculate the terms which were not dependent on  $\Sigma_{\text{prior}}$ , as they do not affect the location of maxima, which substantially simplifies the above expression.

To maximise the marginal likelihood we tried several local optimisation algorithms. MATLAB's `fmincon`, which uses an interior point algorithm, 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 we tested it on maps sampled from priors with known parameters and it performed well (see the Robust estimation of hyperparameters section and Fig. 3).

#### Comparison metric

The metric used to compare maps, for the purpose of comparing relative map quality, was the absolute value of the complex correlation coefficient of two maps represented as one-dimensional complex vectors, which provides a simple and tractable measure between 0 and 1 for overall map comparison.

#### Simple overrepresentation quantification

To aid comparison with the original analysis of Sengpiel et al. (1999), we calculated a simple scalar value for overrepresentation  $\rho$ . This was defined by first binning each pixel in a vector averaged map by its nearest preferred orientation out of  $S = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ . The value  $\rho$  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,  $n_\theta$  is the number of pixels preferring  $\theta$ , and  $\theta'$  is either the reared orientation or the maximally represented orientation in control maps. This metric varies linearly from zero for maps with no overrepresentation, to three for maps solely containing the reared orientation. Maps with an overrepresentation of another orientation yield negative values of  $\rho$ .

Other simple metrics, such as  $\rho$  but with more bins (e.g.,  $S = \{0^\circ, 22.5^\circ, \dots, 157.5^\circ\}$ ), or the magnitude of the average of all vectors in a map, could also be used. We chose  $\rho$  as it is simple, matches well with the quantification used in Sengpiel et al. (1999), and was either equally as successful or better than these other measures in predicting

**Table 1**

Summary of the empirical datasets. 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.'s (2011) datasets.

ID	Cat name	Hemisphere(s)	Treatment	Blocks/ stimulus	Pixel size ( $\mu\text{m}$ )
C1	C29_97	Left	135° stripe	10	22
C2	C45_97	Left & right	0° stripe	8	22
C3	C28_97	Left	90° stripe	6	22
C4		Right	90° stripe	8	22
C5	C16_97	Left	90° stripe	12	22
C6		Right	90° stripe	12	22
C7	C69_97	Right	45° stripe	14	22
C8	C27_97	Left	Normal	10	22
C9		Right	Normal	12	22
C10	N032207	Right	Normal	14	22
C11	N033006	Left & right	Normal	12	22
C12	N013103	Left & right	Normal	12	22
C13	C842	Left & right	Normal	6	22
Macke et al. (2011)	–	–	Normal	100	30
Our data	–	Right	Normal	80	24

overrepresentation when used to construct support vector machines (data not shown, also see the Prediction and classification section).

*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 (Fig. 4B). To formalise this notion we used the absolute relative difference between the max and argmin of the two components, written  $\Delta\text{max}$  and  $\Delta\text{argmin}$ , respectively (Fig. 4E). Formally,  $\Delta\text{max}$  is defined as

$$\Delta\text{max} = \frac{|\text{max}_{\text{real}} - \text{max}_{\text{imaginary}}|}{\frac{1}{2}(\text{max}_{\text{real}} + \text{max}_{\text{imaginary}})} \tag{4}$$

and  $\Delta\text{argmin}$  as

$$\Delta\text{argmin} = \frac{|\text{argmin}_{\text{real}} - \text{argmin}_{\text{imaginary}}|}{\frac{1}{2}(\text{argmin}_{\text{real}} + \text{argmin}_{\text{imaginary}})} \tag{5}$$

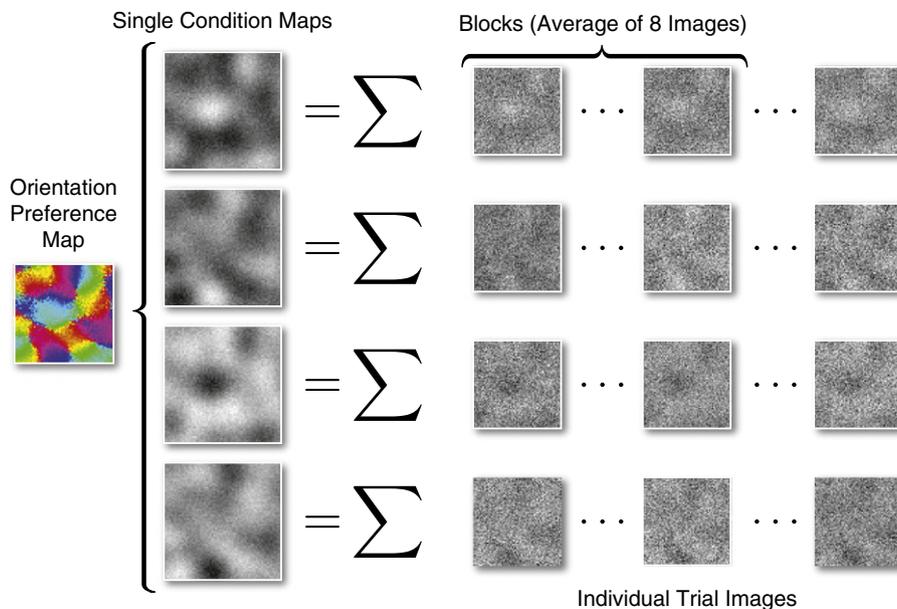
Intuitively, these measures represent the difference between the wavelengths and scalings of the two components of maps, a non-zero value of either of which will create an uneven distribution of orientation preference, by preferentially affecting one component (i.e., real or imaginary) of the map.

*Data*

We used three different empirical datasets for the analyses in this work, all consisting of optical imaging of either cat or ferret primary visual cortex: the images from ferrets used in Macke et al. (2011), images of a single cat collected by us for use in the development of this method (see below for full experimental methods), and images from cats used in Sengpiel et al. (1999) (which contains full details of the experimental procedures). The datasets are summarised in Table 1. In both the data from Sengpiel et al. (1999) and our experiment, cortical areas 17 and 18 were imaged. In the cat, areas 17 and 18 are both geniculorecipient areas and are therefore regarded as primary visual cortex (Payne and Peters, 2001). We will refer to these areas as either areas 17/18 or as primary visual cortex throughout the paper.

For the ferret data we used the area of the images defined as region 2 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. Thirteen different image sets in this form were available for 17 hemispheres of 10 cats (of either sex; Table 1). A schematic showing the form of the block data is given in Fig. 1. Note that for this averaged data the noise estimation is for the noise on the block data. Because each block is the average of 8 trials,



**Fig. 1.** Schematic showing the form of the available data from Sengpiel et al. (1999). Individual trial images were available for our data and the data from Macke et al. (2011), but only block data (the average of sets of 8 images) were available for the data from Sengpiel et al. (1999).

this will underestimate the variance of the noise on the original data by a factor of 8. Since the fitting is performed on the block data, the noise estimate on the block data is used. However, care should be taken if comparing the noise level between the block datasets and other datasets.

We performed an extra experiment on a single cat, which provided data with both more trials at more stimulus orientations as well as a higher spatial resolution than the other available datasets. We used this data primarily as a comparison, to investigate the effects of averaging data.

