Psychological Review
2017, Vol. 124, No. 2, 154–167

0033-295X/17/$12.00 http://dx.doi.org/10.1037/rev0000049

The Neural Representation of the Gender of Faces in the Primate Visual System: A Computer Modeling Study

Thomas Minot, Hannah L. Dury, Akihiro Eguchi, Glyn W. Humphreys, and Simon M. Stringer
University of Oxford

We use an established neural network model of the primate visual system to show how neurons might learn to encode the gender of faces. The model consists of a hierarchy of 4 competitive neuronal layers with associatively modifiable feedforward synaptic connections between successive layers. During training, the network was presented with many realistic images of male and female faces, during which the synaptic connections are modified using biologically plausible local associative learning rules. After training, we found that different subsets of output neurons have learned to respond exclusively to either male or female faces. With the inclusion of short range excitation within each neuronal layer to implement a self-organizing map architecture, neurons representing either male or female faces were clustered together in the output layer. This learning process is entirely unsupervised, as the gender of the face images is not explicitly labeled and provided to the network as a supervisory training signal. These simulations are extended to training the network on rotating faces. It is found that by using a trace learning rule incorporating a temporal memory trace of recent neuronal activity, neurons responding selectively to either male or female faces were also able to learn to respond invariantly over different views of the faces. This kind of trace learning has been previously shown to operate within the primate visual system by neurophysiological and psychophysical studies. The computer simulations described here predict that similar neurons encoding the gender of faces will be present within the primate visual system.

Keywords: ventral visual pathway, face processing, gender recognition, neural network model, trace learning

There is evidence for separate subpopulations of neurons in the primate visual system that encode face images according to identity or expression (Hasselmo, Rolls, & Baylis, 1989; Perrett, Hietanen, Oram, & Benson, 1992). Given the assumption in neuropsychology that gender categorization is separate from identity or expression recognition, it is possible that other subpopulations exist that distinguish between faces along other dimensions (e.g., Bruce & Young, 1986; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003; Zhao, Chellappa, Phillips, & Rosenfeld 2003). Indeed, a small number of functional magnetic resonance imaging (fMRI) studies have found gender-selective neurons in the fusiform face area (FFA; Contreras, Banaji, & Mitchell, 2013; Kaul, Rees, & Ishai, 2011; Ng, Ciaramitaro, Anstis, Boynton, & Fine, 2006; Podrebarac, Goodale, van der Zwan, & Snow, 2013). Freeman, Rule, Adams, and Ambady (2010) investigated categorical representations of gender with fMRI, finding activity in the fusiform gyrus and FFA in response to objective differences between genders, and orbitofrontal cortex activity during subjective decisions about the gender of a face.

The majority of evidence for these subpopulations has arisen from psychophysical studies, however, in which perceptual aftereffects from adaptation to a stimulus are thought to be represented by selective neural populations (Fang & He, 2005). Adaptation studies show that prior viewing of a particular category biases subsequent judgments of neutral exemplars against the initial category. For example, viewing a downward flowing waterfall results in stationary objects appearing to move upward (Frisby, 1979). With faces, prior viewing of a face with distorted features causes subsequent distorted faces to be judged as more normal (Rhodes et al., 2003; Rhodes, Jeffery, Watson, Winkler, & Clifford, 2004). Regarding gender, adaptation to faces of a particular gender induces perceptual aftereffects to faces of the same sex (Baudouin & Brochard, 2011; Bestelmeyer et al., 2008; Jones, DeBruine, Little, & Welling, 2010; Little, DeBruine, & Jones, 2005). Two distinct aftereffects reflect two distinct perceived categories—male and female—which are assumed to be represented by separate subpopulations of neurons in the ventral visual system (Sergent, Ohta, & Macdonald, 1992).

This article was published Online First January 9, 2017.

Thomas Minot, Oxford Centre for Theoretical Neuroscience and Artificial Intelligence, University of Oxford; Hannah L. Dury and Akihiro Eguchi, Oxford Centre for Theoretical Neuroscience and Artificial Intelligence and Department of Experimental Psychology, University of Oxford; Glyn W. Humphreys, Department of Experimental Psychology, University of Oxford; Simon M. Stringer, Oxford Centre for Theoretical Neuroscience and Artificial Intelligence and Department of Experimental Psychology, University of Oxford.

