Neurobiology of Learning and Memory 136 (2016) 147–165

Contents lists available at ScienceDirect

Neurobiology of Learning and Memory

journal homepage: www.elsevier.com/locate/ynlme

Neural network model develops border ownership representation through visually guided learning

Akihiro Eguchi *, Simon M. Stringer

Oxford Centre for Theoretical Neuroscience and Artificial Intelligence, Department of Experimental Psychology, University of Oxford, Oxford, UK

ARTICLE INFO

Article history:
Received 22 June 2016
Revised 7 September 2016
Accepted 7 October 2016
Available online 13 October 2016

Keywords:
Primate vision
Border ownership
Neural network model

ABSTRACT

As Rubin's famous vase demonstrates, our visual perception tends to assign luminance contrast borders to one or other of the adjacent image regions. Experimental evidence for the neuronal coding of such border-ownership in the primate visual system has been reported in neurophysiology. We have investigated exactly how such neural circuits may develop through visually-guided learning. More specifically, we have investigated through computer simulation how top-down connections may play a fundamental role in the development of border ownership representations in the early cortical visual layers V1/V2. Our model consists of a hierarchy of competitive neuronal layers, with both bottom-up and top-down synaptic connections between successive layers, and the synaptic connections are self-organised by a biologically plausible, temporal trace learning rule during training on differently shaped visual objects. The simulations reported in this paper have demonstrated that top-down connections may help to guide competitive learning in lower layers, thus driving the formation of lower level (border ownership) visual representations in V1/V2 that are modulated by higher level (object boundary element) representations in V4. Lastly we investigate the limitations of our model in the more general situation where multiple objects are presented to the network simultaneously.

1. Introduction

As Rubin's famous vase (Fig. 1) demonstrates, our visual perception tends to assign luminance contrast borders to one or other of the adjacent image regions, as if they serve as occluding contours (von der Heydt, Zhou, & Friedman, 2003). This is an example of *feature binding* in vision, in this case binding a luminance contrast border to a particular object. Representing such binding relationships between visual features is essential to the ability of the visual system to interpret and *make sense* of complex visual scenes. Experimental evidence for the neuronal coding of such border-ownership in the primate visual system has arisen in a neurophysiology study carried out by Zhou, Friedman, and von der Heydt (2000).

Zhou et al. (2000) have shown that the responses of simple cells in earlier cortical stages of visual processing such as V1 and V2, which respond preferentially to oriented edges, are also modulated by which side of an object or figure the edge occurs on. This is the case even when the figure/background cues lie well outside the classical receptive field of the neuron, which in area V1 is approximately 1 degree in size. Such neurons are referred to as *border ownership cells*. Sugihara, Qiu, and von der Heydt (2011) later reported that the border ownership signal emerges with a latency of 61 ms, which is about 13 ms later than the onset of orientation selectivity. This suggests that the global image context specifying border ownership modulates the activity of these neurons. In other words, there must be a mechanism that enables the contextual information to be conveyed to these early stage visual neurons in V1 and V2. It has been proposed that these kinds of border ownership responses in area V1 represent a form of feature binding, and so may be important for understanding how primate vision may solve the problem of feature binding more generally.

Some theoreticians have suggested that the context integration required for border ownership representations in V1 and V2 can be achieved via lateral propagation of signals within a layer via horizontal fibres (Baek & Sajda, 2005; Nishimura & Sakai, 2004; Zhaoping, 2005). However, Sugihara et al. (2011) have argued that the conduction velocity of horizontal fibres is too slow (most of them being between 0.1 and 0.4 m/s (Angelucci & Bullier, 2003)) to produce the border ownership signals within the short latency observed in neurophysiology studies. Furthermore, Sugihara et al. (2011) showed that varying the distance between the target border and the visual features that carry contextual information about the 'owner' of the border does not in fact influence the latency before

* Corresponding author.
E-mail address: akihiro.eguchi@psy.ox.ac.uk (A. Eguchi).

http://dx.doi.org/10.1016/j.nlm.2016.10.007
1074-7427/© 2016 The Author(s). Published by Elsevier Inc.

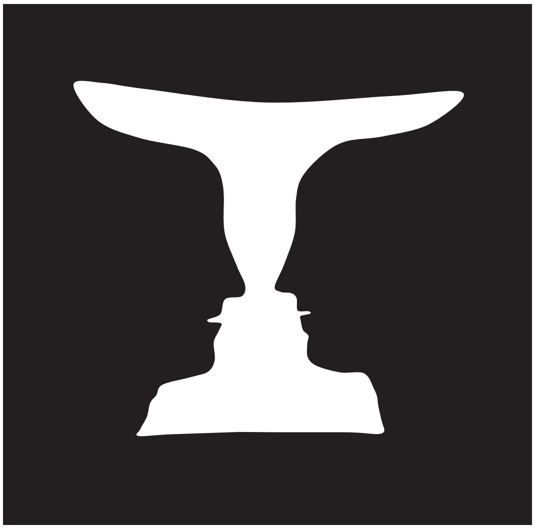

Fig. 1. Rubin's Vase (Rubin, 1915).

the border ownership signals arise. Therefore, they concluded that context influence by horizontal signal propagation alone is highly unlikely.

On the other hand, the feedforward (bottom-up) and feedback (top-down) connections between successive visual stages have fast-conducting axons, with conduction velocities of between 2 and 6 m/s, which is about ten times faster than cortical horizontal fibres (Angelucci & Bullier, 2003). Accordingly, both Craft, Schtze, Niebur, and von der Heydt (2007) and Jehee, Lamme, and Roelfsema (2007) have proposed models that involve hypothetical 'grouping circuits' within a higher cortical layer that capture the contextual information about local boundary elements, and these contextual signals are then relayed down through feedback connections to modulate responses in an earlier layer. They proposed that the larger receptive fields in the higher layer allow the network to employ 'grouping circuits' without having to rely on slow lateral propagation of signals. Nevertheless, it still remains a challenge to understand exactly how such neural circuits may be learned. The objective of the current study is to investigate the learning mechanisms that underpin the development of border ownership cells in the primate visual brain, in terms of synaptic modification guided by visual experience and consequent neural adaptation throughout a hierarchy of cortical stages. Moreover, given the proposed role of border ownership cells in feature binding, which is essential for integrating the visual features within a scene, the simulations described below provide a step towards understanding how the brain learns to make sense of the visual world.

One higher visual area that might provide appropriate top-down modulatory signals is V4, which contains neurons that represent the localised boundary contour elements of objects (Layton, Mingolla, & Yazdanbakhsh, 2012). The responses of these neurons are sensitive to both the shape of the boundary element and where the element is with respect to the centre of mass of the object (Pasupathy & Connor, 2001; Pasupathy & Connor, 2002). Hence each of the neurons encodes that a specific border element belongs to a particular object - i.e. a kind of border ownership representation. A subpopulation of these neurons will provide a distributed representation of the entire boundary of the object. Furthermore, the neurons are able to respond invariantly as the object is shifted across different locations on the retina over a modest range.

The visually-guided development of such V4 cells has been previously investigated in a computational modelling study with an established neural network model, VisNet, of the primate ventral visual pathway (Eguchi, Mender, Evans, Humphreys, & Stringer, 2015). The network architecture consisted of a hierarchy of cortical visual layers, with each layer modelled as a competitive neural network (Wallis & Rolls, 1997). Whenever an image was presented to the network, visual signals propagated through feedforward plastic synaptic connections between successive layers. Within each competitive layer, the excitatory cells competed with each other to respond to the current visual stimulus. In the brain, competition between excitatory cells is implemented via inhibitory interneurons. Although to save computational expense in VisNet, competition between excitatory neurons is modelled more directly using local filters. During an initial period of training with visual objects, the feedforward synaptic connections between successive layers of the network are continually modified using local, biologically plausible, associative learning rules. The competition within each layer then forces individual neurons to learn to respond selectively to a particular stimulus class, with different neurons responding to different kinds of stimulus. Competitive learning is a very simple unsupervised learning paradigm that allows neurons to discover important features of the stimulus input patterns (Rumelhart & Zipser, 1985). Eguchi et al. (2015) showed that the gradual increase in the receptive field size of neurons through successive layers of the visual system (Gross, Bender, & Rocha-Miranda, 1969; Pettet & Gilbert, 1992) allows V4 neurons access to local image information specifying how localised luminance contrast contours belong to adjacent object regions. As a result, cells in the higher layer of their hierarchical competitive neural network model developed neuronal response properties similar to those reported by Pasupathy and Connor (2001, 2002) when the model was trained on a number of real world objects.

In this paper, we extend the previous purely feedforward model of Eguchi et al. (2015) by incorporating both feedforward (bottom-up) and feedback (top-down) connections. This extended model architecture is used to investigate how the edge-detecting simple cells in the earliest layer of the network, which corresponds to visual areas V1/V2 in the primate brain, may develop border ownership representations via top-down modulation from neurons in the output layer, which corresponds to visual area V4. The necessary feedforward and feedback synaptic connectivity within the network is set up by visually-guided learning using a biologically plausible, local, trace learning rule (Foldiak, 1991) as the network is trained on a collection of differently shaped visual object stimuli. We go on to show how these border ownership signals in the earliest layer evolve dynamically during the 300 ms time course of a stimulus presentation, as reported by Sugihara et al. (2011) and Jehee et al. (2007). We then investigate the limitations of the model in the more general situation where multiple objects are presented to the network simultaneously.

1.1. Hypothesis

Eguchi et al. (2015) have shown that when an established hierarchical neural network model of the primate ventral visual pathway, VisNet (Wallis & Rolls, 1997), is trained on 177 images of real world objects, which rotated in plane through 360° and shifted across a 3 × 3 grid of nine different retinal locations, the neurons in the higher layers of the model learn to represent local boundary contour elements. Individual neurons are tuned to boundary elements with a specific curvature at a particular location with respect to the centre of mass of the object. Moreover, the neurons respond invariantly as an object is translated across different retinal locations. These are the same neuronal response properties as observed in area V4 of the primate visual system by Pasupathy and Connor (2002). Although they have reported that the translation invariant responses of V4 neurons are only over a modest range, we can simply suppose that the size of simulated retina in the model matches to the covered range.

The version of the VisNet architecture used in the previous study incorporated only feedforward (bottom-up) connections between successive layers of the network (Eguchi et al., 2015). No feedback (top-down) connections were included in the model even though these are known to exist in the primate ventral visual pathway. It has previously been suggested that the top-down connections might implement attention to objects during visual search (Deco & Lee, 2002; Wagatsuma, Oki, & Sakai, 2013) and were incorporated into a variant of VisNet model to simulate top-down biasing effects (Deco & Rolls, 2004). However, in this previous study the top-down connections were only implemented after training, and so did not influence the visual representations that developed during visually-guided learning. In contrast, in our present paper the top-down connections are also present during training, and thus play a key role in the development of border ownership representations in the early layers. In particular, we propose that the global image context specifying border ownership is conveyed to the early stage visual neurons by top-down connections between layers in order to drive the development of border ownership cells in the early cortical areas as reported by Zhou et al. (2000).

Accordingly, we hypothesised that learning in the extended VisNet architecture introduced in this paper would operate as follows. First, during visually-guided learning in which VisNet is trained on images of differently shaped objects, neurons in the later stages of visual processing such as V4 will learn to encode boundary contour elements through learning in the feedforward connections as previously demonstrated by Eguchi et al. (2015). Next, with continued visually-guided training on the same object images, we expect that strong polysynaptic feedback connections may subsequently develop from those neurons in the later stages of visual processing to neurons in earlier stages such as V1 and V2. These strengthened top-down connections might then modulate the responses of neurons in V1 and V2 according to where their preferred edge element occurs within an object.

More precisely, let us consider a subset Φ^{V4}_{Left} of neurons in V4 that have learned, by the visually-guided competitive learning mechanisms, to encode a vertical straight contour on the left of an object across different retinal locations. This subset of V4 neurons may also develop strengthened top-down polysynaptic connections to a subset of simple cells in V1 and V2 that originally signal the presence of any vertical straight contour within their small classical receptive field. This will force the subset of V1/V2 neurons to preferentially respond when the vertical straight contour is part of the left boundary of an object (top-down signals) at a particular retinal location (bottom-up signals).

Fig. 2(a) shows a case example in which an object with a straight vertical border on its *left* is presented with this border positioned at retinal location *1*. The figure illustrates how the subset Φ^{V4}_{Left} of V4 neurons, which represent a vertical straight edge on the left of an object, may modulate the responses of a subset of V1/V2 simple cells $\Phi^{V1/V2}_{Left,Loc1}$ that represent the presence of a vertical contour at retinal location 1. Fig. 2(b)–(d) shows similar case examples in which the vertical straight edge may occur on either the left or right boundary of the object, with the vertical straight edge positioned in either retinal location 1 or location 2.

In summary, we hypothesise that the observations of Zhou et al. (2000), in which the responses of V1 and V2 neurons are modulated by which side of a figure the edge occurs on, may be replicated by incorporating both bottom-up and top-down associatively modifiable connections within VisNet. This will allow neurons in the early layers to develop their firing responses through visually-guided competitive learning driven by a combination of both bottom-up and top-down visual signals. The neural circuits developed after visually-guided learning in VisNet are expected to be similar to the hypothetical 'grouping circuits' proposed in a previous modelling study of border ownership representation with top-down connections carried out by Craft et al. (2007). However, the focus of our current study is to investigate exactly how such neural circuits may be learned when the network is trained on visual images of differently shaped objects.

2. Materials & methods

2.1. VisNet model

The simulation studies presented in this paper are conducted with a modified version of an established neural network model, VisNet, of the primate ventral visual pathway, which was originally developed by Wallis and Rolls (1997). The original feedforward (bottom-up) version of the network architecture is shown in Fig. 3(a) and (c). It is based on the following: (i) a series of hierarchical competitive networks with local graded lateral inhibition; (ii) convergent feedforward connections to each neuron from a topologically corresponding region of the preceding layer, leading to an increase in the receptive field size of neurons through the visual processing areas; and (iii) synaptic plasticity based on a local associative trace learning rule (6) and (7), which is explained below. The hierarchical series of 4 neuronal layers of VisNet have been loosely related to the following successive stages of processing in the ventral visual pathway: V2, V4, the posterior inferior temporal cortex (TEO), and the anterior inferior temporal cortex (TE) (see Rolls (2012) for a comprehensive review of studies performed using VisNet). In the current simulations reported below, the number of the layers has been reduced to three since a large number of border ownership neurons were found to develop in the third layer of VisNet, which corresponds to TEO in the earlier study (Eguchi et al., 2015).

In the simulations described in this paper, the VisNet architecture was extended to incorporate additional feedback (top-down) connections, which have the similar topological connectivity as the feedforward connections except in the opposite direction (Fig. 3(b)). Both the feedforward and feedback connections to individual cells are derived from a topologically corresponding region of the preceding layer, using a Gaussian distribution of connection probabilities. These distributions are defined by a radius which will contain approximately 67% of the connections from the preceding layer. The values used in the current studies are given in Table 1. The gradual increase in the receptive field of cells in successive layers 1–3 reflects the known physiology of the primate ventral visual pathway (Freeman & Simoncelli, 2011; Pasupathy, 2006; Pettet & Gilbert, 1992).

Furthermore, in order to investigate the precise temporal dynamics of the top-down modulation, we have converted the original discrete time model, which has been used for past VisNet studies, into a time-continuous model with differential equations that are given below.