#### Prior sampled maps

For the purposes of verifying the implementation and testing the validity of the marginal likelihood optimisation and the support vector machine predictions, as well as training the linear predictor, we used prior sampled maps. A Gaussian process prior defines a distribution over maps, and we generated these maps by sampling from prior distributions with specified hyperparameters, then converting the maps into the appropriate form (such as simulated individual trial images), 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 Gaussian noise with the prior covariance function to generate the two components of a map.

The training data used to train the linear predictor consisted of  $2 \times 10^5$  maps sampled from our stripe-reared prior (see the [Results and Prediction and classification](#) sections) in the way described above, for a range of the parameters  $\alpha_{\text{real}}$ ,  $\sigma_{\text{real}}$ ,  $\alpha_{\text{imaginary}}$ ,  $\sigma_{\text{imaginary}}$  and  $r$ , to overrepresent one of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  or  $135^\circ$  to varying degrees. The values of  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$  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  $\rho$ .

#### Elastic net maps

To determine general rules for the covariance functions of stripe-reared maps, and to provide an alternative dataset for performing support vector machine predictions, maps with overrepresented orientations were generated using the elastic net model. This model has been shown to produce overrepresentation of the reared orientation resembling that found in experimental maps ([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 ([Carreira-Perpiñán and Goodhill, 2004](#)). It folds a two-dimensional net in higher-dimensional space in an attempt to minimise a tradeoff between the coverage of the input space and the continuity of the cortical representation. To minimise computation time, we only included visual field position and 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 only of visual field position, and orientation preference and selectivity. Twelve maps were generated with the orientations  $0^\circ$ ,  $15^\circ$ , ...,  $165^\circ$  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 other orientations. 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 the same custom MATLAB software 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>.

#### Data preprocessing

In most of the data used here the imaging covered areas outside of 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 primary visual cortex were considered. The presence of robust orientation preferences can be used to reliably determine primary visual cortex, and we used this approach to mask all of the image sets used here (already completed by [Hunt et al. \(2009\)](#) for the data from [Sengpiel et al. \(1999\)](#)), except for the Macke dataset, as the chosen region already contained only a small number of cells outside primary visual cortex. 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 Gaussian filtered to remove noise, and to be consistent with how they are usually processed ( $\sigma_{\text{lowpass}}$  and  $\sigma_{\text{highpass}}$  were 175 and 275  $\mu\text{m}$  for our data, 75 and 175  $\mu\text{m}$  for data from [Macke et al. \(2011\)](#), and 50 and 150  $\mu\text{m}$  for the data from [Sengpiel et al. \(1999\)](#), all based on the estimated map wavelengths). The data used for the Gaussian process method was not filtered in any way. When calculating the value of  $r$  for vector averaged maps ([Fig. 5D](#)), we calculated the mean as the average of pixel values weighted by tuning strength, to reduce the effects of noise in unresponsive areas of cortex in our estimates.

For consistency, all datasets were normalised to have a zero mean and unit variance, calculated across all pixels in all images from a single experiment. Performing this across the entire set of images, 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 primary visual cortex in each stimulus response to have a zero mean, and applied that transformation to the area within primary visual cortex. This method makes the assumption that the areas outside primary visual cortex 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 the [Experimental methods](#) section). We obtained a single image representing each trial by calculating

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

where  $r_i$  is the image frame  $i$ . The two terms are the mean of the final 5 stimulus frames and the mean of the 10 pre-stimulus frames, respectively. As there are small changes in the images due to respiration and blood-flow, taking the mean of a smaller number of frames produces a more reliable representation of the cortical response than taking the average of all 25 stimulus frames. Subtracting the mean of the pre-stimulus frames has the effect of removing the baseline activity from the images. We presented the animal with 32 directed stimuli covering the full  $360^\circ$  of direction stimulus space 40 times each (see the [Experimental methods](#) section). As we were interested in orientation preference and not in direction preference, we pooled the responses to parallel directions, for a total of 80 responses to each of 16 orientations in the range  $0^\circ$  to  $180^\circ$ .

#### Experimental methods

We performed optical intrinsic signal imaging of primary visual cortex in a single normally sighted juvenile cat (3.8 kg, female) reared from birth in an outdoor enclosure subject to the prevailing night-day cycle. All surgical and experimental procedures were approved by the institutional Animal Experimentation Ethics Committee 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.

#### *Anaesthesia and surgical procedures*

The animal was prepared for acute physiological recordings as described previously (van Kleef et al., 2010). Anaesthesia was induced by intramuscular injection of ketamine (20 mg/kg) and xylazine (1 mg/kg). The animal was then intubated and anaesthesia was maintained for the duration of the experiment by inhalation of gaseous halothane (0.5–1.0%) in a 60:40 mixture of N<sub>2</sub>O and O<sub>2</sub>. Eye movements were minimised by continuous intravenous infusion of gallamine triethiodide (10 mg/kg/h). The animal was mechanically ventilated to maintain end tidal CO<sub>2</sub> concentration between 3.5 and 4%. The animal received a constant intravenous infusion containing Hartmann's 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). The eyes were fitted with neutral power rigid gas-permeable contact lenses and drops (1% atropine; 10% phenylephrine) were administered daily to cause pupillary dilatation and to retract the nictitating membrane.

For imaging, a craniotomy (10 mm × 8 mm; A4 to P6, L1 to L9) was performed to expose 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.

At the conclusion of the experiment the animal was killed by intravenous injection of an overdose of barbiturate (sodium pentobarbital; Lethobarb, 150 mg/kg).

#### *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 (Ratzlaff and Grinvald, 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 the 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 grey screen.

#### *Visual stimuli*

Visual stimuli consisted of luminance defined oriented square wave gratings presented within a circular aperture on an isoluminant grey 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 (covering the central 60° of the visual field) drifting gratings of fixed spatial and temporal frequency (0.15 cycles per degree and 2 Hz respectively) drifting in one of 32 directions equi-spaced

between 0° and 360°. Each stimulus direction, together with a blank condition (no grating), was presented 40 times with the order of presentation randomised across trials to avoid any systematic bias in the observed responses.

## **Results**

The standard method for estimating orientation preference maps from imaging data is vector averaging, which involves averaging the experimental data together and subjective spatial filtering to deal with noise. Macke et al. (2011) recently proposed a map estimation method based on Gaussian process regression which substantially improves on the results of vector averaging. To investigate the effects of single orientation rearing on maps, we developed a generalised form of the Gaussian process method more appropriate for data from abnormally-reared animals, and applied it to three different sets of data: data from an experiment we performed, the data from Macke et al. (2011), and the data from Sengpiel et al. (1999) (see the **Materials and methods** section).

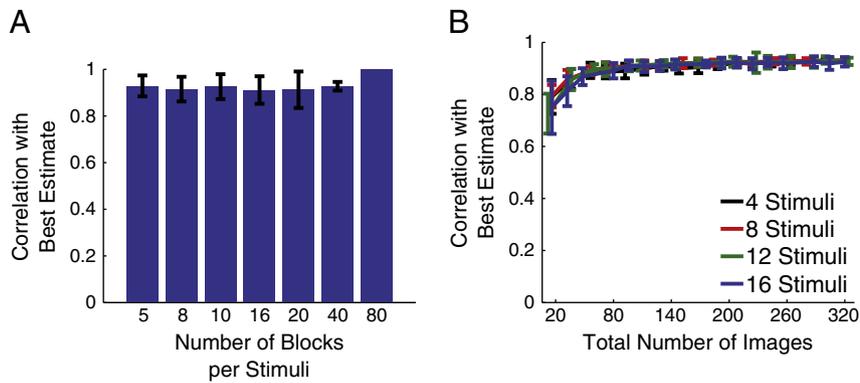
#### *The quality of averaged data and optimal parameters for data acquisition*

A feature of the Sengpiel et al. (1999) data was that the image sets were only available as block images rather than the individual experimental trial images (see the **Materials and methods and Data** sections). Therefore, we first confirmed that the use of this averaged data would not substantially affect the results of our Gaussian process analyses compared to data containing all individual images. To show this we used our own dataset which contained a large number of trials for many orientations.