All authors except Glyn W. Humphreys have contributed toward the manuscript and read and approved the final manuscript. Glyn W. Humphreys was involved in preparing the originally submitted manuscript but died during the subsequent review process. All authors declare no conflicts of interest relevant to this article.

Correspondence concerning this article should be addressed to Hannah L. Dury, Department of Experimental Psychology, University of Oxford, Tinbergen Building, 9 South Parks Road, Oxford OX1 3UD, United Kingdom. E-mail: hannah.dury@psy.ox.ac.uk

Using a biologically plausible model of the primate ventral visual stream, known as VisNet, we describe for the first time how such subpopulations of neurons might arise. The primary aim of this study is to provide a computational example of how the brain might develop separate neural representations of male and female faces. The VisNet model has previously been shown, given particular inputs, to develop clusters of neurons in the output layer of the network that respond exclusively to a particular face-related category, such as face identity or facial expression (Eguchi, Humphreys, & Stringer, 2016; Tromans, Harris, & Stringer, 2011). In the present study, realistic face images are presented to the model, and it is expected that similar clusters of neurons will develop, responding only to female faces or male faces. A second simulation investigates whether these clusters can also develop after presentation of faces that have been rotated. Experience tells us that the human visual system would usually be able to categorize the gender of person not just when facing them head-on, but also turned to the side and in profile, so we expect the model to do the same.

A number of computational models that can categorize faces based on gender have already been developed. Two such commonly used models are support vector machines (SVM) and principal component analysis (PCA; for summaries see Moghaddam & Yang, 2000; Sun, Bebis, Yuan, & Louis, 2002). SVM finds the optimal hyperplane that correctly separates the largest number of data points, which in the case of gender, gives two separable classes divided by the hyperplane. The hyperplane is then applied to new images, giving highly accurate gender categorization (Moghaddam & Yang, 2000). PCA produces a set of eigenvectors from training images that explain the maximum possible variance between images (Sun et al., 2002). These eigenvectors can be thought of as a set of features representing gender, to which new faces are compared in order to correctly categorize their gender.

The key difference between the current model and these previous models is the focus on biological plausibility, as learning is unsupervised within a biologically realistic neural network architecture. The model is never explicitly informed which face image belongs to which category (i.e., male or female) either during training or testing, but instead exploits the statistics of the images to create two separable categories. Rather than relying on supervised learning, VisNet requires no gender labeling of the face images seen during training and updates its synaptic weights locally using a Hebbian-like learning rule. In Simulation 1, a basic Hebb rule is applied to the model. In Simulation 2, a trace learning rule is implemented in order for the model to associate successive views of a rotating face (Foldiak, 1991; Rolls, Cowey, & Bruce, 1992).

Hypotheses

We hypothesized that neurons could learn to respond selectively to either male faces or female faces, regardless of other facial attributes such as identity, through a biologically plausible process of unsupervised competitive learning. Competitive learning relies on inhibitory interactions between neurons within a neuronal layer, which in the brain will be implemented by inhibitory interneurons, to effect competition between neurons. Then a local Hebb (associative) learning rule is used to modify the strengths of the feedforward excitatory connections between successive neuronal layers as images are presented to the network during training. Competitive learning drives the development of neurons that respond selectively to particular stimulus categories, where the members of a stimulus category will share some similar features that can be used to bind those stimuli together. It has previously been shown in VisNet simulations that this kind of competitive learning architecture can encourage individual neurons in higher layers to learn to respond selectively to stimulus categories such as facial identity or expression (Eguchi et al., 2016; Tromans et al., 2011). Both identity and face expression were found to be determined by the spatial relationships between facial features.

Likewise, the fact that humans can usually determine a person's gender from a face image implies that there must be certain geometric features within a face that allow us to discriminate between genders. Gosselin and Schyns (2001) have previously used eye-tracking to determine that humans find the eye and mouth regions particularly important in gender categorization. We therefore hypothesized that the same competitive learning process may exploit the geometric similarities between male faces and corresponding similarities between female faces to produce neurons that learned to encode facial gender. That is, some neurons would learn to respond selectively to all male faces, while other neurons would respond to all female faces. In our simulations described below, tracing the strengthened synaptic connections from the input retina to the gender discriminating neurons in the output layer allowed us to see which geometric features of faces differentiates between gender. The hypothesized process of competitive learning is unsupervised in the sense that no external teacher is needed to label the gender of the faces during training in order to guide learning of gender categories. This is essential to the biological plausibility of the simulations.