2.1.1. Pre-processing of the visual input by Gabor filters

Before the visual images are presented to the VisNet's input layer 1, they are preprocessed by a set of Gabor filters, previously implemented by Deco and Rolls (2004), which accord with the general tuning profiles of simple cells in V1 (Cumming & Parker, 1999; Jones & Palmer, 1987; Lades et al., 1993). The filters provide a unique pattern of filter outputs for each transform of each visual object, which is passed through to the first layer of VisNet. These filters are known to provide a good fit to the firing properties of V1 simple cells, which respond to local oriented bars and edges

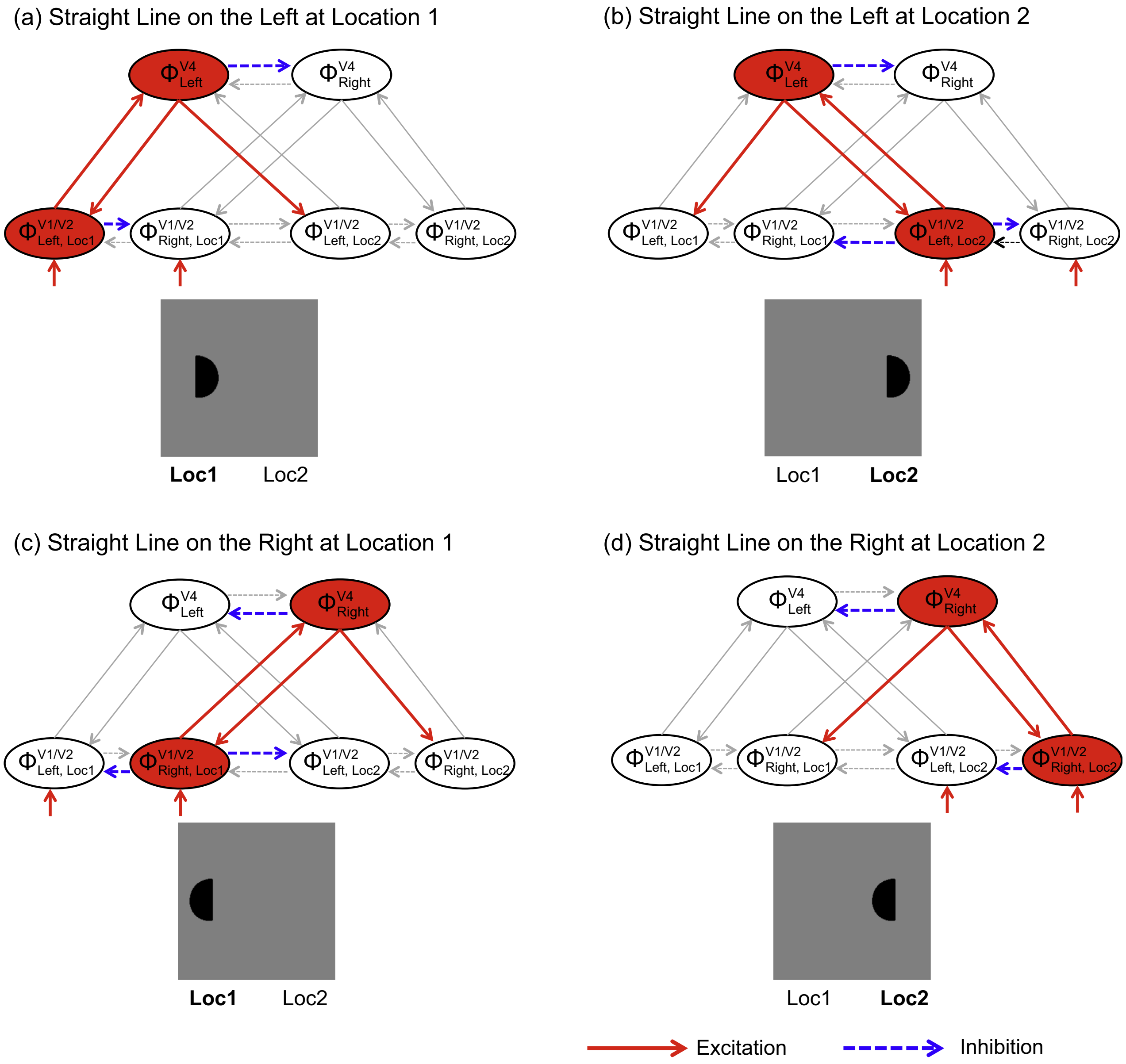

Fig. 2. Hypothesised modulation of edge detecting simple cells in lower layers V1/V2 by top-down signals from higher layer V4 neurons representing boundary contour elements. The figure shows the steady state activations of all neurons after sufficient time (e.g. ⩾61 ms) has elapsed after stimulus presentation to allow visual signals to propagate from the retina up to V4 and then back down to modulate V1/V2 responses. The following four cases are shown. (a) An object with a straight vertical border on its *left* is presented with this border positioned at retinal location *1*. Ascending visual input initially stimulates both subsets of V1/V2 neurons, $\Phi^{V1/V2}_{Left,Loc1}$ and $\Phi^{V1/V2}_{Right,Loc1}$, representing a vertical straight edge at retinal location 1. However, in layer V4, only those V4 neurons Φ^{V4}_{Left} representing a vertical straight edge on the left of an object are preferentially stimulated by the current visual input. Note that these V4 neurons receive additional feedforward (bottom-up) input signals from other V1/V2 neurons (not shown in the figure) which represent local image context, and these additional context signals are required to guide the selective responses of the V4 neurons. How V4 neurons may develop such response properties through self-organisation of the feedforward connections has been previously modelled by Eguchi et al. (2015). The subset of V4 neurons Φ^{V4}_{Left} then stimulates via feedback (top-down) connections those two subsets of V1/V2 neurons $\Phi^{V1/V2}_{Left,Loc1}$ and $\Phi^{V1/V2}_{Left,Loc2}$ which receive strengthened connections from Φ^{V4}_{Left} and are consequently modulated by a straight vertical edge on the left of an object. However, only the particular subset of V1/V2 cells $\Phi^{V1/V2}_{Left,Loc1}$, which represent a vertical bar at retinal location 1 where the vertical bar forms the left hand border of an object, receive the greatest combination of bottom-up and top-down input. Consequently, these V1/V2 neurons fire maximally, representing the border ownership of the vertical edge at this location. (b) An object with a straight vertical border on its *left* is presented with this border positioned at retinal location *2*. In this case, the subset of V1/V2 cells $\Phi^{V1/V2}_{Left,Loc2}$, which represent a vertical bar at retinal location 2 where the vertical bar forms the left hand border of an object, receive the greatest combination of bottom-up and top-down input and fire maximally. (c) An object with a straight vertical border on its *right* is presented with this border positioned at retinal location *1*. This time the subset of V1/V2 cells $\Phi^{V1/V2}_{Right,Loc1}$, which represent a vertical bar at retinal location 1 where the vertical bar forms the right hand border of an object, receive the greatest combination of bottom-up and top-down input and fire maximally. (d) An object with a straight vertical border on its *right* is presented with this border positioned at retinal location *2*. Now the subset of V1/V2 cells $\Phi^{V1/V2}_{Right,Loc2}$, which represent a vertical bar at retinal location 2 where the vertical bar forms the right hand border of an object, receive the greatest combination of bottom-up and top-down input and fire maximally.

within the visual field (Cumming & Parker, 1999; Jones & Palmer, 1987). The input filters used are computed by the following equations:

$$g(x,y,\lambda,\theta,\psi,b,\gamma)=\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda}+\psi\right) \tag{1}$$

with the following definitions:

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\theta + y\cos\theta \\ \sigma &= \frac{\lambda(2^b+1)}{\pi(2^b-1)}\sqrt{\frac{\ln 2}{2}} \end{aligned} \tag{2}$$

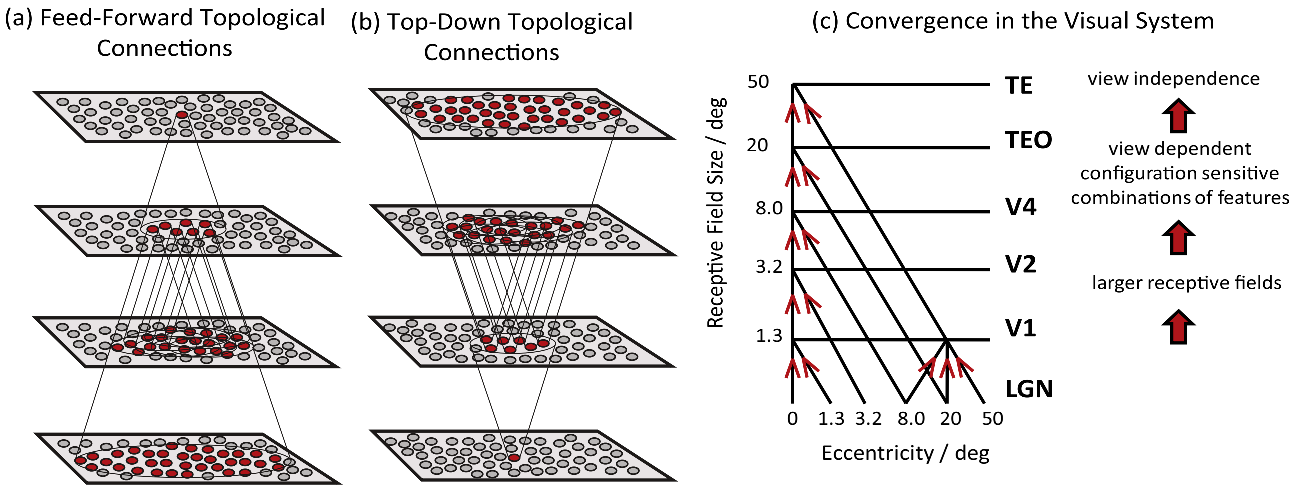
(a) Feed-Forward Topological Connections
(b) Top-Down Topological Connections
(c) Convergence in the Visual System
Receptive Field Size / deg
50
20
8.0
3.2
1.3
TE
TEO
V4
V2
V1
LGN
0
1.3
3.2
8.0
20
50
Eccentricity / deg
view independence
view dependent configuration sensitive combinations of features
larger receptive fields

Fig. 3. (a) The original four-layer feedforward (bottom-up) version of the VisNet architecture. The figure shows the feedforward connectivity, where each neuron receives connections from a topologically corresponding region of the preceding layer. The convergence of feedforward connections through the network is designed to provide fourth layer neurons with information from across the entire input retina. The new VisNet architecture implemented in this paper was extended to incorporate additional feedback (top-down) connections, which have the similar topological connectivity as the feedforward connections except in the opposite direction as shown in (b). (c) Convergence in the visual system V1: visual cortex area V1;TEO: posterior inferior temporal cortex, TE: anterior inferior temporal cortex (IT).

Table 1
Parameters used for simulations with VisNet.

Layer	1	2	3
(a) Parameters for VisNet model			
Dimensions	64 × 64	64 × 64	64 × 64
Number of feedforward fan-in connections	201	100	100
Fan-in Radius (feedforward)	12	12	18
Number of feedback fan-in connections	5	5	–
Fan-in Radius (feedback)	12	12	–
Sparseness of activations (set by adjusting sigmoid threshold α)	33%	33%	50%
Sigmoid slope (β)	31.5	46.1	1.48
Learning rate (k)	1.0	1.0	1.0
Excitatory Radius (σ_E)	1.4	1.1	0.8
Excitatory Contrast (δ_E)	5.35	33.15	117.57
Inhibitory Radius (σ_I)	2.76	5.4	8.0
Inhibitory Contrast (δ_I)	1.6	1.5	1.5
(b) Parameters for Gabor filtering			
Phase shift (ψ)	$0, \pi, -\pi/2, \pi/2$		
Wavelength (λ)	2		
Orientation (θ)	$0, \pi/4, \pi/2, 3\pi/4$		
Spatial bandwidth (b)	1.5 octaves		
Aspect ratio (γ)	0.5		
(c) Parameters for differential model			
Activation time constant (τ_h) [s]	0.1		
Trace time constant (τ_t) [s]	0.5		
Presentation time per stimulus transform [s]	1.0		
Numerical step size (Δt) [s]	0.01		

where x and y specify the position of a light impulse in the visual field (Petkov & Kruizinga, 1997). The parameter λ is the wavelength ($1/\lambda$ is the spatial frequency), σ controls number of such periods inside the Gaussian window based on λ and spatial bandwidth b, θ defines the orientation of the feature, ψ defines the phase, and γ sets the aspect ratio that determines the shape of the receptive field. In the experiments in this paper, an array of Gabor filters is generated at each of 256 × 256 retinal locations with the parameters given in Table 1.

The outputs of the Gabor filters are passed to the neurons in layer 1 of VisNet according to the synaptic connectivity given in Table 1. That is, each layer 1 neuron receives connections from 201 randomly chosen Gabor filters localised within a topologically corresponding region of the retina (this number has been used to be consistent with the original VisNet study (Wallis & Rolls, 1997)). These distributions are defined by a radius shown in Table 1.

2.1.2. Activations of neurons and competition within the network

Within each of the neural layers 1–3 of the network, the activation h_i of each neuron i is governed by the following differential equation:

$$\tau_h \frac{dh_i(t)}{dt} = -h_i(t) + \sum_j w_{ij}(t) r_j(t) \tag{3}$$

where τ_h is the time constant, r_j is the firing rate of presynaptic neuron j, and w_{ij} is the strength of the synapse from neuron j to neuron i. The value of τ_h used in the simulations is 0.1, which is larger than the typical values used for spiking network, 0.01. However, since we do not implement spikes of the neurons and the synaptic learning rule does not depend on the precise timing like STDP, we decided to use the larger time constant for this particular model for the speed of its computation. In this paper, the full differential model, which comprises Eqs. (3), (6) and (7) given below, is numerically simulated using a Forward Euler finite difference scheme with a fixed numerical timestep Δt given in Table 1.

In this paper, we have run simulations with a self-organising map (SOM) (Kohonen, 1982; von der Malsburg, 1973) implemented within each layer. In the SOM architecture, short-range excitation and long-range inhibition are combined to form a Mexican-hat spatial profile and is constructed as a difference of two Gaussians as follows:

$$I_{a,b} = -\delta_I \exp\left(-\frac{a^2+b^2}{\sigma_I^2}\right) + \delta_E \exp\left(-\frac{a^2+b^2}{\sigma_E^2}\right) \tag{4}$$

Here, to implement the SOM, the activations h_i of neurons within a layer are convolved with a spatial filter, $I_{a,b}$, where δ_I controls the inhibitory contrast and δ_E controls the excitatory contrast. The width of the inhibitory radius is controlled by σ_I while the width of the excitatory radius is controlled by σ_E. The parameters a and b index the distance away from the centre of the filter. The lateral inhibition and excitation parameters used in the SOM architecture are given in Table 1. These values were previously found to optimize the performance of the VisNet model (Rolls, 2000; Tromans, Harris, & Stringer, 2011).