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, as measured by the correlation between each map and the map estimated with all available data, induced by using averaged trials. Block data was formed from subsets of this dataset and used to estimate maps using the original Gaussian process method as described by Macke et al. (2011). For example, in the 5 blocks per stimuli case, the 80 trials for each stimuli were split into 5 groups, and the images in each group were averaged together to form a single image. The splits were made randomly a total of 40 times to obtain a distribution of map quality values. The results indicate that using averaged data does not make a substantial difference to the maps (Fig. 2A), at least in terms of overall map quality, and so the use of such data in further analyses is justified.

In addition, we investigated how best to design orientation preference map imaging experiments given a limited number of trials. What is the best 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 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 (Fig. 2B). 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 small numbers of trials (Kruskal–Wallis test,  $p < 0.05$  for total number of images less than 100,  $p > 0.05$  for larger totals). 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. However, for any sufficiently large number of trials (greater than approximately 100), the choice of trial dispersal among stimuli is of little consequence to the quality of the resulting map.

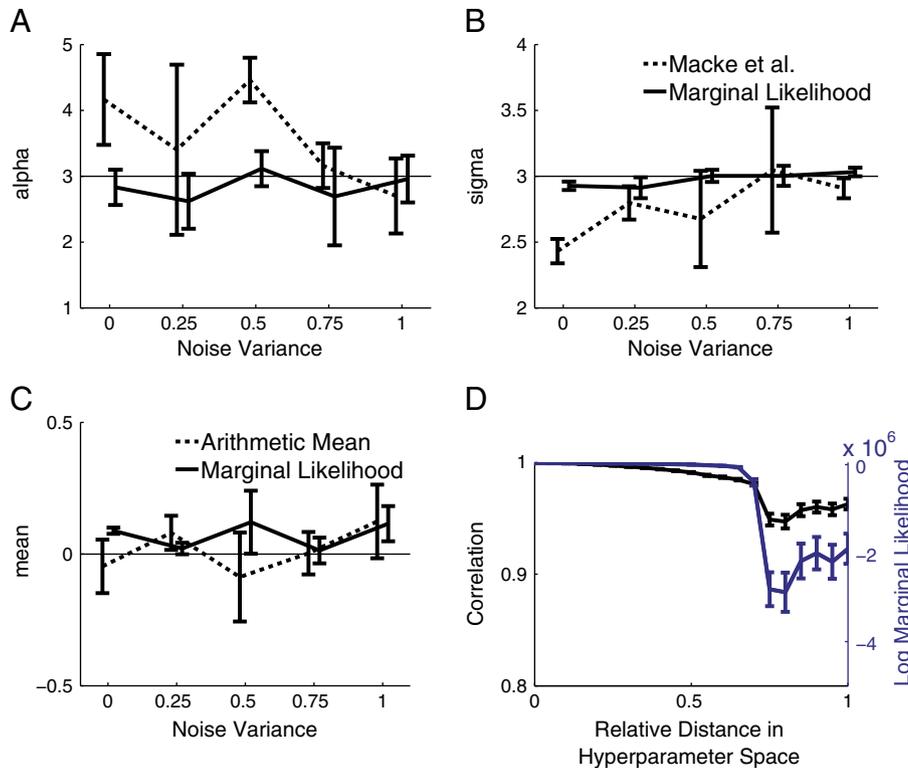


**Fig. 2.** Using averaged data does not reduce map quality substantially, nor does the dispersal of experimental trials among stimuli. (A) A comparison between map quality (measured as the correlation with the map generated using all of the data, 80 trials each of 16 stimuli) and amount of data used, 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  $\pm$  std dev. for 40 random samples. (B) A comparison between map quality (again measured as the correlation with the map generated using all of the data, 80 trials each of 16 stimuli) 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  $\pm$  std dev. is shown for 40 random subsets of images. There is a statistically significant difference in map quality between different numbers of stimuli only when using less than approximately 100 total images (Kruskal–Wallis test,  $p < 0.05$ ). Note that the right-most data points are less than one, as out of 1280 total trials, only at most 320 were used here.

### 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. For very noisy data, the autocorrelation function can also be very noisy, in which case filtering of the map is required. However, this introduces a subjective step into the estimation of hyperparameters. We therefore implemented a computationally expensive but more robust method for determining the hyperparameters, which involves finding



**Fig. 3.** Hyperparameters can be reliably inferred from the data by maximising the marginal likelihood. (A–B) Estimated hyperparameters obtained by maximising the marginal likelihood (solid) or using the method of Macke et al. (2011) (dashed), for maps with known parameters sampled from Gaussian process priors, as increasing degrees of noise are added (see the Materials and methods and Prior sampled maps sections). The horizontal lines show the true prior parameter values. The marginal likelihood estimates are of substantially higher accuracy and lower variance than those obtained by the simpler technique of fitting the prior covariance to the maps. (C) Estimates of the prior mean obtained by maximising the marginal likelihood (solid) or by directly calculating the arithmetic mean of the data (dashed), for prior sampled maps with known mean and added noise. The accuracy of the techniques is now comparable, with the marginal likelihood estimates having a generally lower variance. For simplicity, we only show the results for one set of parameters ( $\alpha = 3$ ,  $\sigma = 3$ ,  $\mu = 0$ ); other parameter sets showed similar results, as did the extension to the six parameter model (see the Results and Shifting covariance functions sections). The variance in the maps was approximately  $\alpha = 3$ , making the largest noise variance approximately one third of the map values. Data shown are mean  $\pm$  std dev. for 10 applications of the algorithms to 10 different maps, per noise variance. Note that some of the variance is due to variability between the prior sampled maps themselves. (D) The variability in map correlation (black) and marginal likelihood (blue) as a function of distance in hyperparameter space, for a prior sampled map. Map correlation compares with the map corresponding to the maximum marginal likelihood, and distance is Euclidean distance from the optimal parameters in hyperparameter space, divided by the optimal parameters to normalise distance across the different parameters. The peak of the marginal likelihood is relatively flat, and a large range of hyperparameters gives very similar posterior map estimates. Data shown are mean  $\pm$  std error.

the parameter set which maximises the marginal likelihood (Rasmussen and Williams, 2006) (see the Materials and methods and Marginal likelihood maximisation sections). This results in an inferred set of parameters which determine the prior which is most likely to have generated the data.

To validate this technique, we compared its chosen parameters to the true parameters of prior sampled maps (see the Materials and methods and Prior sampled maps sections). We found a close and relatively invariable match, even as noise was added to the data (Figs. 3A–C). By comparison, directly estimating the parameters from the data in the same way as Macke et al. (2011) produced far more variable and less accurate results (Figs. 3A–C).