We also hypothesized that the incorporation of short range excitatory synaptic connections within each neuronal layer to implement a self-organizing map (SOM) architecture (Kohonen, 1982; Von der Malsburg, 1973) would drive the development of map-like firing properties within the higher layers. In this case, neurons responding to male faces would be physically clustered together, and neurons responding to female faces would also be clustered together. This was indeed found to be the case in the first simulation experiment reported below.

If a person turns their head to the side, we do not suddenly lose the ability to determine their gender. We hypothesized that neurons responding selectively to either male or female faces would also learn to respond invariantly to different views of a rotating face if the network is trained using a trace learning rule (Foldiak, 1991; Wallis & Rolls, 1997). Such learning rules incorporate a memory trace of recent neuronal activity, which encourages neurons in higher layers to bind together visual stimuli that tend to occur close together in time. It is assumed that in the natural visual world, the different views of a particular face are usually seen in temporal proximity, for example, during head rotation. In this case, trace learning will bind these different views of a face together onto the same subset of output neurons, developing an invariant representation of that face.

This kind of trace learning has been previously demonstrated to operate within the primate visual system by neurophysiological studies in macaques (Li and DiCarlo, 2008; Rolls et al., 1992). Temporal continuity produced by sequential presentation of stimuli can facilitate learning in humans, giving psychophysical evi-

dence of trace learning (Perry, Rolls, & Stringer, 2006). Trace learning is also unsupervised in that the network is not provided with an explicit training signal to label the gender or any other attribute of the faces during training, which is important to maintain biological plausibility. Trace learning has been incorporated within previous VisNet simulations, in which neurons learned to respond to different views of rotating faces (Wallis & Rolls, 1997). We hypothesized that binding different views of rotating faces together by trace learning would drive the development of view invariant neuronal responses. This could operate in tandem with the aforementioned process of competitive learning, in which different neurons learn to respond to either male or female faces. We therefore ran a second simulation experiment, in which the male and female faces were each presented three times; facing straight ahead, turned 45°, and turned 90°. By implementing a trace learning rule in place of the Hebb rule, we hypothesized that the model would be able to associate the three rotated views as belonging to the same male or female face. In this case, individual neurons should learn to respond selectively to either male or female faces even when seen across the three different views. We expect the same simulation using a Hebb rule to fail to learn invariant representations of gender.

Method

Model Architecture

The simulation studies presented below are conducted with an established biologically plausible neural network model, VisNet, of the primate ventral visual pathway shown in Figure 1. The standard network architecture consists of a hierarchy of four competitive neural layers corresponding to successive stages of the ventral visual pathway. During training with visual objects, the strengths of the feed-forward synaptic connections between successive neuronal layers are modified by biologically plausible local learning rules, where the change in the strength of a synapse depends on the current or recent activities of the pre- and postsynaptic neurons. A variety of such learning rules may be implemented with different learning properties.

One simple well-known learning rule is the Hebb rule:

$$\delta w_{ij} = k r_i^{\tau} r_j^{\tau} \tag{1}$$

where δw_{ij} is the change of synaptic weight w_{ij} from presynaptic neuron j to postsynaptic neuron i, $r^{\tau}{}_i$ is the firing rate of postsynaptic neuron i at timestep τ, $r^{\tau}{}_j$ is the firing rate of presynaptic neuron j at timestep τ, and k is the learning rate constant. Alternatively, a trace learning rule (Foldiak, 1991; Wallis & Rolls, 1997) may be implemented, which incorporates a memory trace of recent neuronal activity:

$$\delta w_{ij} = k \bar{r}_i^{\tau-1} r_j^{\tau} \tag{2}$$

where $\bar{r}_i^{\tau}$ is the trace value of the firing rate of postsynaptic neuron i at timestep τ. The trace term is updated at each timestep according to

$$\bar{r}_i^{\tau} = (1 - \eta) r_i^{\tau} + \eta \bar{r}_i^{\tau-1} \tag{3}$$

where η may be set anywhere in the interval [0, 1], and for the simulations described below, η was set to 0.8. The effect of this learning rule is to encourage neurons to learn to respond to visual input patterns that tend to occur close together in time. If the different views of a particular male or female face are seen in temporal proximity, for example, as the head rotates, then the trace learning rule will bind these different views onto the same subset of output neurons leading to view-invariant neuronal representations.