Next, the contrast between the activations of neurons within each layer is enhanced by passing the activations of the neurons through a sigmoid transfer function as follows:

$$r = f^{sigmoid}(h') = \frac{1}{1 + \exp\left(-2\beta(h' - \alpha)\right)} \tag{5}$$

where h' is the activation after applying the SOM filter, r is the firing rate after contrast enhancement, and α and β are the sigmoid threshold and slope respectively. The parameters α and β are constant within each layer although α is adjusted within each layer of neurons to control the sparseness of the firing rates. For example, to set the sparseness to 5%, the threshold is set to the value of the 95th percentile point of the activations within the layer. The parameters for the sigmoid activation function are shown in Table 1.

2.1.3. Modification of synaptic weights during training

During training with visual objects, while the connectivity pattern is fixed, the strengths of the feedforward and feedback synaptic connections between successive neuronal layers are modified by a trace learning rule (Foldiak, 1991; Wallis & Rolls, 1997), which incorporates a memory trace of recent neuronal activity:

$$\frac{dw_{ij}(t)}{dt} = k\overline{r_i}(t)r_j(t) \tag{6}$$

where $r_j(t)$ is the firing rate of pre-synaptic neuron $j, \overline{r_i}(t)$ is the memory trace value of the firing rate of post-synaptic neuron i, w_{ij} is the synaptic weight from pre-synaptic neuron j to post-synaptic neuron i, and k is the learning rate constant. The memory trace value $\overline{r_i}(t)$ is updated according to the equation:

$$\tau_t \frac{\overline{r_i}(t)}{dt} = -\overline{r_i}(t) + r_i(t) \tag{7}$$

where $r_i(t)$ is the firing rate of post-synaptic neuron i, and τ_t is a trace time constant which is given in Table 1. The effect of the trace learning rule (6) is to encourage neurons to learn to respond to visual input patterns that tend to occur close together in time. The utility of this temporal binding is as follows. If, during training, each object is presented to the network in a sequence of different retinal locations clustered together in time before switching to the next object, then this enables neurons in higher layers to learn to respond to their preferred visual stimulus with shift invariance across different retinal locations as described in the earlier simulation study of Eguchi et al. (2015).

During the numerical simulation, to prevent the same few neurons always winning the competition, the synaptic weight vector $\mathbf{w}_i$ for each neuron i is normalised to unit length after each learning update for each training image by setting

$$\mathbf{w}_i = \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|} \tag{8}$$

where $\|\mathbf{w}_i\|$ is the length of the vector $\mathbf{w}_i$ given by

$$\|\mathbf{w}_i\| = \sqrt{\sum\nolimits_j w_{ij}^2} \tag{9}$$

Neurophysiological evidence for synaptic weight normalisation is provided by Royer and Paré (2003).

In the original discrete-time version of VisNet, the synaptic weights are trained layer by layer (Wallis & Rolls, 1997). However, it is important to note that in the current time-continuous version of VisNet, all the synapses across the layers are trained simultaneously. This means that every time step, each neuron calculates the weighted sum of the pre-synaptic activations, at both feed-forward and top-down synapses, to update the activation h (Eq. (3)). Next the neuronal firing rates within each layer are simultaneously determined by applying the SOM filter (Eq. (4)) and then the contrast enhancement (Eq. (5)). The trace learning rule (Eqs. (6) and (7)) is then applied at all of the synapses simultaneously to update the synaptic weights. In other words, in the current VisNet model, the training of the backprojections starts at the same time as the forward projections, with the bottom-up and top-down afferent connections to all of the layers being trained simultaneously.

2.2. Analysis techniques

Information theory is used to quantify how selective neurons are for members of a particular stimulus category. If a neuron responds invariantly to the members of a particular stimulus category but not to stimuli from other stimulus categories, then the neuron carries a high level of information about the presence of its preferred stimulus category.

For example, we have previously used information theory to quantify how well neurons have learned to respond selectively to a particular visual stimulus with translation invariance across different retinal locations (Eguchi et al., 2015). If the responses r of a neuron carry a high level of information about the presence of a particular stimulus s across different retinal locations, then this implies that the neuron will respond selectively to the presence of that stimulus regardless of where the stimulus is presented on the retina. In this way, information theory can provide a direct measure of both the selectivity of a neuron for a particular stimulus, as well as how translation-invariant the neuronal responses are as the stimulus is shifted across the retina.

In this paper, we continue to use information theory to assess the stimulus selectivity and translation invariance of neurons in the layer 3 that have learned to respond to localised object boundary elements with translation invariance, as previously investigated by Eguchi et al. (2015). However, in this new study we also apply information theory to assess how well simple cells in layer 1 have learned to represent border ownership through top-down modulation. We therefore use information theory to assess whether some layer 1 simple cells learn to respond selectively to a vertical straight edge on the *left* boundary of an object, while other simple cells learn to respond to a vertical straight edge on the *right* boundary of an object, regardless of the overall object shape. The simple cells in layer 1 have a small fan-in from the retina and are tuned to specific retinal locations, and consequently do not respond invariantly over different retinal locations. Instead, the simple cells should ideally respond invariantly over different global object shapes, as long as there is a straight vertical edge in the correct location on the object boundary.

Two information measures were used to assess network performance (see Rolls, Treves, Tovee, & Panzeri (1997) and Rolls & Milward (2000)). These two measure use the responses from either individual neurons (single-cell information analysis) or small ensembles of neurons (multiple-cell information analysis), each of which will be discussed in turn.

2.2.1. Single-cell information

A single cell information measure was applied to individual cells to measure how much information is available from the responses of a single cell about which stimulus category is present.

For border ownership simple cells in layer 1, there are two stimulus categories presented at each of two retinal training locations 1 and 2 (i.e., in total four stimulus categories). These two categories are: (i) a vertical straight edge which is on the left hand boundary of an object and (ii) a vertical straight edge which is on the right hand boundary of an object. Therefore, to score high single cell information, a layer 1 neuron must respond selectively either to all object shapes with a vertical straight bar on the left or all object shapes with a straight vertical bar on the right, but only for one of the two retinal locations.

On the other hand, we are interested in measuring translation invariant responses of the cells in layer 3 as V4 neurons. Accordingly, although the responses of neurons in layer 3 are assessed using the same two stimulus categories, we calculated how well those cells learned to respond invariantly to stimuli presented in both retinal locations 1 and 2. In other words, there are in total two stimulus categories in this case instead of four in the case of layer 1. Therefore, to score high single cell information, a layer 3 neuron must respond either to all object shapes with a vertical straight bar on the left or all object shapes with a straight vertical bar on the right, and do so for both of the two retinal locations.

To be informative in the context of this study, the responses of a given neuron (r) should be specific to the presence of a straight vertical edge at a particular side (s = left/right), and independent of the remaining global shape of the object (in layers 1 and 3) or retinal location (in layer 3). The amount of stimulus specific information that a specific cell carries is calculated from the following formula with details given by Rolls and Milward (2000):

$$I(s,\vec{R}) = \sum_{r\in\vec{R}} P(r|s)\log_2 \frac{P(r|s)}{P(r)} \tag{10}$$

Here s is a particular stimulus and $\vec{R}$ is the set of responses of a cell to the set of stimuli.

The maximum information that an ideally developed cell could carry is given by the formula:

$$\text{Maximum cell information} = \log_2(n)\text{bits} \tag{11}$$

where n is a number of different stimulus categories. For example, in the case of translation invariant representation in Layer 3 with two stimulus categories, the maximum information possible is 1 bit.

2.2.2. Multiple-cell information

While useful in assessing the tuning properties of individual neurons, the single-cell information measure cannot give a complete assessment of VisNet's performance with respect to recognition of the full set of stimulus categories. If all cells learned to respond to the same stimulus category (according to the single-cell measure) then there would be relatively little information available about the whole set stimulus categories $\vec{S}$. To address this issue, we also calculated a multiple-cell information measure, which assesses the amount of information that is available about the whole set of stimulus categories from a *subpopulation* of neurons. This measure quantifies the network's ability to tell which stimulus is currently presented to the network based on the set of responses, $\vec{R}$, of a subpopulation of cells.

In brief, we would like to calculate the mutual information between the stimulus categories and the neuronal responses – the average amount of information obtained (across all stimuli) from the responses of the neuronal ensemble, about which stimulus category was present after a single presentation of a stimulus. However, due to the difficulty in adequately sampling this high dimensional neural response space, it is challenging to construct accurate probability distributions for directly calculating the mutual information. Instead, a decoding procedure is used to estimate which stimulus s' gave rise to the particular firing rate response vector on each trial. A probability table is then constructed between the real stimuli, s, and the decoded stimuli, s'. From this probability table, the multiple-cell information is then calculated as follows.

$$I_{\vec{C}}(S,S') = \sum_{s,s'} P(s,s')\log_2 \frac{P(s,s')}{P(s)P(s')} \tag{12}$$

$$P(s') = \sum_{s\in S} P(s'|R_{\vec{C}}(s)) \times P(R_{\vec{C}}(s))) \tag{13}$$

$$P(s,s') = P(s'|R_{\vec{C}}(s)) \times P(R_{\vec{C}}(s)) \tag{14}$$

Here, S represents the set of the stimulus categories presented to the network, and $\vec{C}$ defines the set of cells used in the analysis. For each analysis, the ensembles of cells are sampled from the pool of the cells which consists of five cells that had, as single cells, the most information about each stimulus category (i.e., the size of the pool is 5 × *number_of_the_stimulus_category*). From the set of cells $\vec{C}$, the firing responses $R_{\vec{C}}$ ($R = r(c)|c \in \vec{C}$) to each stimulus in S are used as the basis for the Bayesian decoding procedure.

For a given set of cells, the probabilities generated by the decoding procedure are factored into a confusion matrix, which matches up the actual input stimulus category in $\vec{S}$ with the predicted stimulus category in $\vec{S'}$. Here, $P(s_i')$ represents the probability that the predicted stimulus category s'_i is actually the stimulus category s_i that is currently presented to the network. A higher value of $P(s,s')$ relative to $P(s)P(s')$ indicates a stronger relationship between s and s'. This information provides the basis for calculating the multiple-cell information analysis. More details of the decoding procedure is provided in Rolls and Milward (2000).

3. Simulation results

3.1. Study 1: simulation of the visually-guided development of border ownership representations

In this simulation study, VisNet was initially trained and tested on the same abstract visual object shapes (familiar objects) shown in Fig. 4(a). The model was then also cross-validated by testing the same trained network on the novel visual objects (novel objects) shown in Fig. 4(b), which were not presented to the network during initial training. The familiar objects were hexagons and semicircles, which were either black or light grey. Black objects were presented against a light grey background, while light grey objects were presented against a black background. Each object had a vertical straight edge either on its left boundary (Fig. 4(a2,a4)) or right boundary (Fig. 4(a1,a3)). Although in the natural environment, the object does not normally jump from one location to the other instantaneously, the region activated on the retina does constantly shifts around due to the rapid eye movement called saccades. To simulate this effect, during training and testing, each object was presented in two locations on the left (Location 1) and right (Location 2) of the 256 × 256 retina.

Whenever an object was presented on the left of the retina, the vertical straight edge on its (left or right) boundary was precisely aligned with retinal Location 1 (Fig. 4(a1,a2)). This enabled us to explore the top-down modulation of the subpopulation of simple cells in Layer 1 tuned to vertical straight edges at this specific retinal location. In a similar manner, whenever the object was presented on the right of the retina, the vertical straight edge on its (left or right) boundary was aligned with retinal Location 2 (Fig. 4(a3,a4)). Again, this permitted us to explore the top-down modulation of simple cells in Layer 1 tuned to vertical straight edges at this retinal location.

During training, the familiar objects shown in Fig. 4(a) were presented to the network one at a time shifting across the two retinal locations while the feedforward and feedback synaptic connections between successive layers were modified using the trace learning rule (6) and (7). The trace learning rule in the feedforward connections drives the development of neuronal responses in the higher layers that are translation invariant across different retinal locations by encouraging postsynaptic neurons to learn to respond to subsets of input patterns that tend to occur close together in time. As long as, during training, each object is presented across different retinal locations in temporal proximity, then the trace

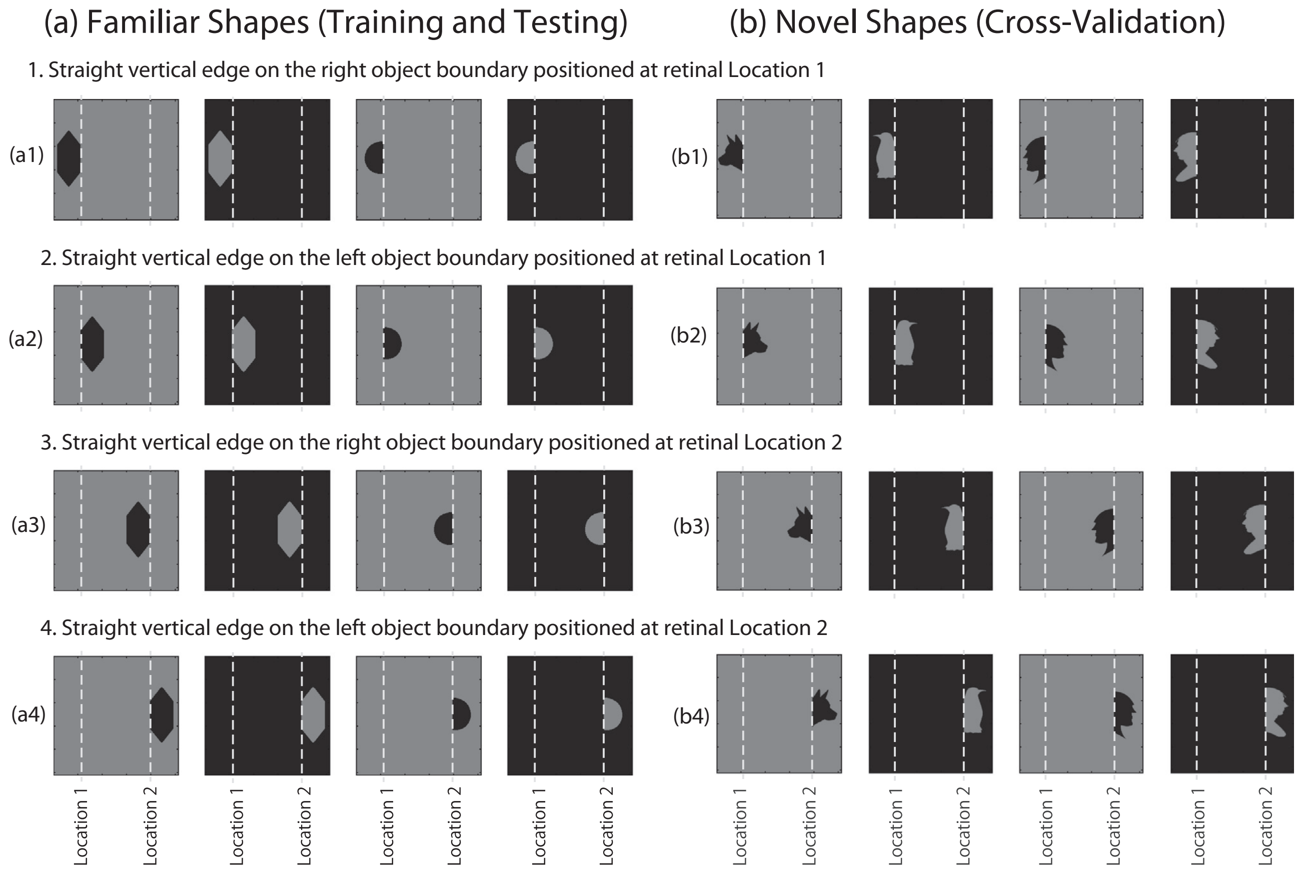
(a) Familiar Shapes (Training and Testing)
(b) Novel Shapes (Cross-Validation)
1. Straight vertical edge on the right object boundary positioned at retinal Location 1
2. Straight vertical edge on the left object boundary positioned at retinal Location 1
3. Straight vertical edge on the right object boundary positioned at retinal Location 2
4. Straight vertical edge on the left object boundary positioned at retinal Location 2
(a1)
(a2)
(a3)
(a4)
(b1)
(b2)
(b3)
(b4)
Location 1
Location 2

Fig. 4. The visual object stimuli used for the simulation study. (a) A set of abstract familiar shape stimuli used to both train and test the network model (shaded hexagons and semicircles). The objects were black when presented on a light grey background or light grey when presented on a black background. Each object had a vertical straight edge either on its left boundary (a1, a3) or right boundary (a2, a4). During training and testing, each object was presented in two locations on the left and right of the retina. Whenever an object was presented on the left of the retina, the vertical straight edge on its (left or right) boundary was precisely aligned with retinal Location 1 (a1, a3). Similarly, whenever the object was presented on the right of the retina, the vertical straight edge on its (left or right) boundary was aligned with retinal Location 2 (a2, a4). (b) A set of novel stimuli used to cross-validate the performance of the network after it had been trained on the familiar set of stimuli (a). The four novel stimuli were a dog's head, a penguin, and two differently shaped human heads. Each novel stimulus has a vertical straight edge on one side. The four novel objects are each presented in two retinal locations in a similar manner to the familiar shapes (a). This gives a total of eight novel stimulus presentations.

learning rule will produce output neurons that have learned to respond selectively to a particular object feature in a translation invariant manner. Therefore, during training, we selected each object in turn and presented that object in the two different retinal locations before moving on to the next object.