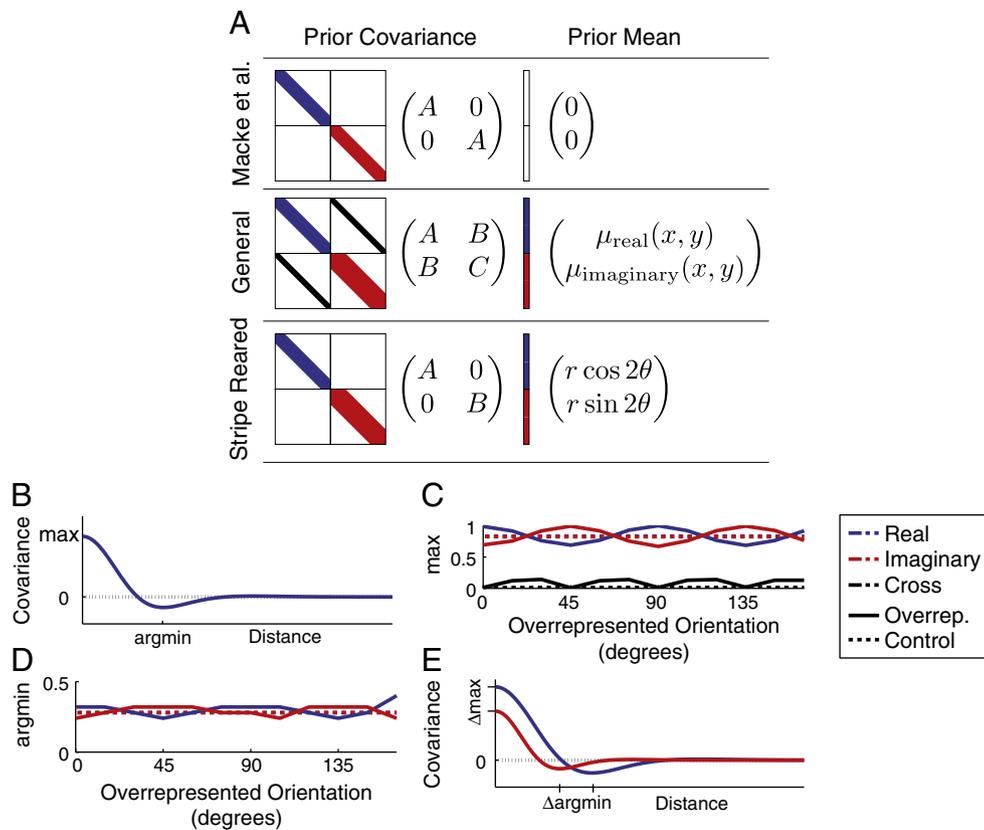
We found that the peaks of the marginal likelihood were usually relatively flat (Fig. 3D), indicating that a large range of parameter values fit the data well. It is indeed the case that a large range of hyperparameters produces very similar posterior map estimates (Fig. 3D). In most orientation preference experiments, a large number of images are captured relative to the number of data points one might have in a more typical regression problem, which reduces the accuracy requirements on the prior. However, this only applies to the case where one is primarily interested in deriving posterior map estimates, as in Macke et al. (2011), hence why their technique worked well for their purposes. When using the hyperparameters themselves to analyse maps, rather than

interrogating the estimated posterior maps, accurate parameter estimates become very important, especially as more parameters are added to the prior (see below).

Generalising the Gaussian process 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 matched, and also that the a priori mean at each pixel was zero (shown schematically in Fig. 4A, first row). These assumptions hold for normally-reared animals, but anisotropic rearing introduces correlations between the real and imaginary components, differences in their individual covariance functions, and a non-zero mean at each pixel (Fig. 4A, second row).

To investigate this empirically, we generated maps using the elastic net model with and without overrepresentation of a much larger number of orientation angles than available in experimental data (see the Materials and methods section). 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 (Figs. 4C–D). These results also show that it is the relative difference between the covariance functions of the two components



**Fig. 4.** The structure of the prior covariance and mean, and the nature of their changes in elastic net maps with overrepresentation. (A) The first row of the table shows the model used by Macke et al. (2011), with a zero mean and matched and independent real (blue) and imaginary (red) 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 (black). The final row shows a compromise between the first two, with a non-zero mean parameterised by the overrepresented angle  $\theta$  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) The maximum value of the autocovariance functions of the real (red) and imaginary (blue) components, and the maximum absolute value of the covariance between the two components (black), of elastic net maps, showing a dependence on the overrepresented orientation (one map for each overrepresented orientation; other maps showed similar results (data not shown)). Dotted lines show the same for maps without overrepresentation. (D) The distance at which the autocovariance is at a minimum for the same maps, again showing a dependence on the overrepresented orientation, and no change in control maps. Note that the argmin of the cross-correlation is not shown as it is a non-increasing function. These results inspired the generalised prior (panel A, second row). (E) Definitions of the terms  $\Delta\text{max}$  and  $\Delta\text{argmin}$  (Eqs. 4 and 5). Note that the chosen overrepresented orientations in panels C and D 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.

which is important, not their absolute values, leading to the definitions of  $\Delta_{\max}$  and  $\Delta_{\arg\min}$  (Eqs. 4 and 5; Fig. 4E). Intuitively, these measure the difference in scaling and wavelength of the real and imaginary components of maps, respectively. 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. Thus, a more general treatment than that of Macke et al. (2011) is required when investigating maps with possible overrepresentation.

### The analysis of stripe-reared maps

#### 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  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$ ) 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. However, this method only allows a granular analysis of the overrepresentation on a small set of predefined orientations, and artefacts and noise in the images may lead to large areas which are erroneously determined to prefer some orientation. Nonetheless, we calculated a measure of this type to allow for comparison with previous work. We used a scalar measure  $\rho$ , based on the difference between the number of pixels preferring the reared or maximally represented orientation and the other orientations (see the [Materials and methods and Simple overrepresentation quantification](#) sections). Using this measure there was a significant overrepresentation of the reared orientation in the stripe-reared maps in the Sengpiel et al. (1999) dataset compared to the control animals (Fig. 5A, points;  $p = 0.0002$ , Mann–Whitney  $U$  test,  $U = 121$ ), confirming the results of the original study. However, as expected, the values of this measure are not robust to changes in the filtering parameters used to smooth the data, as it only measures overrepresentation by counting individual pixels (Fig. 5A, error bars).

#### Shifting value distributions

Our investigations of the elastic net model predict that in an orientation preference map with overrepresentation, the value distributions of the two components of the map will 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, using the value distributions of the vector averaged maps from the Sengpiel et al. (1999) dataset. In control animals, the two map components had nearly identical distributions (Fig. 5B). However, in stripe-reared maps the mean of the component corresponding to the reared orientation had a positive shift (Fig. 5C), with the corresponding components of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  and  $135^\circ$  being positive real, positive imaginary, negative real and negative imaginary, respectively. We assumed that

this pattern holds for intermediate orientations, allowing the means of the value distributions to be parametrised in polar coordinates by

$$\begin{aligned}\mu_{\text{real}} &= r \cos 2\theta \\ \mu_{\text{imaginary}} &= r \sin 2\theta,\end{aligned}$$

where  $r$  and  $\theta$  are the degree and the orientation of overrepresentation, respectively. We chose this parameterisation as it allows for features of overrepresentation, that is the degree and orientation, to be explicitly represented.