To prevent the same few neurons always winning the competition, the synaptic weight vectors are normalized to unit length after each learning update for each training pattern. Neurophysiological evidence for synaptic weight normalization is provided by Royer and Para (2003).

VisNet is a hierarchical neural network model of the primate ventral visual pathway, which was originally developed by Wallis and Rolls (1997). The standard network architecture is shown in Figure 1. It is based on the following: (a) A series of hierarchical competitive networks with local graded lateral inhibition; (b) Convergent connections to each neuron from a topologically corresponding region of the preceding layer, leading to an increase in

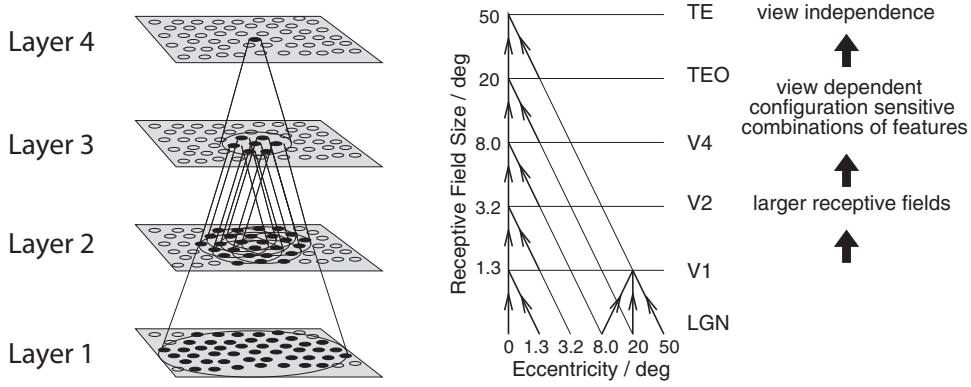

Figure 1. Left: Stylized image of the four-layer VisNet architecture. Convergence through the network is designed to provide fourth layer neurons with information from across the entire input retina. Right: Convergence in the visual system. V1: visual cortex area V1; TEO = posterior inferior temporal cortex; TE = inferior temporal cortex (IT).

the receptive field size of neurons through the visual processing areas; and (c) Synaptic plasticity based on a local associative learning rule such as the Hebb rule or trace rule.

In past work, the hierarchical series of 4 neuronal layers of VisNet have been related to the following successive stages of processing in the ventral visual pathway: V2, V4, the posterior inferior temporal cortex, and the anterior inferior temporal cortex. However, this correspondence has always been quite loose because the ventral pathway may be further subdivided into a more fine-grained network of distinct subregions.

The forward 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 that 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 reflects the known physiology of the primate ventral visual pathway (Freeman & Simoncelli, 2011; Pasupathy, 2006; Pettet & Gilbert, 1992).

Preprocessing of the Visual Input by Gabor Filters

Before the visual images are presented to VisNet's input layer 1, they are preprocessed by a set of input filters that accord with the general tuning profiles of simple cells in area V1. The filters provide a unique pattern of filter outputs for each transform of each face, which is passed through to the first layer of VisNet. In this paper, the input filters used are Gabor filters. 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 within the visual field (Cumming & Parker, 1999; Jones & Palmer, 1987). The input filters used are computed by the following equations:

$$g(x, y, \lambda, \theta, \psi, b, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos\left(2\pi\frac{x'}{\lambda} + \psi\right) \tag{4}$$

with the following definitions:

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta \\ y' &= -x\sin\ \theta + y\cos\ \theta \\ \sigma &= \frac{\lambda(2^b+1)}{\pi(2^b-1)}\sqrt{\frac{\ln 2}{2}} \end{aligned} \tag{5}$$

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 2.

Table 1
Dimensions of the Four Neuronal Layers of VisNet and Synaptic Connectivity Between Successive Layers

Layer	Dimensions	Number of connections	Radius
Layer 4	256 × 256	200	24
Layer 3	256 × 256	200	18
Layer 2	256 × 256	200	12
Layer 1	256 × 256	201	12
Retina	256 × 256 × 16		

Note. The columns show the number of neurons within each layer, the number of afferent feedforward synaptic connections per neuron, and the radius in the preceding layer from which 67% of the connections are received.