3.1.1. Steady state firing properties of cells in layers 1 and 3 at the end of each stimulus presentation

In this section we analyse the *steady state* firing responses of Layer 1 and Layer 3 neurons at the end of each stimulus presentation before and after training with the same object stimuli used for training (familiar objects) shown in Fig. 4(a) as well as with the novel object stimuli shown in Fig. 4(b) to cross-validate the developed response properties.

We first tested the firing properties of the output (Layer 3) neurons to investigate whether these neurons had learned to respond selectively to the presence of a vertical straight edge on either the left boundary or right boundary of an object. Such neurons had to respond invariantly across different global object shapes (i.e. hexagon or semicircle), different kinds of object shading (i.e. black or light grey), and different trained retinal locations (i.e. Location 1 or Location 2). The same set of stimuli used to train the network shown in Fig. 4(a) was presented to VisNet during testing, and the firing rate of each neuron in the output layer of the network was recorded. In order to quantify the performance, information analysis was conducted as described in Section 2.2.

In this analysis, there are two different stimulus categories ($n = 2$) as explained in Section 2.2. In Fig. 4(a), stimuli from the first category with a vertical straight edge on the left are shown in rows (b) and (d), while stimuli from the second category with a vertical straight edge on the right are shown in rows (a) and (c). Since each category member was defined by its shape (hexagon or semicircle), shading (black or light grey), and retinal location (Location 1 or Location 2), there were $2^3 = 8$ members of transforms of each of the two stimulus categories. Individual Layer 3 neurons had to respond invariantly over the eight transforms of its preferred stimulus category, and not respond to any members of the other stimulus category, in order to carry maximum information about its preferred category.

Fig. 5 shows the information analysis of the steady state response properties of Layer 3 neurons at the end of each stimulus presentation. Results are presented before and after training. Plot (a) shows the single cell information analysis. The maximum amount of information possible for the simulation is $\log_2(n)$ where n is the number of stimulus categories = 2, that is 1 bit. Before training, no neurons reached 1 bit of information and in fact most neurons carried much less than 1 bit. However, after training, nearly all the neurons carried 1 bit of information. This result confirms that nearly all of the Layer 3 neurons had successfully

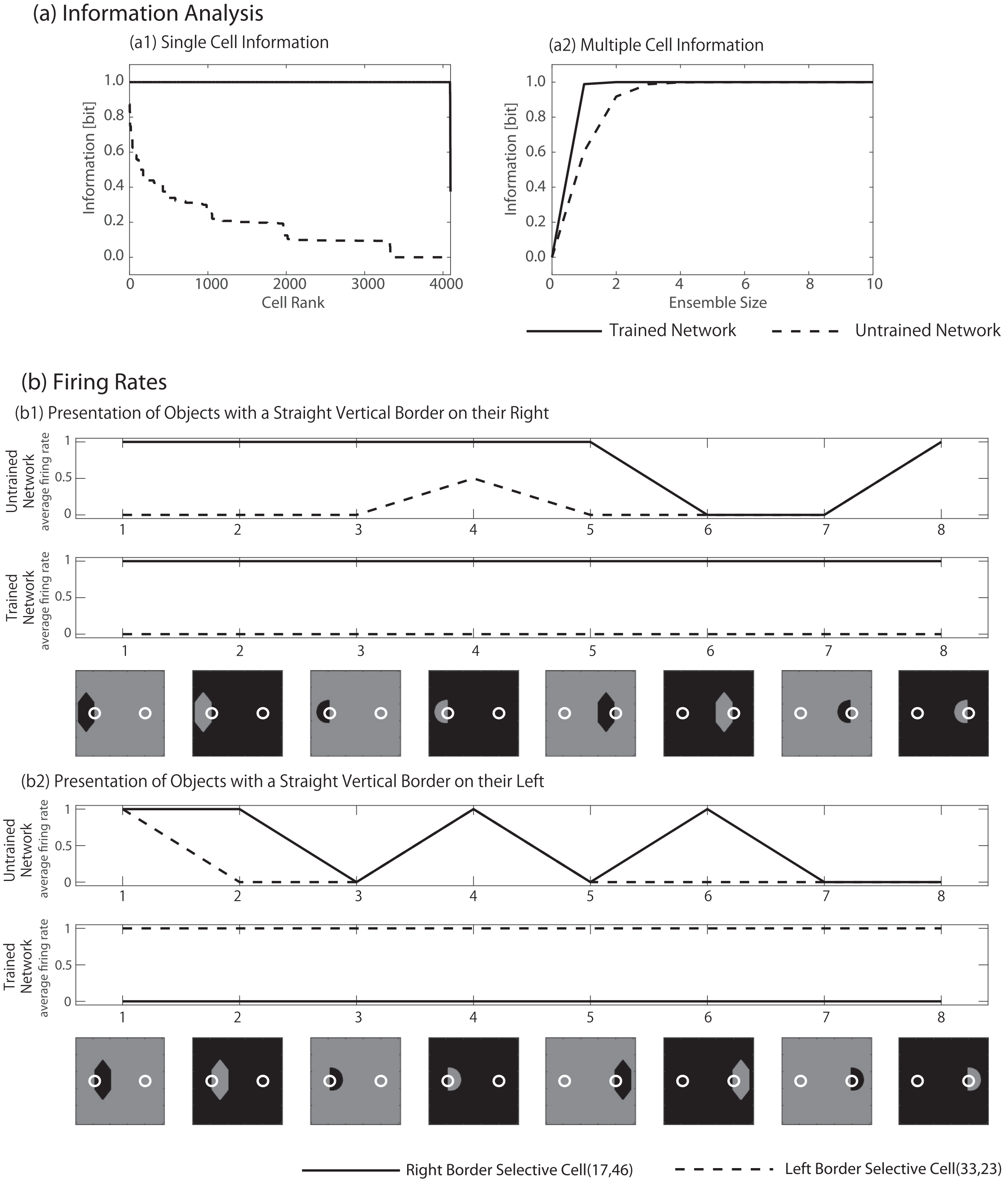

Fig. 5. The steady state response properties of Layer 3 neurons at the end of each stimulus presentation of familiar shapes that were used to train the network (shown in Fig. 4 (a)). (a) **Information analysis:** We computed the information carried by the output (3rd layer) neurons about whether the vertical straight edge was on the left or right boundary of each object presented to the network before and after training. Plot (a1) shows the maximum single cell information carried by each of the 4096 neurons in Layer 3 about which one of the two stimulus categories was presented, where all of the neurons in Layer 3 are plotted along the abscissa in rank order. The result shows that nearly all of the Layer 3 neurons learned to respond selectively to a vertical straight edge either on the left or on the right of an object boundary, regardless of the global shape, shading or retinal location of the object. Plot (a2) shows the multiple cell information carried by different sized (i.e. up to ten neurons) random ensembles of Layer 3 neurons that individually had high levels of single cell information. It is evident that training has led to an increase in the multiple cell information, which after training asymptotes to the maximum level of 1 bit with only one neuron included in the analysis. (b) **Firing rate responses of two Layer 3 neurons that has maximum single cell information:** plot (b1) shows the responses of two Layer 3 neurons to all eight objects with a vertical straight edge on their right boundary, and plot (b2) shows the responses of the same two Layer 3 neurons to all eight objects with a vertical straight edge on their left boundary. These results show that neuron (17, 46) learned to respond selectively to all objects with a vertical straight edge on the right, while neuron (33, 23) learned to respond to all objects with a vertical straight edge on the left.

learned to respond selectively to a vertical straight edge either on the left or on the right of an object boundary, regardless of the global shape, shading or retinal location of the object.

Plot (b) shows the multiple-cell information analysis. It is evident that training has led to an increase in the multiple cell information, which after training asymptotes to the maximum level of

1 bit with only one neuron included in the analysis. This is possible because, in the case of just two stimulus categories, the low or high firing responses of a single perfectly discriminating neuron will provide 1 bit of information about both stimulus categories. However, further inspection of the responses of Layer 3 neurons confirmed that some neurons had learned to respond selectively to objects with a straight vertical edge on the left boundary, while other neurons had learned to respond to a straight vertical edge on the right object boundary. This confirmed that the population of Layer 3 neurons learned to represent both of these stimulus categories.

Fig. 5(b) shows the steady state firing rate responses of two typical Layer 3 neurons (17, 46) and (33, 23) at the end of each stimulus presentation. The firing rate responses are plotted before and after training. Plot (b1) shows the responses of the two Layer 3 neurons to all eight object stimuli from the second stimulus category, i.e. objects with a vertical straight edge on their right boundary. While plot (b2) shows the responses of the same two Layer 3 neurons to all eight object stimuli from the first stimulus category, i.e. objects with a vertical straight edge on their left boundary. The white circle plotted on each stimulus gives the idea of the size of the fan-in radius of the neurons in the input layer of the network. The results show that, after training, neuron (17, 46) had learned to respond selectively to all objects with a vertical straight edge on the right, while neuron (33, 23) had learned to respond to all objects with a vertical straight edge on the left. These observed firing rate responses in Layer 3 were similar to those experimentally observed in area V4 of the primate visual system (Pasupathy & Connor, 2001, 2002) and demonstrated in the previous simulation study of Eguchi et al. (2015). These are the kind of neuronal response characteristics needed to provide top-down modulation of border ownership neurons in Layer 1 (corresponding to V1/V2).

Since the study above uses exactly the same two shapes (hexagon and semicircle) for training and testing the network, there is a possibility that the responses of the developed cells are specific to the set of actual trained objects and might not generalise to novel objects not encountered during training. Therefore, in order to cross-validate the response characteristics of these neurons, the four novel shapes shown in Fig. 4(b) are presented to the same trained network and the firing rates are recorded. In other words, the network was trained with the objects shown in Fig. 4(a) and then tested with a set of four different novel shapes shown in Fig. 4(b).

Fig. 6 shows the firing rate responses of the two Layer 3 neurons, which were previously tested on familiar objects in Fig. 5 (b), at the end of each novel stimulus presentation before and after training. Similar to the original set of shapes used to train the network, each shape contains a vertical straight edge on either the right or left and is presented at two different retinal locations (i.e., Location 1 or Location 2). Fig. 6(a) shows the responses of two Layer 3 neurons to all eight novel object stimuli from the second stimulus category, i.e. objects with a vertical straight edge on their right boundary. Before training, neuron (17, 46) and neuron (33, 23) both responded quite erratically to the different object stimuli. However, after training, neuron (17, 46) responded to all of the objects with a vertical straight edge on their right, while neuron (33, 23) did not respond to any of these stimuli. Plot (b) shows the responses of the same two Layer 3 neurons to all eight novel object stimuli from the first stimulus category, i.e. objects with a vertical straight edge on their left boundary. Before training, neurons (17, 46) and (33, 23) responded quite erratically to the different object stimuli. However, after training, neuron (33, 23) responded to all of the objects with a vertical straight edge on their left, while neuron (17, 46) did not respond to any of these stimuli. Taken together, these results show that neuron (17, 46) learned to respond selectively to all novel objects with a vertical straight edge on the right, while neuron (33, 23) learned to respond to all novel objects with a vertical straight edge on the left. Thus, the neurons continued to respond selectively to the presence of a vertical straight edge on either the left or the right of an object even if the objects are novel. This result confirms that the representations developed in the output layer of VisNet are not specific to the set of trained objects, but are in fact more generally selective to the presence of a vertical straight edge on either the left boundary or right boundary of an object.

We next tested whether Layer 1 neurons had developed the kind of border ownership representations reported by Zhou et al. (2000). In other words, we tested whether the feedback (top-down) connections newly implemented in VisNet enabled the activity in Layer 3 (corresponding to the experimentally observed neural responses in primate visual area V4) to successfully modulate the responses of neurons in Layer 1 (corresponding to visual areas V1/V2) such that the Layer 1 simple cells representing vertical straight edges at either retinal Location 1 or 2 responded selectively depending on whether the vertical straight edge was on the left or right boundary of the object.

In order to quantify the performance of Layer 1 neurons, we computed the information carried by the steady state responses of these cells at the end of each stimulus presentation. The results of this analysis are presented in Fig. 7(a), where we show the information carried by Layer 1 neurons before and after training. Layer 1 neurons are not expected to develop translation invariance across different retinal locations due to the small fan-in of connections from the retina. Therefore, we computed information that was specific to either retinal Location 1 or Location 2. Specifically, the analysis calculated the information carried by the Layer 1 neurons about whether the vertical straight edge in the object stimulus presented to the network was an example from one of four stimulus categories: (i) the vertical straight edge is positioned at retinal Location 1 and is on the left boundary of the object presented there, (ii) the vertical straight edge is positioned at retinal Location 1 and is on the right boundary of the object presented there, (iii) the vertical straight edge is positioned at retinal Location 2 and is on the left boundary of the object presented there, and (iv) the vertical straight edge is positioned at retinal Location 2 and is on the right boundary of the object presented there. Since there are $n = 4$ stimulus categories, perfectly discriminating neurons carry a maximum of $\log_2(n) = 2$ bits of information.

Fig. 7(a1) shows the single cell information analysis. The plot shows the maximum information carried by each of the 4096 neurons in Layer 1 about which one of the four stimulus categories was presented. It can be seen that training the network has led to a large increase in the number of neurons carrying the maximum 2 bits of information. After training, 145 cells learned to carry the maximum single cell information. These Layer 1 neurons thus provide the kind of border ownership representations experimentally observed in cortical visual area V1 by Zhou et al. (2000). Plot (a2) shows the multiple-cell information analysis.