To check that this is an appropriate parameterisation of the mean, we least-squares fit the means of the vector averaged map component distributions (Figs. 5B–C) to find a value for  $r$ , the degree of overrepresentation for each map (Fig. 5D). As expected, the mean value of  $r$  for the stripe-reared maps is significantly greater than that of the control maps ( $p = 0.028$ , Mann–Whitney  $U$  test,  $U = 63$ ), although there is a less clear separation between the two groups than with  $\rho$  (Fig. 5A). However, fitting the values of  $r$  and  $\theta$  to the value distributions in this way provides an improved technique for estimating overrepresentation compared with that used in Sengpiel et al. (1999), as here we are not limited to considering only a discrete set of orientations.

Adding these two parameters,  $r$  and  $\theta$ , as hyperparameters to the Gaussian process model allows us to infer them from the data by maximising the marginal likelihood (see the [Materials and methods and Marginal likelihood maximisation](#) sections). Performing this, with a prior otherwise the same as Macke et al.'s (2011), provides surprising results (Fig. 5E). 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  $\theta$  were, however, in good agreement with expectations, with all orientations being within approximately  $25^\circ$  of the reared orientation, with the exception of map C4 (see below), while having no parameters which were fit subjectively (Fig. 5F). This large degree of overlap in the distributions shows that just shifting value distributions is not enough to account for all of the overrepresentation in the maps.

The fitted parameters of map C4 indicated overrepresentation of the orientation orthogonal to that in which the animal was reared. Further investigation revealed that the region outside primary visual cortex in the map violated our normalisation assumption of (potentially biased) Gaussian noise (see the [Materials and methods and Data preprocessing](#) sections), thus not allowing us to properly normalise the responses between stimulus conditions. No other method of normalisation could be found which could ameliorate this issue. It is worth noting that the measure  $\rho$  only indicates an overrepresentation of the reared orientation when this map is filtered appropriately. An unfiltered vector averaged map of the C4 data has a negative  $\rho$  value, indicating that the orthogonal orientation is overrepresented. Map C3 contains data from the other

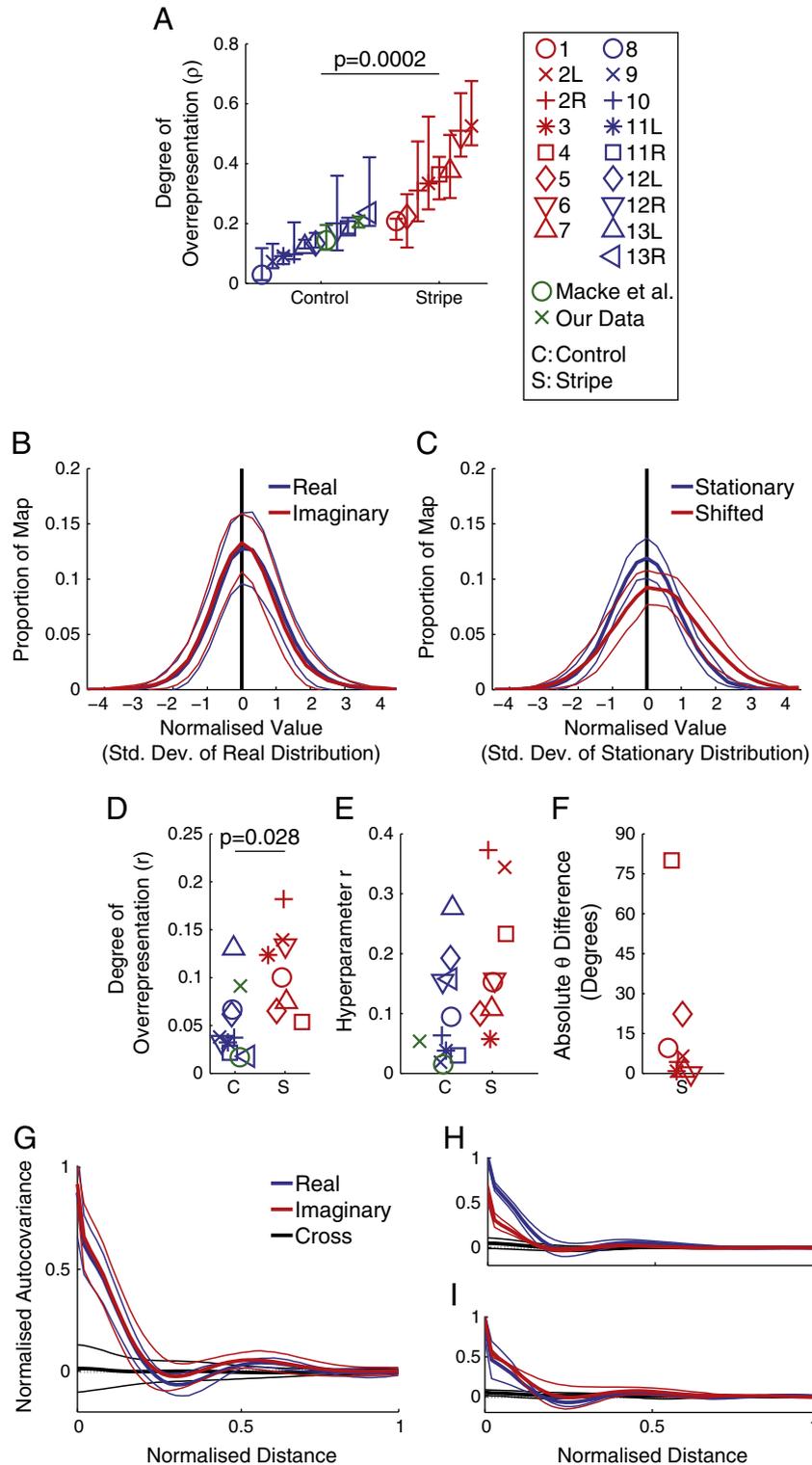
**Fig. 5.** Stripe-rearing affects the value distributions of maps and their covariance functions. (A) Values of the overrepresentation metric  $\rho$ , based on counting the number of pixels preferring each orientation (see the [Materials and methods](#) section), showing a significant difference between groups ( $p = 0.0002$ , Mann–Whitney  $U$  test,  $U = 121$ ). Error bars show the range of values  $\rho$  takes over a range of realistic filtering parameters ( $\sigma_{\text{lowpass}}$  and  $\sigma_{\text{highpass}}$  ( $\mu\text{m}$ ): 150–225 and 275–350 for our data, 50–125 and 175–250 for data from Macke et al. (2011), and 25–100 and 150–225 for data from Sengpiel et al. (1999)); based on the estimated wavelengths of the maps), showing that it is not a very robust measure. Symbols show the value for the median of the range of filtering parameters given, which was used to conduct the statistical test. Data points are ordered by their  $\rho$  value within groups. (B) The normalised value distributions of the real and imaginary components of all normally-reared maps ( $n = 5$  cats, 9 hemispheres), with the thick and thin lines showing mean  $\pm$  std dev. The 95% confidence interval on  $\mu_{\text{imaginary}} - \mu_{\text{real}}$  is  $[-0.012, -0.009]$  (Student's  $t$ -test,  $t = -13.3$ ) which is close to zero as expected, but shows a small difference in the means due to noise. (C) 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  $\mu_{\text{shifted}} - \mu_{\text{stationary}}$  is  $[0.050, 0.054]$  (Student's  $t$ -test,  $t = 46.2$ ), almost five times as large as the difference in the normally-reared maps. (D) Values of  $r$  from least-squares fitting the mean parameterisation to the vector averaged maps' value distributions, showing a generally larger value of  $r$  for the stripe-reared maps than the control maps ( $p = 0.028$ , Mann–Whitney  $U$  test,  $U = 63$ ). (E) Marginal likelihood fitted values of  $r$ , showing generally larger value for stripe-reared maps, but with a less clear distinction between groups than in panel D. (F) The absolute difference between the reared orientation and the fitted value of  $\theta$ , showing that the most overrepresented orientations are close to the reared orientations, with the exception of a single map (discussed in the [Results and Shifting value distributions](#) sections). (G) 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 B. (H–I) The normalised covariance functions of (H) all  $0^\circ$  and  $90^\circ$  stripe-reared maps ( $n = 3$  cats, 6 hemispheres) and (I) all  $45^\circ$  and  $135^\circ$  stripe-reared maps ( $n = 2$  cats, 2 hemispheres). In the  $0^\circ$  and  $90^\circ$  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  $45^\circ$  and  $135^\circ$  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 Fig. 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.