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 localized within a topologically corresponding region of the retina.

Calculation of Cell Activations Within the Network

Within each of the neural layers 1 to 4 of the network, the activation h_i of each neuron i is set equal to a linear sum of the inputs r_j from afferent neurons j in the preceding layer weighted by the synaptic weights w_{ij}. That is,

$$h_i = \sum_j w_{ij} r_j \tag{6}$$

where r_j is the firing rate of neuron j, and w_{ij} is the strength of the synapse from neuron j to neuron i.

SOM

In this paper, we have run simulations with a 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:

$$I_{a,b} = -\delta_I \exp\left(-\frac{a^2+b^2}{\sigma_I^2}\right) + \delta_E \exp\left(-\frac{a^2+b^2}{\sigma_E^2}\right) \tag{7}$$

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 center of the filter. The filter is applied to each layer with wrap-around of opposing edges, to give a toroidal structure. The lateral inhibition and excitation parameters used in the SOM architecture are given in Table 3.

Contrast Enhancement of Neuronal Firing Rates Within Each Layer

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

$$r = f^{sigmoid}(h') = \frac{1}{1 + \exp\left(-2\beta(h' - \alpha)\right)} \tag{8}$$

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 between each

Table 2
Parameters for Gabor Input Filters

Parameter (Symbol)	Value
Phase shift (Ψ)	0: White on black bar π: Black on white bar −π/2: Black (left) and white (right) bar π/2: Black (right) and white (left) bar
Wavelength(λ)	2 pixels
Orientation(θ)	0, π/4, π/2, 3π/4
Spatial bandwidth (*b*)	1.5 octaves
Aspect ratio (γ)	.5

layer of neurons to control the sparseness of the firing rates. For example, to set the sparseness to 4%, the threshold is set to the value of the 96th percentile point of the activations within the layer. The parameters for the sigmoid activation function are shown in Table 4. These are general robust values found to operate well from an optimization procedure performed in earlier work (Stringer, Perry, Rolls, & Proske, 2006; Stringer & Rolls, 2008; Stringer, Rolls, & Tromans, 2007). In the brain, competition and contrast enhancement would be affected by inhibitory interneurons. While the actual calculation performed within each layer of VisNet during contrast enhancement is not biologically plausible, the method described here will give a similar overall effect as a population of inhibitory neurons and also allows for control of sparseness.

Training the Network: Learning Rules Used to Modify Synaptic Weights

At the start of training, the feedforward synaptic weights between successive layers are initialized with random values. In addition, the visual images of faces used to train the network are initially preprocessed by the Gabor input filters. During training, the outputs of the Gabor filters are passed to layer 1 of VisNet. In layer 1, the cell firing rates are computed as described above. Activity is then propagated sequentially through layers 2 to 4 using the same mechanisms at each layer. As activity is propagated through successive layers, the synaptic weights w_{ij} in the feedforward synaptic connections between successive layers are modified according to the Hebb learning rule (a) for the first simulation experiment with static faces, or the trace learning rule (b) for the second simulation experiment with rotating faces. A Hebb rule is also used for the second simulation as a check for the necessity of using a trace learning rule. The effect of the trace learning rule is to encourage postsynaptic neurons to learn to respond to input patterns that tend to occur close together in time, which should not occur with the Hebb rule. VisNet has previously used trace learning to develop cells that respond to rotating faces with view invariance (Wallis & Rolls, 1997).

Table 3
SOM Parameters

Layer	1	2	3	4
Excitatory radius (σ_E)	1.4	1.1	.8	1.2
Excitatory contrast (δ_E)	5.35	33.15	117.57	120.12
Inhibitory radius (σ_I)	2.76	5.4	8.0	12.0
Inhibitory contrast (δ_I)	1.5	1.5	1.6	1.4

Table 4
Parameters for Sigmoid Activation Function

Layer	1	2	3	4
Percentile	80	80	80	80
Slope (β)	190	40	75	26

Face Stimuli

The stimuli used to train and test the network were realistic images of faces, created using the 3-D face modeling software package Facegen (www.facegen.com). FaceGen builds artificial 3-D face images from templates taken from 273 high resolution 3-D face scans from real faces. The images are averaged, and PCA is used to extract a set of variances from the mean representing facial categories such as shape, color, and gender. This in turn gives a normal distribution from which a random coefficient can be chosen, creating a random, realistic face based on a range of alterable features.