Although one may be confused with the unexpectedly good decoding performance even in the untrained network, this can be explained by the topologically established synaptic connections and the feedback connections from the neurons in the higher layer which has larger size of receptive field. However, as long as there is some statistical correlations between the input pattern and the output, the multiple-cell information analysis can better capture the information than the single-cell information analysis. Therefore, it is important to see whether the performances improved after the training or not. In this case, although the change is not as obvious as the case of the layer 3, it is still evident that training has led to an increase in the multiple cell information, which after training asymptotes to the maximum level of 2 bits with only two neurons included in the analysis.

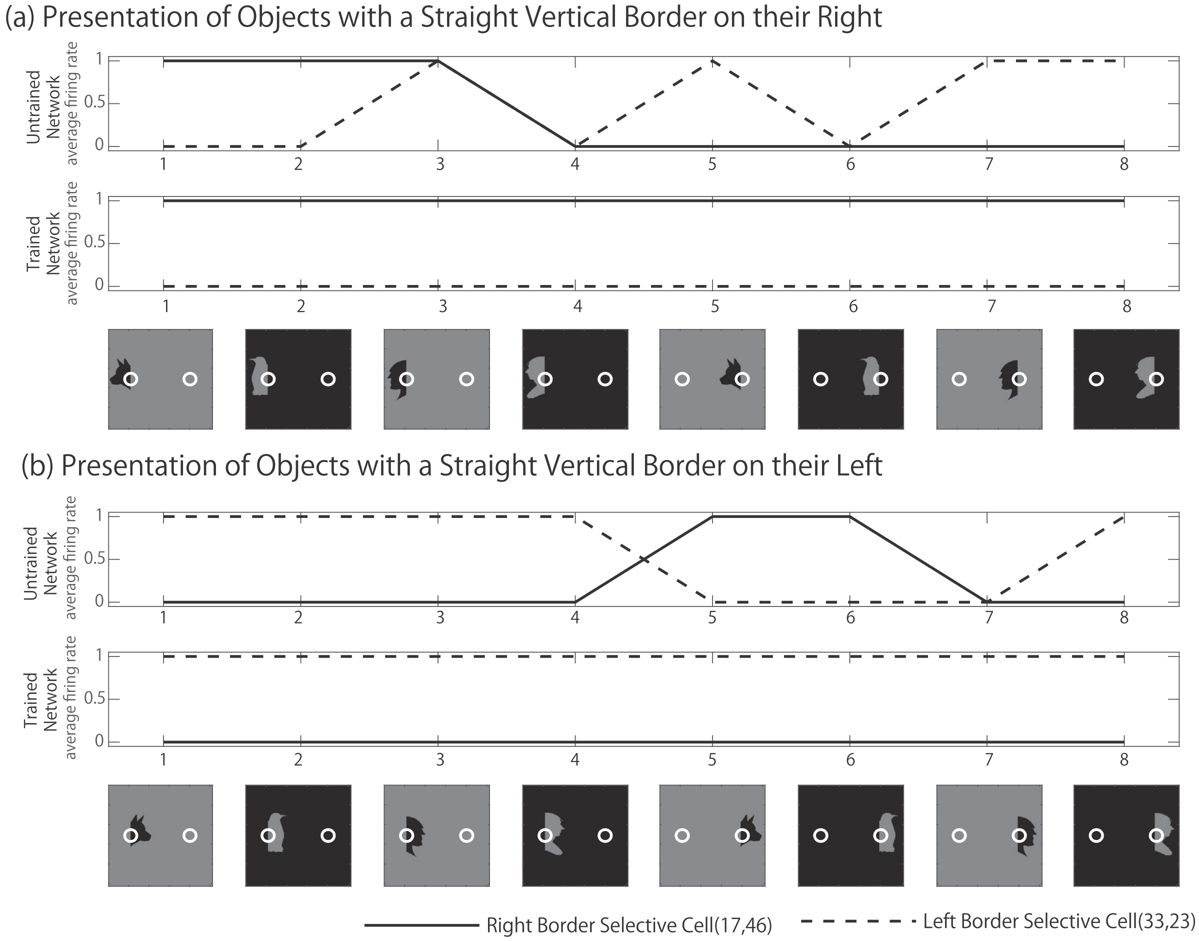

Fig. 6. Cross-validation of the developed firing properties of neurons in Layer 3 with the set of novel shapes not presented during training as shown in Fig. 4(b). Each of the four novel objects is presented in two retinal locations giving a total of eight novel stimulus presentations. This figure shows the firing rate responses of the same two Layer 3 neurons that were previously tested on familiar objects in Fig. 5(b). Plot (a) shows the responses of the two Layer 3 neurons to all eight novel stimulus presentations with a vertical straight edge on their right boundary, and plot (b) shows the responses of the same two Layer 3 neurons to all eight novel stimulus presentations with a vertical straight edge on their left boundary. These results show that neuron (17, 46) learned to respond selectively to all objects with a vertical straight edge on the right, while neuron (33, 23) learned to respond to all objects with a vertical straight edge on the left. These results confirm that the developed firing properties of the cells are not specific to the set of trained objects and generalise to the set of novel objects not presented during training.

Fig. 7(b) shows the steady state firing rate responses of four typical Layer 1 neurons at the end of each stimulus presentation before and after training. Plot (b1) shows the responses of the four Layer 1 neurons to all eight object stimuli with a vertical straight edge on their right boundary. After training, neuron (22, 8) responded selectively to all of the objects with a vertical straight edge on their right boundary aligned with retinal Location 1, while neuron (42, 43) responded to all of the objects with a vertical straight edge on their right boundary aligned with retinal Location 2. Plot (b2) shows the responses of the same four Layer 1 neurons to all eight object stimuli with a vertical straight edge on their left boundary. After training, neuron (34, 53) responded selectively to all of the objects with a vertical straight edge on their left boundary aligned with retinal Location 1, while neuron (32, 12) responded to all of the objects with a vertical straight edge on their left boundary aligned with retinal Location 2.

The developed firing properties were next cross-validated by testing the same trained network on the novel set of shapes shown in Fig. 4(b). Fig. 8(b) shows the firing rate responses of the same four Layer 1 neurons that were previously tested on familiar objects in Fig. 7(b). Plot (b1) shows the responses of the four Layer 1 neurons to all eight object stimuli with a vertical straight edge on their right boundary. The first four stimuli 1–4 shown along the abscissa have the object presented in retinal Location 1, while the next four stimuli 5–8 have the object presented in retinal Location 2. Plot (b2) shows the responses of the same four Layer 1 neurons to all eight object stimuli with a vertical straight edge on their left boundary. The first four stimuli 1–4 shown along the abscissa have the object presented in retinal Location 1, while the next four stimuli 5–8 have the object presented in retinal Location 2. It can be seen that each of the four Layer 1 neurons responds selectively to one of the four stimulus categories: (i) Location 1/ left boundary, (ii) Location 1/ right boundary, (iii) Location 2/ left boundary, and (iv) Location 2/ right boundary. These results confirm that the border ownership representations developed in Layer 1 are not specific to the set of trained objects, and are in fact able to generalise to the set of novel object shapes. Thus, different Layer 1 neurons had learned to respond selectively to the presence of a vertical straight edge on either the left boundary or right boundary of an object when the edge is aligned with a particular retinal location. These are the same kinds of border ownership representations reported in the neurophysiology study of primate visual area V1 carried out by Zhou et al. (2000).

3.1.2. Dynamical firing properties of cells in layer 1 during each stimulus presentation: time course of the emergence of border ownership signals

Sugihara et al. (2011) reported that the representation of border ownership in primate visual area V1, i.e. the selective modulation of the responses of V1 neurons that encode vertical straight edges

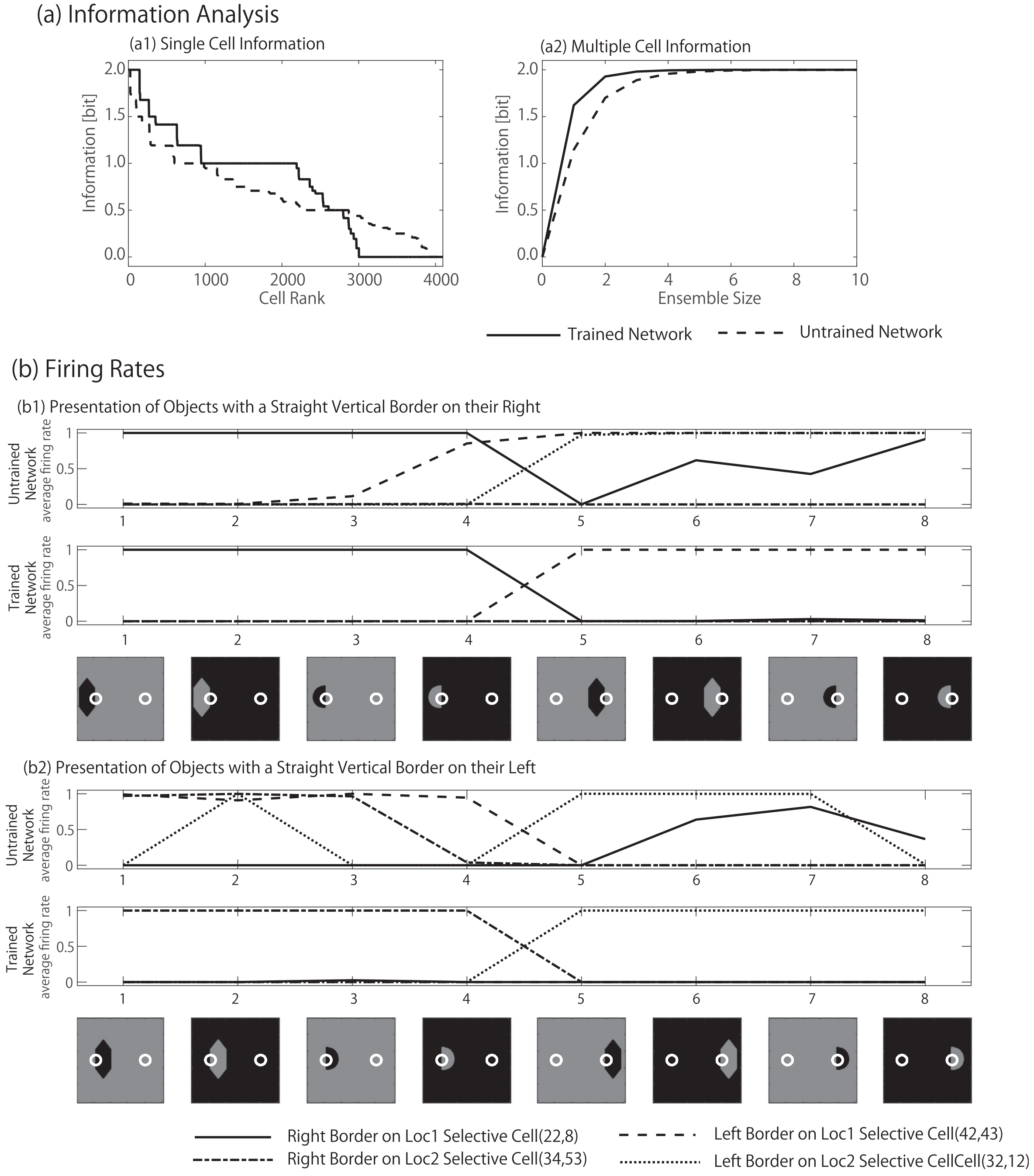

Fig. 7. Steady state response properties of Layer 1 neurons at the end of each stimulus presentation of the familiar shapes that were used to train the network (Fig. 4(a)). (a) **Information analysis:** Since Layer 1 neurons are not expected to develop translation invariance across different retinal locations, we computed the information carried by these neurons about whether the vertical straight edge in the object stimulus was from one of four stimulus categories: (i) Location 1/ left boundary, (ii) Location 1/ right boundary, (iii) Location 2/ left boundary, and (iv) Location 2/ right boundary. Since there are n = four stimulus categories, perfectly discriminating neurons carry a maximum of 2 bits of information. Plot (a1) shows the maximum single cell information carried by each of the 4096 neurons in Layer 1 about which one of the four stimulus categories was presented, where all of the neurons in Layer 1 are plotted along the abscissa in rank order. The result shows that these Layer 1 neurons have learned to respond with perfect selectivity to one of the four stimulus categories, thus providing the kind of border ownership representations experimentally observed in cortical visual area V1 by Zhou et al. (2000). Plot (a2) shows the multiple cell information carried by different sized (i.e. up to ten neurons) random ensembles of Layer 1 neurons that individually had high levels of single cell information. It is evident that training has led to an increase in the multiple cell information, which after training asymptotes to the maximum level of 2 bits with only two neurons included in the analysis. (b) **The firing rate responses of four Layer 1 neurons with maximum single cell information:** plot (b1) shows the responses of the four Layer 1 neurons to all eight object stimuli with a vertical straight edge on their right boundary. The first four stimuli 1–4 shown along the abscissa have the object presented in retinal Location 1, while the next four stimuli 5–8 have the object presented in retinal Location 2. Plot (b2) shows the responses of the same four Layer 1 neurons to all eight object stimuli with a vertical straight edge on their left boundary. The first four stimuli 1–4 shown along the abscissa have the object presented in retinal Location 1, while the next four stimuli 5–8 have the object presented in retinal Location 2. These results show that different Layer 1 neurons had learned to respond selectively to each of the four stimulus categories. These are the same kinds of border ownership representations found experimentally in primate visual area V1 by Zhou et al. (2000).

by whether the edge appears on the left or right boundary of an object, begins to appear at around 61 ms after the presentation of the visual stimulus. We hypothesise that this gradual emergence of the border ownership signal in area V1 is due to the time it takes for visual signals to propagate up to higher visual areas such as V4, where neurons may represent a vertical straight edge on either the