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.

*Shifting covariance functions*

Another possible source of the overrepresentation not accounted for by the shift in value distribution means is change in 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 (Fig. 5G), as expected, but showed a slight difference in both magnitude and wavelength dependent on the reared orientation in the stripe-reared cases, in line with the predictions made by the elastic net maps (Figs. 5H–I). There was no substantial correlation between the two map components, which again is consistent with the elastic net maps, since for Cartesian axis aligned orientations (i.e., 0°, 45°, 90° and 135°, the only rearing orientations examined in Sengpiel et al. (1999)), the predicted correlation between components is zero (Fig. 4C).



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:  $\alpha_{\text{real}}$ ,  $\sigma_{\text{real}}$ ,  $\alpha_{\text{imaginary}}$  and  $\sigma_{\text{imaginary}}$ , which allow the independent specification of the maximum and argument of the minimum of the covariance functions of both components (Fig. 4B). Following the results from the elastic net maps (Figs. 4C–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  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$  (Eqs. 4 and 5; Fig. 4E). These measures represent the difference in scaling and wavelength between the real and imaginary components of the maps, respectively.

Inferring all six hyperparameters of the model from the data by maximising the marginal likelihood, and calculating the measures  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$ , provided a clearer picture of the differences between control and stripe-reared maps (Figs. 6A–B). Control maps showed generally small values of  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$ , and stripe-reared maps generally larger values, but interestingly this was not the case for all maps, as with the hyperparameter  $r$  (Fig. 5E). These results indicate that stripe-rearing affects multiple aspects of the structure of orientation maps, and does so to varying degrees in different animals, and that the structural changes in maps which induce overrepresentation are indeed as predicted by the elastic net model. Interestingly, all stripe-reared maps had a high value of at least one of the measures ( $r$ ,  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$ ), motivating us to investigate whether some combination of these measures may give the best indication of overall overrepresentation.

#### Prediction and classification

We therefore combined the values  $r$ ,  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$  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 Gaussian process priors with a range of values of  $r$ ,  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$  (see the [Materials and methods and Prior sampled maps](#) sections). Calculating  $\rho$  for these maps, and then solving the linear system

$$a\Delta_{\text{max}} + b\Delta_{\text{argmin}} + cr + d = \rho$$

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 chose to use  $\rho$  as a measure of overrepresentation here, as while it is still a granular measure, on Gaussian process posterior maps the problems with its dependency on filtering parameters and noise are alleviated. We used this model to predict overrepresentation in the maps, which resulted in a significant grouping of the maps by rearing condition, as expected (Fig. 6C;  $p = 0.011$ , Mann–Whitney  $U$  test,  $U = 98$ ). These results can be compared to those from Sengpiel et al. (1999) (Fig. 5A), and to the least-squares fitting of  $r$  (Fig. 5D), with both of which there is a good match. This result demonstrates that even though we have broken down the structural causes of overrepresentation into separate measures, we are still overall capturing the majority of the overrepresentation present in the maps.

The fitted hyperparameters can also be used to classify maps by rearing condition without relying on the measure  $\rho$  by training a linear support vector machine (SVM) on the hyperparameters. We trained and tested the SVMs on maps sampled from prior Gaussian processes with varying parameters (96 control and 96 with overrepresentation; see the [Materials and methods and Prior sampled maps](#) sections), using either  $\rho$  or the three parameters  $r$ ,  $\Delta_{\text{max}}$  and  $\Delta_{\text{argmin}}$ , which were estimated from the maps (using the empirical covariance functions and means, as the maps were noise-free). The SVMs were trained on a random half of the maps, and tested on the other half, and this was repeated 100 times. The classifier using the three parameters performed