In the first experiment, all images faced straight ahead, as shown in Figure 2. A total of 400 face images were built for each simulation in Experiment 1; 200 male and 200 female. Half of these images were used in training, the other half in testing. In the second experiment, the faces were presented in different orientations. One hundred faces of each gender were created for each simulation, each presented facing 0° (straight ahead), turned 45°, and turned 90° giving a total of 600 images, as illustrated in Figure 3. Again, half of these images were used in training and the other half in testing.

Within each of the two experiments, we performed two separate simulations with different distributions of gender-related facial features. FaceGen's PCA algorithm determines the features in a face most correlated with a particular category. For example, a particular size or shape of the nose might be particularly correlated with female faces. Consequently, regarding gender, FaceGen uses values from −4 to 4, where −4 refers to *very female,* 4 refers to *very male,* and 0 is *ambiguous*. In the first simulation of both experiments, we selected extreme values of −4 and 4. In the second simulations, we selected a more realistic distribution of

Figure 2. Examples of face images used in Experiment 1 in which faces are seen from the front view. Left: Example of a FaceGen generated female face. Right: Example of a male face.

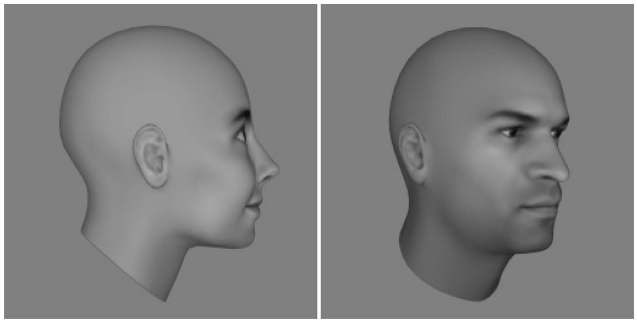

Figure 3. Examples of face images used in Experiment 2 in which faces are seen rotating through different orientations. Left: Example of a FaceGen generated female face, turned 90°. Right: Example of a male face, turned 45°.

gender values, shown in Figure 4. This weighted most of the images toward a particular gender, but also included some more gender-ambiguous stimuli.

Protocols for Training and Testing the Network

In order to train the network during each simulation, a set of face images were presented to it in a randomized order. A total of 200 face stimuli were used for training in each of the two simulations of the first experiment, while 300 face stimuli were used for training in the two simulations of the second experiment. The presentation of all face stimuli comprised one training epoch. At each presentation of a face to the network, the activation of individual neurons within the first layer is calculated, then their firing-rate is calculated and finally their afferent feedforward synaptic weights updated according to whichever learning rule is currently used (Hebb or trace). After the first layer has been trained for a fixed number of epochs, activity is then propagated to the second layer and the training process repeated. In this manner, the network is trained one layer at a time, starting with layer 1 and finishing with layer 4. In both experiments we used 50, 100, 100, 75 training epochs for layers 1, 2, 3, and 4, respectively. Across all layers and in both experiments, the learning rate of the model was set to 0.1, and the sparseness of firing rates was set to 0.8. Parameters were tuned to give best performance. However, in additional simulations (not shown), it was found that model performance was quite robust as these parameter values were varied. The numbers of training epochs per layer were determined by reducing the number of epochs until model performance became degraded. The numbers of training epochs are therefore the minimum required to allow for adequate learning.

After training, the network was tested on a new set of face stimuli. The firing rates from all neurons within the output (fourth) layer of the network were recorded in response to each face stimulus.

Information Theoretic Performance Measures

The network's performance was assessed using two information theoretic measures: single- and multiple-cell information about which stimulus category, that is, male or female face, was shown. Full details on the application of these measures to VisNet are given by Rolls and Milward (2000). These measures reflect the extent to which cells respond invariantly to a particular stimulus category such as a male or female face over a number of transforms such as different facial orientations or identities, but respond in contrasting ways to the different stimulus categories.