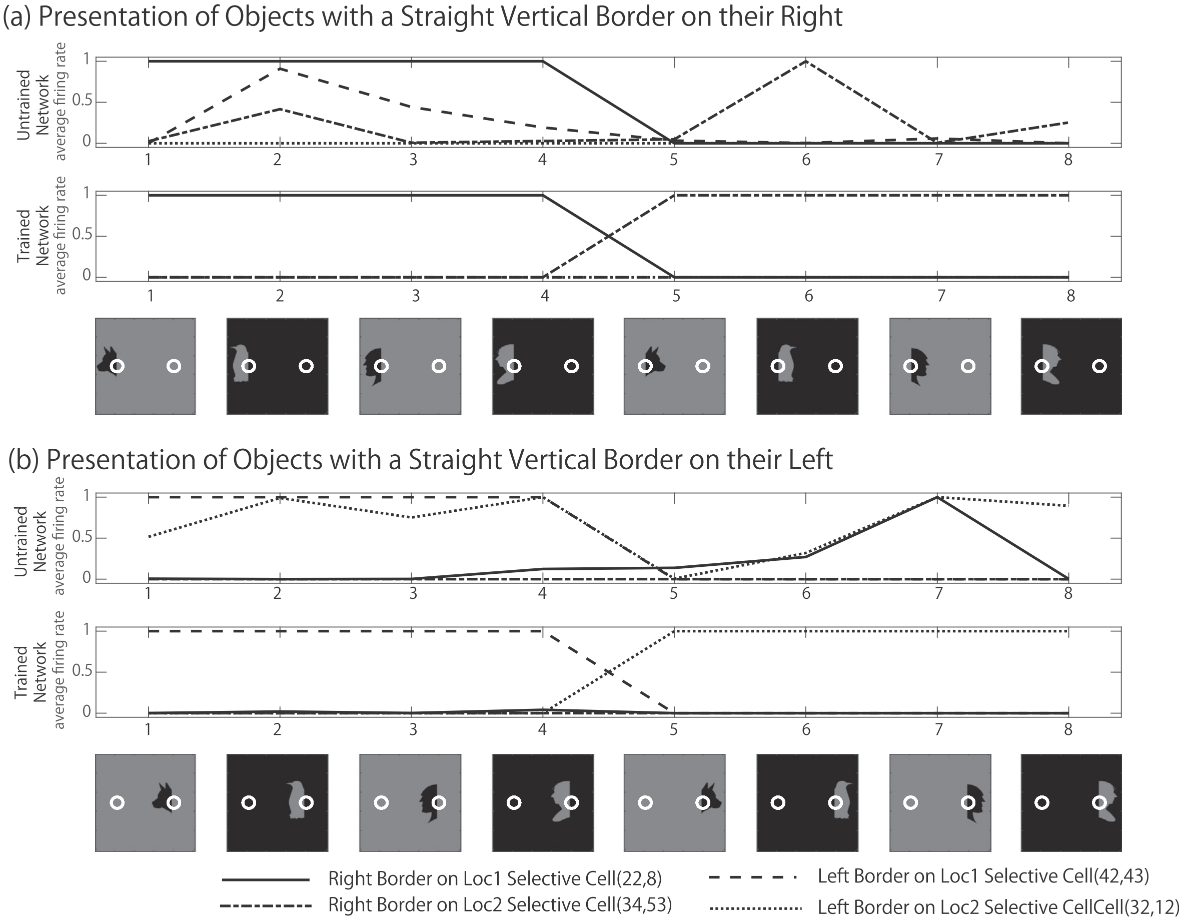

Fig. 8. Cross-validation of the developed firing properties of neurons in Layer 1 with the set of novel shapes not presented during training as shown in Fig. 4(b). Each of the four novel objects is presented in two retinal locations giving a total of eight novel stimulus presentations. This figure shows the firing rate responses of the same four Layer 1 neurons that were previously tested on familiar objects in Fig. 7(b). Plot (a) shows the responses of the four Layer 1 neurons to all eight novel object stimuli with a vertical straight edge on their right boundary. The first four stimuli 1–4 shown along the abscissa have the object presented in retinal Location 1, while the next four stimuli 5–8 have the object presented in retinal Location 2. Plot (b) shows the responses of the same four Layer 1 neurons to all eight novel object stimuli with a vertical straight edge on their left boundary. The first four stimuli 1–4 shown along the abscissa have the object presented in retinal Location 1, while the next four stimuli 5–8 have the object presented in retinal Location 2. It is evident that each of the four Layer 1 neurons has learned to respond to one of the four stimulus categories: (i) Location 1/left boundary, (ii) Location 1/right boundary, (iii) Location 2/left boundary, and (iv) Location 2/right boundary. These results confirm that the border ownership representations developed in Layer 1 were not specific to the set of trained objects and generalise to the set of novel objects not presented during training.

left or right of an object boundary across different retinal locations, and then to propagate back down to modulate the activities of neurons in area V1. We investigated this proposal computationally by recording the temporal evolution of the responses of border ownership neurons in Layer 1 of the trained VisNet model during 300 ms stimulus presentations.

Fig. 9 shows the dynamical evolution through time of the border ownership representations conveyed by four typical neurons in Layer 1 during the 300 ms time course of stimulus presentations. The results are shown after training has established border ownership representations in Layer 1. Each row shows results for one of the four neurons, where each neuron is tuned to a different border ownership category as follows: (Row 1) the neuron is tuned to a vertical straight edge on the right object boundary aligned with retinal Location 1, (Row 2) the neuron is tuned to a vertical straight edge on the left object boundary aligned with retinal Location 1, (Row 3) the neuron is tuned to a vertical straight edge on the right object boundary aligned with retinal Location 2, and (Row 4) the neuron is tuned to a vertical straight edge on the left object boundary aligned with retinal Location 2. Column (a) shows the average responses of each neuron to the members of its preferred stimulus category (solid line) and the members of its three non-preferred stimulus categories (dashed line) plotted over the 300 ms time courses of the stimulus presentations. It can be seen that the firing responses of all four neurons begin to strongly differentiate between their preferred and non-preferred stimulus categories by about 50 ms after the start of stimulus presentation. By the end of the stimulus presentation at 300 ms, the responses of the neurons fully differentiate between their preferred and non-preferred stimulus categories. Column (b) shows the average single cell information carried by the four neurons about their preferred stimulus category plotted over the 300 ms time courses of the stimulus presentations. Consistent with the firing rate plots, there is a monotonic increase in the information carried by each of the four neurons during the 300 ms time course of stimulus presentation.

The simulation results show how the border ownership representations gradually emerge in Layer 1 over the time course of 300 ms during stimulus presentation. Near the beginning of the stimulus presentation, the Layer 1 neurons merely represent the presence of a straight vertical edge at a particular retinal Location 1 or 2. The Layer 1 neurons have not begun to carry information about border ownership at this point. However, as the visual signals propagate up to Layer 3 and back down again to Layer 1, these top down signals from Layer 3 begin to strongly modulate the activities of Layer 1 neurons at around 50 ms. The effect of this

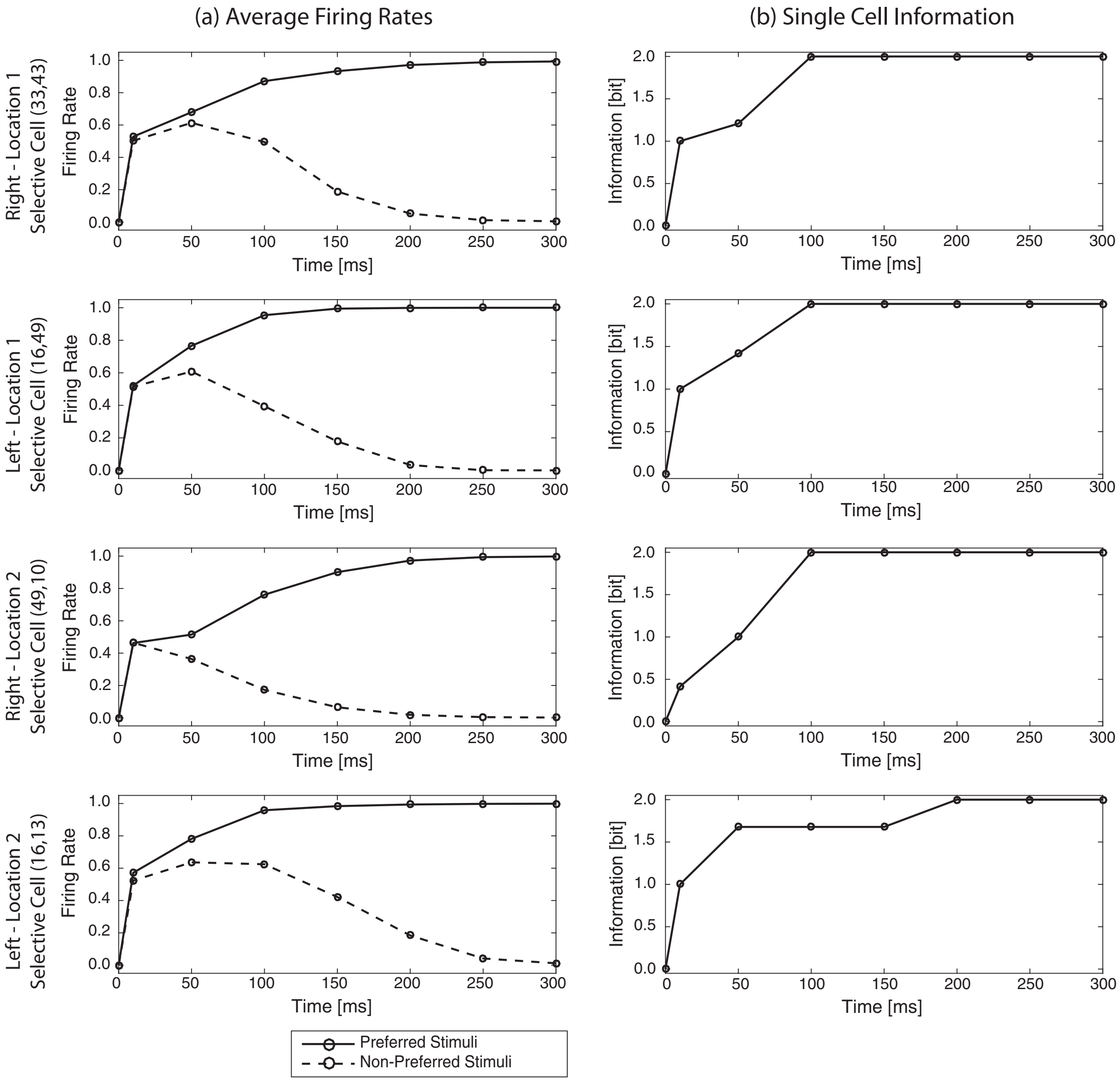

Fig. 9. The temporal evolution of border ownership representations conveyed by four typical Layer 1 neurons during the 300 ms time course of stimulus presentations. The network was trained and tested with the objects shown in Fig. 4. Results are shown after training. Each row shows results for one of the four neurons, where each neuron is tuned to a different border ownership category as follows. Row 1 (top row): a neuron tuned to a vertical straight edge on the right object boundary aligned with retinal Location 1, Row 2: a neuron tuned to a vertical straight edge on the left object boundary aligned with retinal Location 1, Row 3: a neuron tuned to a vertical straight edge on the right object boundary aligned with retinal Location 2, and Row 4 (bottom row): a neuron tuned to a vertical straight edge on the left object boundary aligned with retinal Location 2. Column (a) shows the average firing rates of the four neurons plotted over the 300 ms time courses of the stimulus presentations. Each subplot shows the average responses of the neuron to the members of its preferred stimulus category (solid line) and the members of its three non-preferred stimulus categories (dashed line). For all four neurons, it can be seen that their firing responses begin to strongly differentiate between the preferred and non-preferred stimulus categories at around 50 ms. By the end of the stimulus presentation at 300 ms, the neurons show complete differentiation between the preferred and non-preferred stimulus categories. Column (b) shows the average single cell information carried by the four neurons about their preferred stimulus category plotted over the 300 ms time courses of the stimulus presentations.

top down modulation is to drive the activity of the Layer 1 neurons to represent the border ownership categories. These simulation results are qualitatively similar to the temporal evolution of border ownership representations reported by Sugihara et al. (2011) and Jehee et al. (2007).

3.2. Study 2: failure of the model under more general stimulus conditions

In the above simulations, we have tested the model by presenting a single object to the network at a time. However, the primate visual system is usually presented with multiple objects simultaneously in real world scenes. This more realistic situation actually exposes a weakness in our current rate-coded model. As we explained earlier, Pasupathy and Connor (2002) have reported that the local boundary representations observed in area V4 such as Φ^{V4}_{Left} and Φ^{V4}_{Right} are translation invariant across different retinal positions over a modest range. This may lead to a lack of specificity with respect to retinal location in the contextual information that is back-projected to the earlier layers of the network. This will be problematic, for example, when two objects that contain a straight vertical contour on different object sides (left or right) are presented to the network simultaneously. In this case, both Φ^{V4}_{Left} and Φ^{V4}_{Right} will be activated in the higher V4 layer. However, Φ^{V4}_{Left} and

Φ^{V4}_{Right} will top-down modulate V1/V2 simple cells representing a vertical straight edge on the left and right object boundaries, respectively, across *all* trained retinal locations. Thus, the top-down modulation of V1/V2 neuronal firing rates is not specific to retinal location. This effectively destroys the local border ownership (binding) information carried by the V1/V2 neurons. We elaborate this important argument in more detail next.

3.2.1. Proposed mechanism by which border ownership information carried by V1/V2 neurons in the rate-coded model may be lost when the network is presented with multiple visual objects

Suppose that, during testing of the model, an object that contains a straight vertical contour on the left is presented with that contour positioned at a retinal Location 1, and another object that contains a straight vertical contour on the right is presented with that contour at a retinal Location 2 as shown in Fig. 10. In this case, as explained in the figure, both Φ^{V4}_{Left} and Φ^{V4}_{Right} should become highly activated at the same time. However, during training, the subpopulations Φ^{V4}_{Left} and Φ^{V4}_{Right} are trained to respond invariantly as an object is translated across different retinal locations. This means that both Φ^{V4}_{Left} and Φ^{V4}_{Right} each end up with strong bi-directional polysynaptic connections with subpopulations of V1/V2 simple cells representing a straight vertical contour at all trained retinal locations. In this case, the top-down signals from Φ^{V4}_{Left} and Φ^{V4}_{Right} each modulate the responses of V1/V2 simple cells across both retinal Locations 1 and 2.

In this situation, as we have explained earlier, $\Phi^{V1/V2}_{Left,Loc1}$ will become strongly activated by receiving both the feedforward signals that indicate that a straight vertical contour is present at retinal Location 1 and the feedback signals from Φ^{V4}_{Left} that indicate that the straight vertical contour is on the left side of the object. Similarly, $\Phi^{V1/V2}_{Right,Loc2}$ will become strongly activated by receiving both the feedforward signals that indicate that a straight vertical contour is present at retinal Location 2 and the feedback signals from Φ^{V4}_{Right} that indicate that the straight vertical contour is on the right side of the object.

However, the problem is that the other sets of neurons, $\Phi^{V1/V2}_{Left,Loc2}$ and $\Phi^{V1/V2}_{Right,Loc1}$ may also be strongly activated. This is because both Φ^{V4}_{Left} and Φ^{V4}_{Right} have strong bi-directional polysynaptic connections with subpopulations of V1/V2 simple cells representing a straight vertical contour at both trained retinal Locations 1 and 2. More specifically, $\Phi^{V1/V2}_{Left,Loc2}$ may receive not only the feedforward signals that indicate that a straight vertical contour is present at the retinal location 2, but also the feedback signals from Φ^{V4}_{Left} which are actually activated by the presence of the other object with a straight vertical contour on the left at retinal Location 1. As a result, $\Phi^{V1/V2}_{Left,Loc2}$ may become activated even though no object with a straight vertical contour on the left is ever presented at retinal Location 2. Similarly, $\Phi^{V1/V2}_{Right,Loc1}$ may receive not only the feedforward signals that indicate that the straight vertical contour is present at retinal Location 1, but also the feedback signals from Φ^{V4}_{Right} which are activated by the presence of the other object with a straight vertical contour on the right at retinal Location 2. As a result, $\Phi^{V1/V2}_{Right,Loc1}$ may become activated even though no object with a straight vertical contour on the right is ever presented at retinal Location 1.

The upshot of this is that when the two objects are presented to the model simultaneously, all of the V1/V2 subpopulations $\Phi^{V1/V2}_{Left,Loc1}, \Phi^{V1/V2}_{Right,Loc1}, \Phi^{V1/V2}_{Left,Loc2}$ and $\Phi^{V1/V2}_{Right,Loc2}$ may become active. In this case, these subpopulations of V1/V2 neurons will fail to represent the border ownership (binding) information. This will be a general problem for the current rate-coded formulation of the model when presented with visual input from more realistic scenes containing multiple objects.