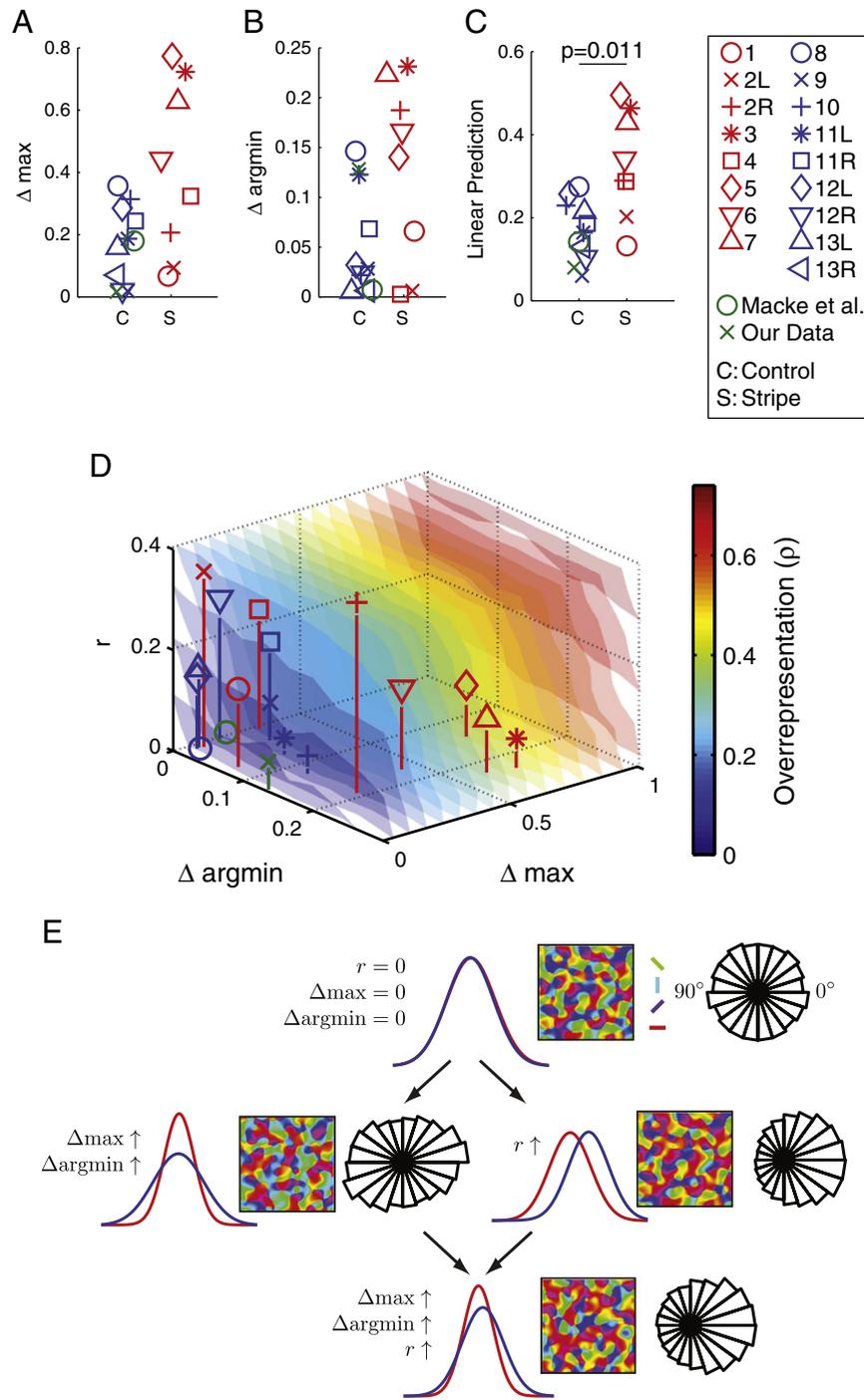
consistently slightly better, correctly classifying 98% of the maps on average, while the  $\rho$  classifier averaged a 95% accuracy level. The distributions of accuracies were significantly different ( $p = 7 \times 10^{-23}$ , Mann–Whitney  $U$  test,  $U = 60735$ ). We repeated this process on maps generated using the elastic net (96 control and 96 with overrepresentation), which showed a much larger difference in the results of the two classifiers (three parameter classifier mean accuracy: 99%,  $\rho$  mean accuracy: 82%,  $p = 5 \times 10^{-35}$ , Mann–Whitney  $U$  test,  $U = 5050$ ). The nature of the elastic net model causes it to generate maps with more variability when compared with prior sampled maps, in the face of which the three parameter classifier performs significantly better. Intuitively this is expected, as  $\rho$  simply compares counts of pixels and is quite susceptible to minor variability in the maps (of which noise is one example), while our three parameters measure overall differences in the maps, which provide a more robust assessment of the changes which lead to overrepresentation, rather than measuring it directly. When trained on the maps from Sengpiel et al. (1999) (cross-validating by leaving out a single map on each training iteration, because of the much smaller dataset than above), the three parameter SVM could correctly classify 14/17 of the maps (2 stripe-reared and 1 control map misclassified), and an SVM trained on the values of  $\rho$  could also correctly classify 14/17 maps (1 stripe-reared map and 2 control maps misclassified). This shared classification rate of approximately 82% falls within the range expected after the analysis of computational maps above, the differences being accounted for by the very noisy nature of the empirical data.

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  $\rho$  for these maps increased linearly, at different rates, for each of the hyperparameters, as predicted by the linear relationship investigated above (Fig. 6D). To investigate why some of the maps lie on different regions of the same iso- $\rho$  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- $\rho$  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 (Fig. 6E). An increase in  $r$  shifts the distributions without affecting their shape, while changing  $\Delta_{\text{argmin}}$  or  $\Delta_{\text{max}}$  changes their shapes without affecting their means. We can get an overall estimate of the level of overrepresentation through our linear model, or by classifying maps with a SVM, but by looking at the different components in this way we can see how the aspects combine to form the final level of overrepresentation. Examining the distributions of orientation and selectivity in the maps it is apparent that different combinations of the hyperparameters create maps with subtle differences in the distributions, as well as spatial structure.

## 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 not optimised to address the effects of rearing in an abnormal visual environment, which is an important method for investigating brain plasticity. Here, we have generalised the method, including introducing a prior which allows some orientations to be overrepresented. Though the



**Fig. 6.** Fitted overrepresentation parameters and map classification and overrepresentation prediction using them. (A) The relative absolute difference in maximum of marginal likelihood fitted covariance functions,  $\Delta \max$ . (B) The relative absolute difference in argmin of marginal likelihood fitted covariance functions,  $\Delta \arg \min$ . (C) Results of applying the linear model to the maps, showing a significant split in the data by rearing condition ( $p = 0.011$ , Mann–Whitney  $U$  test,  $U = 98$ ), a good match with the results in Fig. 5A. Symbols show which map each point represents, as per legend. (D) Distribution of  $\rho$  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  $\rho$  increases linearly with each hyperparameter, as expected, but the values of each map sit at different locations on the planes on constant  $\rho$ , 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  $\Delta \max - \Delta \arg \min$  plane. (E) Changes in prior sampled maps induced 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  $\Delta \max$  or  $\Delta \arg \min$ , with a change in the variance of the components, and overrepresentation of  $0^\circ$  and  $90^\circ$ . (Right) A map with increased  $r$ , showing shifted distributions, and overrepresentation of  $0^\circ$ . (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 Fig. 5C, and note that the map has a subtle but noticeable level of overrepresentation of  $0^\circ$  (red 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.

maps estimated with this generalised method are not distinguishable by eye from those estimated with Macke et al.'s (2011) prior, analysis of the fitted parameters of the prior 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, as well as the overall structure of these changes. There are, in principle,

many ways one could structurally change an orientation preference map to bias its overall preference for a particular orientation. Our demonstration of the particular nature of these changes, that is systematic and rearing-orientation-dependent changes to the means and covariance functions, therefore provides an important contribution to our knowledge of the effects of visual input on the structure of visual cortical maps.

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, essentially equivalent to our  $\rho$  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 is 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, and represent the relative absolute difference between the wavelengths of the real and imaginary components ( $\Delta_{\text{argmin}}$ ), the relative absolute difference between the scaling of these two components ( $\Delta_{\text{max}}$ ), and the difference in the mean of the two components ( $r$ ). Fig. 6E illustrates with simulations how varying each of these parameters independently affects map structure. 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 can provide a reliable prediction of rearing condition.

The greater number of parameters in our model of over-representation compared to the classical analysis technique makes for a mathematically more complex model, which in general can be vulnerable to overfitting the data. For determining whether or not overrepresentation has occurred in a map, a single parameter such as  $\rho$ , which trades robustness for simplicity, may suffice. However, as both our elastic net results (Fig. 4) and fitted parameters (Figs. 5E and 6A–B) show, to investigate the nature of the environmentally-induced changes in maps, three parameters are required.

The diversity in parameter values between animals in the same rearing condition is intriguing (Figs. 5E and 6A–B), 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 visual experience 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 primary visual cortex are more similar than those of unrelated neurons (Li et al., 2012; Ohtsuki 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 of 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)).