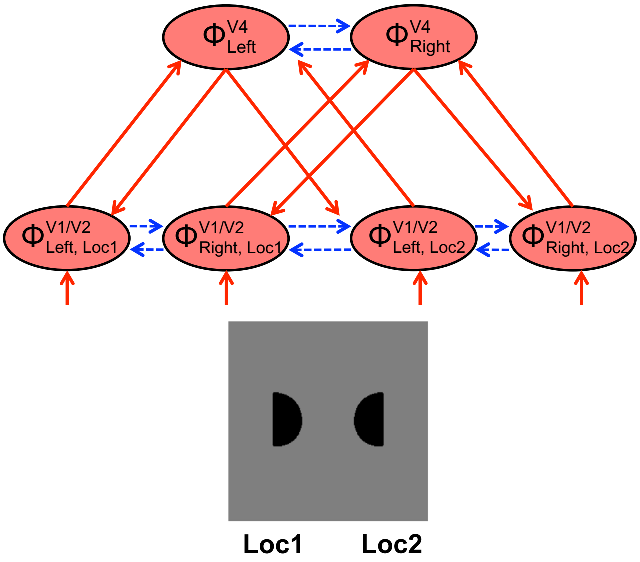

Fig. 10. Hypothesised modulation of edge detecting simple cells in lower layers V1/V2 of the rate-coded model by top-down signals from higher layer V4 neurons representing boundary contour elements when two visual object stimuli are presented simultaneously. Assume that during testing of the model, an object with a straight vertical border on its *left* is presented with this border positioned at retinal location *1*, and another object with a straight vertical border on its *right* is presented with this border positioned at retinal location *2*. Ascending visual input initially stimulates all subsets of V1/V2 neurons, which represent a vertical straight edge at retinal location 1 ($\Phi^{V1/V2}_{Left,Loc1}$ and $\Phi^{V1/V2}_{Right,Loc1}$), and a vertical straight edge at retinal location 2 ($\Phi^{V1/V2}_{Left,Loc2}$ and $\Phi^{V1/V2}_{Right,Loc2}$). In layer V4, V4 neurons that represent a vertical straight edge on the left of an object (Φ^{V4}_{Left}) are stimulated by ascending visual signals from the object in retinal Location 1, while V4 neurons that represent a vertical straight edge on the right of an object (Φ^{V4}_{Right}) are stimulated by ascending visual signals from the object in retinal Location 2. However, the subpopulations Φ^{V4}_{Left} and Φ^{V4}_{Right} have each been trained to respond with translation invariance across all trained retinal locations, and so have developed strong bi-directional (i.e. bottom-up and top-down) polysynaptic connections with subpopulations of V1/V2 simple cells representing all retinal locations. Consequently, Φ^{V4}_{Left} and Φ^{V4}_{Right} will top-down modulate V1/V2 simple cells representing a vertical straight edge on the left and right object boundaries, respectively, across all trained retinal locations. In this case, all of the V1/V2 cells shown in the figure end up receiving a similar amount of bottom-up and top-down excitatory input. Both subpopulations $\Phi^{V1/V2}_{Left,Loc1}$ and $\Phi^{V1/V2}_{Right,Loc1}$ will be active at retinal Location 1, and both subpopulations $\Phi^{V1/V2}_{Left,Loc2}$ and $\Phi^{V1/V2}_{Right,Loc2}$ will be active at retinal Location 2. Thus, when more than one visual object is presented to the model, the V1/V2 neurons $\Phi^{V1/V2}_{Left,Loc1}, \Phi^{V1/V2}_{Right,Loc1}, \Phi^{V1/V2}_{Left,Loc2}$ and $\Phi^{V1/V2}_{Right,Loc2}$ may fail to represent the border ownership (binding) information.

3.2.2. Results

In this section, the model was trained with the set of objects shown in Fig. 4, where these objects were presented to the network one at a time during training as described in the simulations above. However, the network was then tested with *two* objects shown together during each visual presentation, where we used the set of test images shown in Fig. 11. We analysed the steady state firing responses of Layer 1 neurons at the end of each such visual presentation. We compared these results with those reported above in which only a single object was presented to the network at a time during testing. In order to facilitate comparison of the results for the two test situations, in each case we analysed how much information Layer 1 neurons carried about border ownership stimulus categories (straight vertical edges on the left or right object boundaries) that were associated with retinal Location 1.

The set of images used for testing the network with two objects at a time are shown in Fig. 11. There are two different stimulus categories. The first stimulus category, shown in Fig. 11(a), consists of all possible combinations two objects where one of the objects has a vertical straight edge on its *right* boundary which is positioned at retinal Location 1. On the other hand, the second stimulus category, shown in Fig. 11(b), consists of all possible combinations two objects where one of the objects has a vertical straight edge on its *left* boundary which is positioned at retinal Location 1. Each of the two stimulus categories undergoes 16 transforms, which are due to variations in the following four stimulus features: 2 different shapes (semicircle or hexagon) at retinal Location 1 × 2 different shapes (semicircle or hexagon) at retinal Location 2 × 2 sides of an object (left or right) on which a straight vertical edge may occur at retinal Location 2 × 2 kinds of shading contrast between objects and background.

Fig. 12 compares the border ownership information carried by Layer 1 neurons (corresponding to visual areas V1/V2) when the network is tested with objects shown individually (solid line) or when tested on two objects presented together (dashed line). We assess the performance of the network using single-cell information analysis. The information analysis is applied to the steady state firing responses of Layer 1 neurons at the end of each stimulus presentation.

The solid lines in Fig. 12 shows the performance of the model when tested with the objects shown in Fig. 4 presented one at a time during testing. We computed the single cell information carried by each Layer 1 neuron about one of the four stimulus categories separately while we previously plotted them altogether in Fig. 7(a). That is, we computed the information about whether the vertical straight edge in the object stimulus was from one of the following four stimulus categories: (i) a vertical straight edge on the left object boundary positioned at retinal Location 1, (ii) a vertical straight edge on the right object boundary positioned at retinal Location 1, (iii) a vertical straight edge on the left object boundary positioned at retinal Location 2, and (iv) a vertical straight edge on the right object boundary positioned at retinal Location 2. Since there are four stimulus categories, perfectly discriminating neurons carry a maximum of 2 bits of information. However, in this figure the maximum single-cell information has been rescaled to 1. We plot results for the two stimulus categories (i) and (ii), which are associated with retinal Location 1. The information carried by Layer 1 neurons about stimulus category (i) is shown in the right plot, while information carried by Layer 1 neurons about the stimulus category (ii) is shown in the left plot. It can be seen that when the network is tested on a single object at a time, a large number of Layer 1 neurons (i.e. 55 neurons and 71 neurons for the stimulus category (i) and (ii), respectively) reach the theoretical maximum level of information about border ownership.

The dashed lines in Fig. 12 shows the performance of the model when tested with *two* objects shown together during each visual presentation using the test images shown in Fig. 11. Here we computed the single cell information carried by the Layer 1 neurons about the two stimulus categories described in Fig. 11, which are both associated with retinal Location 1. The first stimulus category includes all combinations of two objects where one of the objects has a vertical straight edge on its *right* boundary which is positioned at retinal Location 1, while the second stimulus category includes all combinations of two objects where one of the objects has a vertical straight edge on its *left* boundary positioned at retinal Location 1. Since there are two stimulus categories, neurons may carry up to a maximum of 1 bit of information. The information carried by Layer 1 neurons about the first stimulus category is shown in the left plot, while information carried by Layer 1 neurons about the second stimulus category is shown in the right plot. It can be seen that when the network is tested on two objects at a time, there is a large drop in the levels of single cell information carried by Layer 1 neurons about border ownership compared to when the network is tested on individual objects, with far fewer Layer 1 neurons reaching the theoretical maximum of 1 bit for this test case. This result supported our above prediction that border ownership information carried by Layer 1 (V1/V2) neurons in the rate-coded model may be lost when the network is presented with multiple visual objects during testing.

It can be seen in Fig. 12, however, that a small number of Layer 1 neurons did still reach the maximum of 1 bit of information (19 neurons for both stimulus categories (i) and (ii)). How might this

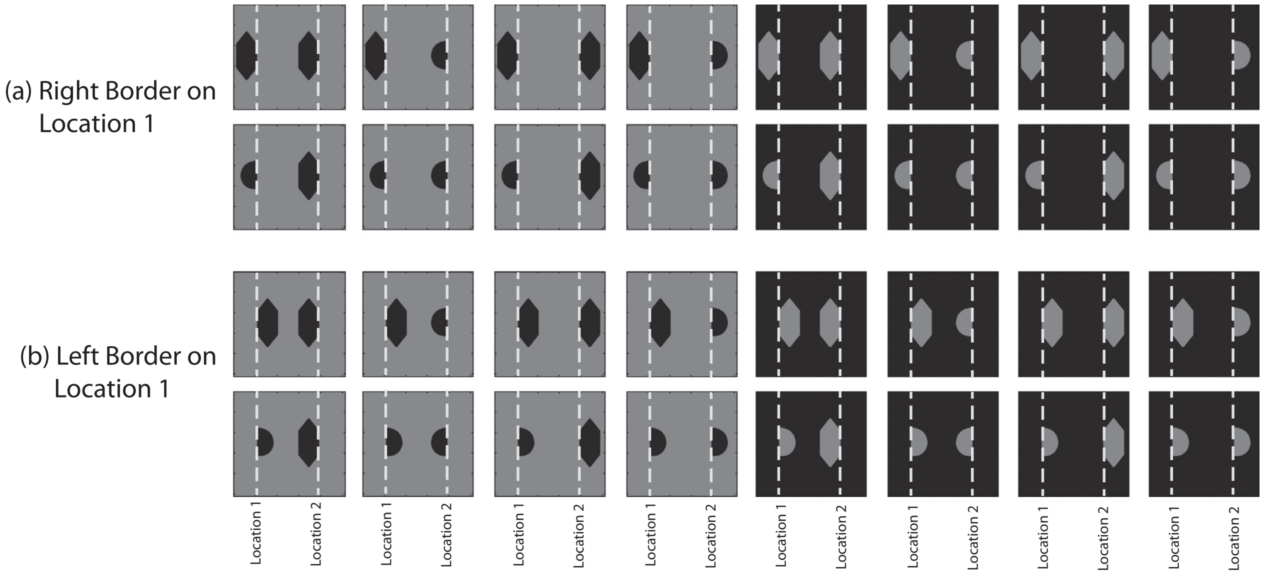

Fig. 11. The set of visual stimuli used to test the performance of the network when two objects are presented simultaneously during testing. There are two categories of visual stimuli. The first stimulus category consists of all possible combinations two objects where one of the objects has a vertical straight edge on its *right* boundary which is positioned at retinal Location 1. The second stimulus category consists of all possible combinations two objects where one of the objects has a vertical straight edge on its *left* boundary which is positioned at retinal Location 1.

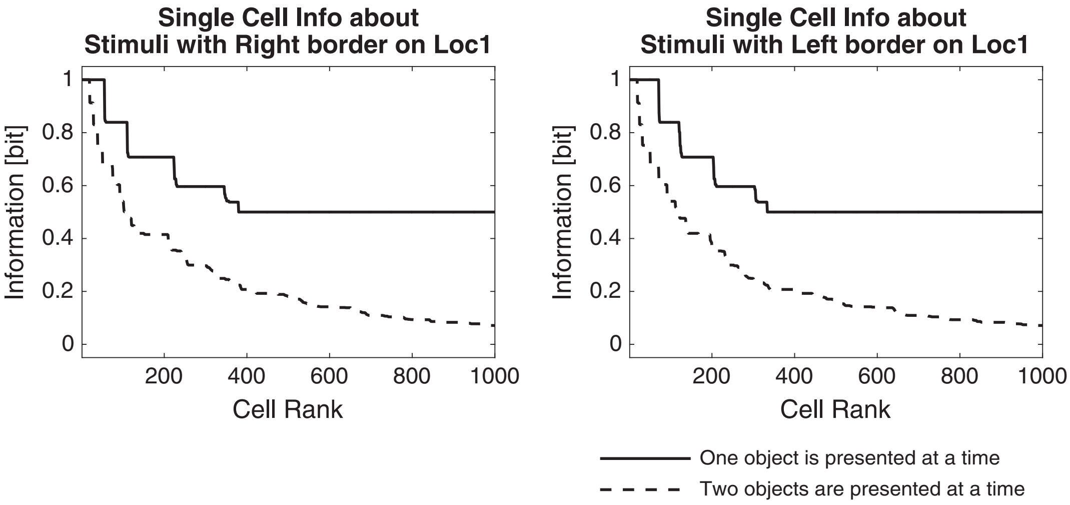

Fig. 12. A quantitative comparison of the border ownership information carried by Layer 1 neurons when the network is tested with objects shown individually (solid line) or when tested on two objects presented together (dashed line). The network performance is assessed using single-cell information analysis. The information analysis is applied to the steady state firing responses of Layer 1 neurons at the end of each stimulus presentation. In both test situations, the model was initially trained with the individual objects shown in Fig. 4, as described earlier in the paper. **Solid lines:** the performance of the model when tested with the objects shown in Fig. 4 presented one at a time during testing. We computed the single cell information carried by each Layer 1 neuron about the four stimulus categories previously described in Fig. 7(a). In this figure the maximum single-cell information ($\log_2(4) = 2$) has been rescaled to 1. The information carried by Layer 1 neurons about stimulus category (i) is shown in the right plot, while information carried by Layer 1 neurons about the stimulus category (ii) is shown in the left plot. It can be seen that when the network is tested on a single object at a time, many Layer 1 neurons reach the theoretical maximum level of information about border ownership. **Dashed lines:** the performance of the model when tested with *two* objects shown together during each visual presentation using the test images shown in Fig. 11. Here we computed the single cell information carried by each Layer 1 neuron about the two stimulus categories described in Fig. 11. Since there are two stimulus categories, neurons may carry up to a maximum of 1 bit of information. The information carried by Layer 1 neurons about the first stimulus category is shown in the left plot, while information carried by Layer 1 neurons about the second stimulus category is shown in the right plot. It can be seen that when the network is tested on two objects at a time, there is a large drop in the levels of single cell information carried by Layer 1 neurons about border ownership compared to when the network is tested on individual objects, with far fewer Layer 1 neurons reaching the theoretical maximum of 1 bit for this test case.

happen if both of the Layer 3 subpopulations Φ^{V4}_{Left} and Φ^{V4}_{Right} were completely translation invariant with strong bi-directional (bottom-up and top-down) polysynaptic connections with subpopulations of Layer 1 (V1/V2) simple cells representing a straight vertical contour at both trained retinal Locations 1 and 2? To understand this, we investigated the firing properties of neurons in Layer 2 after training. Although not shown here, it was found that, due to the limited feedforward fan-in of synaptic connections from the input 'retina', some of the Layer 2 neurons had learned to respond to a vertical straight edge either on the left object boundary or right object boundary at only a single retinal location. These location-specific Layer 2 neurons were then able to directly modulate the Layer 1 neurons representing that particular retinal location. This would allow these Layer 1 neurons to continue to respond selectively to whether a vertical straight edge was on either the left or right boundary of an object presented at that retinal location regardless of the presence of another object simultaneously presented elsewhere. However, this effect was rather minor given that the great majority of Layer 1 neurons lost their border ownership selectivity when two objects were presented during testing.

4. Discussion

We have investigated through computer simulation how top-down connections may play a fundamental role in the development of border ownership representations in the early cortical visual layers V1/V2. In terms of the novelty, this work is different from previous modelling studies that have already proposed hypothetical neural circuits for such coding in that we investigated how such circuits may develop using a biologically plausible, local, trace learning rule to modify the synaptic connectivity during visual experience.

A number of modelling studies have previously considered the role of top-down signals in visual information processing. For example, as discussed in Section 1.1, some authors have proposed that top-down connections might implement attention to objects during visual search (Deco & Lee, 2002; Deco & Rolls, 2004). However, in these previous modelling studies the top-down connections were only introduced after the initial training phase was completed, and hence the self-organisation of the synaptic connections throughout the network relied on purely feedforward visual processing. Consequently, the top-down connections did not affect the visual representations that developed in the network during visually-guided learning. In another modelling study carried out by Renart, Parga, and Rolls (1999), top-down connections were able to influence the recall of visual representations in a linked attractor network comprised of multiple cortical modules (Rolls, 2008). However, the representations in this attractor network were hand specified during an initial stage of supervised learning, and did not self-organise using unsupervised competitive learning. Thus, again, the top-down connections were not able to influence the nature of the visual representations that developed. In our own model presented in this paper, the top-down connections are present during both training and testing. Consequently, the top-down connections played a critical role in the self-organisation of border ownership representations in Layer 1 during the initial unsupervised competitive learning. In this case, each neuron receives signals from both afferent bottom-up and top-down connections, which self-organise simultaneously during learning. This allows the network to develop representations that depend on a precise learned combination of bottom-up and top-down signals.

The simulations reported here have demonstrated how top-down connections may help to guide competitive learning in lower layers, thus driving the formation of lower level (border ownership) visual representations in V1/V2 that are modulated by higher level (object boundary element) representations in V4. More precisely, we have shown that simple cells in area V1 representing a vertical straight edge at a particular retinal location can learn to be modulated by top-down connections from higher level representations of object shape in, for example, area V4 (Pasupathy & Connor, 2001; Pasupathy & Connor, 2002). However, more importantly, we also identified the limitation of the mechanism within a rate-coded model when trying to simulate the results of the neurophysiological studies that have shown that border-ownership selective neurons for single-figure displays generally are so also for multi-figure displays (Martin & von der Heydt, 2015; Qiu, Sugihara, & von der Heydt, 2007). In the second half of the simulation studies, we have investigated how the rate-coded model presented in this paper fails under more general stimulus conditions, in which more than one object stimulus is presented to the network at the same time after training.

The result suggests that the incorporation of additional top-down connections, although necessary, is not sufficient by itself to allow the network to develop robust border ownership representations in the early layers and thus solve this kind of feature binding problem. Our model failed because the current model of the network is not able to specify which features are part of which objects. Therefore, we propose that it is important to have a form of binding neuron (e.g., border ownership neuron in V1/V2) that responds if and only if the neurons representing the low-level feature such as simple oriented bars are actually participating in driving the neurons representing the high-level feature. The binding neuron should not respond if the neurons representing the low-level feature and the neurons representing the high-level feature just happen to be co-active, where the former are not actually driving the latter. Such unrelated co-activation of low and high-level features might occur, for example, because of the presence of multiple similar objects within a complex natural scene. Then, the question is what further biological details is needed to be incorporated into the model to allow it to form such robust border ownership representations under more general stimulus conditions.

A biological detail that is not implemented in the current model is cortical magnification. It is known that mammalian brains process visual input in a highly non-uniform manner. Specifically, the Ganglion cells in the retina sample the visual input at a higher resolution in the fovea than the periphery (Wassle, Grunert, Rohrenbeck, & Boycott, 1990), which gives rise to a distorted visual field representation in V1 where the fovea has a higher "cortical magnification factor", i.e. more V1 neurons processing foveal input than the peripheral visual field (Cowey & Rolls, 1974; Daniel & Whitteridge, 1961). Subsequent neural processing, with a simple Gaussian sampling of the representation that is laid out across the surface of area V1, results in an asymmetry of central V4 receptive fields as well (Motter, 2009). The question is whether cortical magnification may play any role in the development of border ownership representations.

Our laboratory has previously investigated the effects of implementing a cortical magnification factor within a purely feedforward neural network model of primate visual object recognition (Trappenberg, Rolls, & Stringer, 2002). It was found that when the objects were presented against a simple blank background then neurons in the upper cortical layer responded to their preferred objects across a wide region of the retina. In this scenario, trace learning can continue to operate normally as an object translates across different locations on the retina. Neurophysiological evidence for trace learning has been reported by Cox, Meier, Oertelt, and DiCarlo (2005). Moreover, past simulation studies have found that the trace learning mechanism is quite robust to the way in which the eyes saccade around the visual scene, and is in fact enhanced by more randomised exploration of a scene (Rolls & Milward, 2000). Consequently, we would not expect the introduction of a cortical magnification factor into the border ownership simulations reported in this paper to prevent the model from operating in the same qualitative manner as described above. However, in the simulation study of (Trappenberg et al., 2002), it was also found that, with a cortical magnification factor, if the objects were presented against cluttered backgrounds then the receptive fields of neurons in the upper layer shrunk down around the fovea due to competition from the background features. These simulation results reflected what had been previously observed in a primate neurophysiology study carried out by Rolls, Aggelopoulos, and Zheng (2003), in which the receptive fields of object-selective neurons in the primate temporal visual cortex reduced down to approximately the size of the object when it was presented against a natural scene. It should also be noted that the neurophysiology studies of V4 shape selective neurons (Pasupathy & Connor, 2001) investigated the responses of these neurons to shapes that were presented in isolation. Our border ownership model sought to replicate the development of these V4 shape selective firing properties in Layer 3 under similar viewing conditions - that is, the network was trained on one shape at a time presented against a blank background. It remains to be seen how the firing properties of these shape selective neurons in area V4 of the primate brain might be affected when the shapes are presented within natural scenes. The cortical magnification factor may play an important role in this situation. Addressing these issues will require a combination of further neurophysiology and modelling studies.

Another biological detail that is not implemented in the current model is the spike dynamics of neurons. We hypothesise that extending the model with spiking neural network would solve the issue. The current rate-coded model only represents the average firing rate of each neuron, and not the actual timings of the electrical pulses emitted by neurons in the brain. The architecture and operation of neural tissue in the visual cortex of primates differs from the VisNet model implemented in this paper in the following important ways. Firstly, real neurons in the brain communicate by emitting and receiving electrical pulses called action potentials or 'spikes'. Secondly, the way in which synapses are strengthened and weakened during learning is dependent on the timings of the spikes emitted by the pre- and post-synaptic neurons (Bi & Poo, 1998; Markram, Lbke, Frotscher, & Sakmann, 1997). For example, in the brain, a synapse may be strengthened if the pre-synaptic spike occurs about 20 ms before the post-synaptic spike, but weakened if the pre-synaptic spike occurs about 20 ms after the post-synaptic spike. This is known as spike time dependent plasticity (STDP). Thirdly, the electrical pulses can take several milliseconds to travel along an axon from one neuron to the next, with different axonal connections having different time delays.

Physiological studies have shown that neural synchrony is unrelated, or at best weakly related, to contour grouping (Martin & von der Heydt, 2015; Roelfsema, Lamme, & Spekreijse, 2004). On the other hand, if distributions of axonal delays between neurons are incorporated into a model, then this can give rise to a phenomenon known as 'polychronization' (Izhikevich, 2006). This phenomenon involves the network learning many memory patterns, each of which takes the form of a repeating temporal loop of neuronal firings. These temporal memory loops self-organise automatically when STDP is used to modify the strengths of synapses in a recurrently connected spiking network with randomised distributions of axonal conduction delays between neurons. Polychronization can dramatically increase the selectivity of neurons and increase the memory capacity of a network. We hypothesise that such a spiking model may develop border ownership neurons in layer 1 (corresponding to V1/V2) that respond selectively to a vertical straight edge on either the left or right boundary of an object at the neuron's preferred retinal location, regardless of the presence of other objects at different retinal locations. More generally, we propose that these biological elements will be needed to model how the primate visual system solves 'the binding problem' in vision. Consequently, in future work we will explore how border ownership representations may develop in a new spiking neural network version of the VisNet model, which incorporates bottom-up and top-down connections, distributions of axonal transmission delays, and spike time dependent plasticity (STDP).

References

Angelucci, A., & Bullier, J. (2003). Reaching beyond the classical receptive field of V1 neurons: Horizontal or feedback axons? *Journal of Physiology-Paris, 97*(23), 141–154.

Baek, K., & Sajda, P. (2005). Inferring figure-ground using a recurrent integrate-and-fire neural circuit. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13*(2), 125–130.

Bi, G.-q., & Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience, 18*(24), 10464–10472.

Cowey, A., & Rolls, E. T. (1974). Human cortical magnification factor and its relation to visual acuity. *Experimental Brain Research, 21*(5), 447–454.

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). 'Breaking' position-invariant object recognition. *Nature Neuroscience, 8*(9), 1145–1147. PMID: 16116453.

Craft, E., Schtze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology, 97*(6), 4310–4326.

Cumming, B. G., & Parker, A. J. (1999). Binocular neurons in v1 of awake monkeys are selective for absolute, not relative, disparity. *The Journal of Neuroscience, 19* (13), 5602–5618. PMID: 10377367.

Daniel, P. M., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of Physiology, 159*(2), 203–221.

Deco, G., & Lee, T. (2002). *A unified model of spatial and object attention based on inter-cortical biased competition*. Computer Science Department.

Deco, G., & Rolls, E. T. (2004). A neurodynamical cortical model of visual attention and invariant object recognition. *Vision Research, 44*(6), 621–642.

Eguchi, A., Mender, B. M. W., Evans, B., Humphreys, G., & Stringer, S. (2015). Computational modeling of the neural representation of object shape in the primate ventral visual system. *Frontiers in Computational Neuroscience, 9*(100), 100.

Foldiak, P. (1991). Learning invariance from transformation sequences. *Neural Computation, 3*(2), 194–200.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14*(9), 1195–1201.

Gross, C. G., Bender, D. B., & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science, 166*(3910), 1303–1306. PMID: 4982685.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation, 18*(2), 245–282.

Jehee, J. F. M., Lamme, V. A. F., & Roelfsema, P. R. (2007). Boundary assignment in a recurrent network architecture. *Vision Research, 47*(9), 1153–1165.

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*(6), 1187–1211. PMID: 3437330.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43*(1), 59–69.

Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R., & Konen, W. (1993). Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers, 42*(3), 300–311.

Layton, O. W., Mingolla, E., & Yazdanbakhsh, A. (2012). Dynamic coding of border-ownership in visual cortex. *Journal of Vision, 12*(13). 8–8.

Markram, H., Lbke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science (New York, N.Y.), 275*(5297), 213–215.

Martin, A. B., & von der Heydt, R. (2015). Spike synchrony reveals emergence of proto-objects in visual cortex. *The Journal of Neuroscience, 35*(17), 6860–6870.

Motter, B. C. (2009). Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant sized sampling of the V1 surface. *The Journal of neuroscience: The official journal of the Society for Neuroscience, 29* (18), 5749–5757.

Nishimura, H., & Sakai, K. (2004). Determination of border ownership based on the surround context of contrast. *Neurocomputing, 5860*, 843–848.

Pasupathy, A. (2006). Neural basis of shape representation in the primate brain. *Progress in Brain Research, 154*, 293–313. PMID: 17010719.

Pasupathy, A., & Connor, C. E. (2001). Shape representation in area v4: Position-specific tuning for boundary conformation. *Journal of Neurophysiology, 86*(5), 2505–2519.

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area v4. *Nature Neuroscience, 5*(12), 1332–1338.

Petkov, N., & Kruizinga, P. (1997). Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: Bar and grating cells. *Biological cybernetics, 76*(2), 83–96. PMID: 9116079.

Pettet, M. W., & Gilbert, C. D. (1992). Dynamic changes in receptive-field size in cat primary visual cortex. *Proceedings of the National Academy of Sciences, 89*(17), 8366–8370. PMID: 1518870.

Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nature Neuroscience, 10*(11), 1492–1499.

Renart, A., Parga, N., & Rolls, E. T. (1999). Associative memory properties of multiple cortical modules. *Network: Computation in Neural Systems, 10*(3), 237–255.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience, 7*(9), 982–991.

Rolls, E. T. (2000). Functions of the primate temporal lobe cortical visual areas in invariant visual object and face recognition. *Neuron, 27*(2), 205–218. PMID: 10985342.

Rolls, E. (2008). *Memory, attention, and decision-making: A unifying computational neuroscience approach* (1st ed.). Oxford University Press.

Rolls, E. T. (2012). Invariant visual object and face recognition: Neural and computational bases, and a model, VisNet. *Frontiers in Computational Neuroscience, 6*, 35.

Rolls, E. T., Aggelopoulos, N. C., & Zheng, F. (2003). The receptive fields of inferior temporal cortex neurons in natural scenes. *The Journal of Neuroscience, 23*(1), 339–348.

Rolls, E. T., & Milward, T. (2000). A model of invariant object recognition in the visual system: Learning rules, activation functions, lateral inhibition, and information-based performance measures. *Neural Computation, 12*(11), 2547–2572. PMID: 11110127.

Rolls, E. T., Treves, A., Tovee, M. J., & Panzeri, S. (1997). Information in the neuronal representation of individual stimuli in the primate temporal visual cortex. *Journal of computational neuroscience, 4*(4), 309–333.

Royer, S., & Paré, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature, 422*(6931), 518–522. PMID: 12673250.

Rubin, E. (1915). *Synsoplevede figurer* (PhD thesis). Copenhagen: University of Copenhagen.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9*(1), 75–112.

Sugihara, T., Qiu, F. T., & von der Heydt, R. (2011). The speed of context integration in the visual cortex. *Journal of Neurophysiology, 106*(1), 374–385.

Trappenberg, T. P., Rolls, E. T., & Stringer, S. M. (2002). Effective size of receptive fields of inferior temporal visual cortex neurons in natural scenes. *Advances in Neural Information Processing Systems, 1*, 293–300.

Tromans, J. M., Harris, M., & Stringer, S. M. (2011). A computational model of the development of separate representations of facial identity and expression in the primate visual system. *PLoS ONE, 6*(10), e25616.

von der Heydt, R., Zhou, H., & Friedman, H. S. (2003). Neural coding of border ownership: Implications for the theory of figure-ground perception. *Perceptual organization in vision: Behavioral and neural perspectives*, 281–304.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik, 14*(2), 85–100.

Wagatsuma, N., Oki, M., & Sakai, K. (2013). Feature-based attention in early vision for the modulation of figure-ground segregation. *Frontiers in Psychology, 4*.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51*(2), 167–194.

Wassle, H., Grunert, U., Rohrenbeck, J., & Boycott, B. B. (1990). Retinal ganglion cell density and cortical magnification factor in the primate. *Vision Research, 30*(11), 1897–1911.

Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area V2. *Neuron, 47*(1), 143–153.

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience, 20*(17), 6594–6611.