Our initial guide for the changes we would expect to see in stripe-reared maps was the elastic net model, and the orientation preference maps containing overrepresentation it produced. We found that in those maps, both the means and covariance functions of the two map components changed, which we went on to find in the empirical data from Sengpiel et al. (1999) as well. Changes in either the means or the covariance functions are enough to induce overrepresentation by themselves, so it is quite remarkable that the elastic net model shows changes in both, as in the experimental maps. This adds to the list of map statistics and phenomena the model has previously been shown to produce (Carreira-Perpiñán et al., 2005), and reinforces its place as a useful model for investigating visual cortical maps. These results add these structural changes under abnormal rearing to the set of phenomena that models should be able to reproduce. Whether or not other models of map development produce these specific effects is currently unknown.

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 ( $\alpha$  and  $\sigma$ ) 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. While our method still relies on assumptions regarding the spatial form of the maps and the nature of the noise present in the data, these assumptions are clearly stated, which is preferable to the subtle and subjective assumptions made by techniques such as vector averaging.

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.

## Acknowledgments

We thank Chris Williams for helpful discussions. This research was supported by the National Health and Medical Research Council grant 525459, the Australian Research Council through the Centre of Excellence in Vision Science (CE0561903) and the Biotechnology and Biological Sciences Research Council grant BB/J002089/1. The authors declare no competing financial interests.

## References

- Aitkin, L.M., Merzenich, M.M., Irvine, D.R., Clarey, J.C., Nelson, J.E., 1986. Frequency representation in auditory cortex of the common marmoset (*Callithrix jacchus jacchus*). *J. Comp. Neurol.* 252, 175–185.
- Anderson, P.A., Olavarria, J., Van Sluyters, R.C., 1988. The overall pattern of ocular dominance bands in cat visual cortex. *J. Neurosci.* 8, 2183–2200.
- Astic, L., Saucier, D., Holley, A., 1987. Topographical relationships between olfactory receptor cells and glomerular foci in the rat olfactory bulb. *Brain Res.* 424, 144–152.
- Blakemore, C., Cooper, G.F., 1970. Development in the brain depends on the visual environment. *Nature* 228, 477–478.
- Blasdel, G., Salama, G., 1986. Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature* 321, 579–585.
- Bonhoeffer, T., Kim, D.S., Maloney, D., Shoham, D., Grinvald, A., 1995. 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, 1973–1988.
- Cardoso, M.F., Salcedo, R.L., Feyo de Azevedo, S., 1996. The simplex-simulated annealing approach to continuous non-linear optimization. *Comput. Chem. Eng.* 20, 1065–1080.
- Carreira-Perpiñán, M.A., Goodhill, G.J., 2004. Influence of lateral connections on the structure of cortical maps. *J. Neurophysiol.* 92, 2947–2959.
- Carreira-Perpiñán, M.A., Lister, R.J., Goodhill, G.J., 2005. A computational model for the development of multiple maps in primary visual cortex. *Cereb. Cortex* 15, 1222–1233.
- Chen, X., Gabitto, M., Peng, Y., Ryba, N.J.P., Zuker, C.S., 2011. A gustotopic map of taste qualities in the mammalian brain. *Science* 333, 1262–1266.
- Crair, M.C., Ruthazer, E.S., Gillespie, D.C., Stryker, M.P., 1997. Relationship between the ocular dominance and orientation maps in visual cortex of monocularly deprived cats. *Neuron* 19, 307–318.
- Espinosa, J.S., Stryker, M.P., 2012. Development and plasticity of the primary visual cortex. *Neuron* 75, 230–249.
- Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: theory. *NeuroImage* 16, 465–483.
- Hirsch, H.V., Spinelli, D.N., 1970. Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science* 168, 869–871.
- Hübener, M., Shoham, D., Grinvald, A., Bonhoeffer, T., 1997. Spatial relationships among three columnar systems in cat area 17. *J. Neurosci.* 17, 9270–9284.
- Hunt, J.J., Giacomantonio, C.E., Tang, H., Mortimer, D., Jaffer, S., Vorobyov, V., Erickson, G., Sengpiel, F., Goodhill, G.J., 2009. Natural scene statistics and the structure of orientation maps in the visual cortex. *NeuroImage* 47, 157–172.
- Hunter, D.J., 2005. Gene–environment interactions in human diseases. *Nat. Rev. Genet.* 6, 287–298.
- Issa, N.P., Trepel, C., Stryker, M.P., 2000. Spatial frequency maps in cat visual cortex. *J. Neurosci.* 20, 8504–8514.
- Kaschube, M., Wolf, F., Geisel, T., Löwel, S., 2002. Genetic influence on quantitative features of neocortical architecture. *J. Neurosci.* 22, 7206–7217.
- Li, Y., Lu, H., Cheng, P., Ge, S., Xu, H., Shi, S.H., Dan, Y., 2012. Clonally related visual cortical neurons show similar stimulus feature selectivity. *Nature* 486, 118–121.
- Macke, J.H., Gerwinn, S., White, L.E., Kaschube, M., Bethge, M., 2011. Gaussian process methods for estimating cortical maps. *NeuroImage* 56, 570–581.
- Ohtsuki, G., Nishiyama, M., Yoshida, T., Murakami, T., Histed, M., Lois, C., Ohki, K., 2012. Similarity of visual selectivity among clonally related neurons in visual cortex. *Neuron* 75, 65–72.
- Payne, B.R., Peters, A., 2001. The concept of cat primary visual cortex. *The Cat Primary Visual Cortex*. Academic Press.
- Rasmussen, C.E., Williams, C.K.I., 2006. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, MA.
- Ratzlaff, E.H., Grinvald, A., 1991. A Tandem-Lens Epifluorescence Macroscope – Hundred-Fold Brightness Advantage For Wide-Field Imaging. *J. Neurosci. Methods* 36, 127–137.
- Roger, A.S., Schwartz, E.L., 1990. Cat and monkey cortical columnar patterns modeled by bandpass-filtered 2D white noise. *Biol. Cybern.* 62, 381–391.
- Sengpiel, F., Stawinski, P., Bonhoeffer, T., 1999. Influence of experience on orientation maps in cat visual cortex. *Nat. Neurosci.* 2, 727–732.
- Shatz, C.J., Stryker, M.P., 1978. Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *J. Physiol.* 281, 267–283.
- Shmuel, A., Grinvald, A., 1996. Functional organization for direction of motion and its relationship to orientation maps in cat area 18. *J. Neurosci.* 16, 6945–6964.
- Tanaka, S., Ribot, J., Imamura, K., Tani, T., 2006. Orientation-restricted continuous visual exposure induces marked reorganization of orientation maps in early life. *NeuroImage* 30, 462–477.
- Tomita, K., Sperling, M., Cambridge, S.B., Bonhoeffer, T., Hübener, M., 2002. A molecular correlate of ocular dominance columns in the developing mammalian visual cortex. *Cereb. Cortex* 23 (11), 2531–2541.
- van Kleef, J.P., Cloherty, S.L., Ibbotson, M.R., 2010. Complex cell receptive fields: evidence for a hierarchical mechanism. *J. Physiol.* 588, 3457–3470.
- Weliky, M., Bosking, W.H., Fitzpatrick, D., 1996. A systematic map of direction preference in primary visual cortex. *Nature* 379, 725–728.
- Woolrich, M.W., 2012. Bayesian inference in fMRI. *NeuroImage* 62, 801–810.