The single-cell information measure is applied to individual cells in layer 4 and measures how much information was available from the response of a single cell about which stimulus category, that is, male or female face, was shown. For each cell, the single-cell information measure used was the maximum amount of information a cell conveyed about any one stimulus category. This was computed using the following formula with details given by Rolls and Milward (2000) and Rolls, Treves, Tovee, and Panzeri (1997). The stimulus-specific information $I(s, R)$ is the amount of

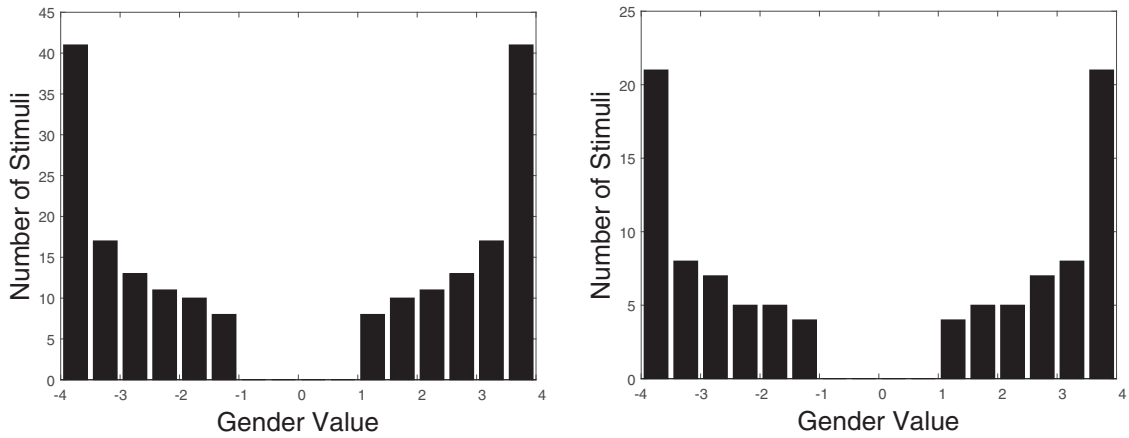

Figure 4. Distributions of gender values used to construct the face images in simulations that included more gender-ambiguous faces. Left: Distribution of gender values chosen for Experiment 1, Simulation 2. Right: Distribution of gender values chosen for Experiment 2, Simulation 2. These distributions are more realistic in that they include a significant proportion of more gender-ambiguous faces between the extreme gender values of −4 (*very female*) and 4 (*very male*).

information the set of responses R has about a specific stimulus s, and is given by

$$I(s, R) = \sum_{r \in R} P(r|s) log_2 \frac{P(r|s)}{P(r)} \quad (9)$$

where r is an individual response from the set of responses R.

However, the single-cell information measure cannot give a complete assessment of VisNet's performance with respect to categorizing male and female faces. If all output cells learned to respond to only one of the two stimulus categories (e.g., male faces), then the single-cell information measures alone would not reveal this. 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 (i.e., both male and female faces) from a population of neurons. Procedures for calculating the multiple-cell information measure are described by Rolls and Milward (2000) and Rolls et al. (1997). In brief, we calculate the mutual information, that is, the average amount of information that is obtained about which stimulus category was shown from a single presentation of a stimulus from the responses of all the cells. That is, the mutual information between the whole set of stimulus categories S and of responses R is the average stimulus-specific information across the stimulus categories. This is achieved through a decoding procedure, in which the stimulus category s' that gave rise to the particular firing rate response vector on each trial is estimated. A probability table is then constructed of the real stimulus category s and the decoded stimulus category s'. From this probability table, the mutual information is calculated as

$$I(s, S') = \sum_{s,s'} P(s, s') log_2 \frac{P(s, s')}{P(s)P(s')} \quad (10)$$

Multiple cell information values were calculated for the subset of cells which had as single cells the most information about which stimulus category (male/female) was shown. In particular, we calculated the multiple-cell information using five cells that had the maximum single-cell information for male faces and five cells with the maximum single-cell information for female faces. The criterion for perfect performance was that the multiple-cell information should reach the level needed to fully discriminate the two stimulus categories, that is $\log_2(S)$ bits. For gender, there are two stimulus categories, meaning that the maximal information is 1 bit.

We also used the following three global measures of network performance.

First, we compared the single-cell information measures of the n highest performing fourth layer neurons in the trained network with the maximum possible information by computing the measure:

$$R_a = \frac{\sum_{i=1}^{n} I_i^{trained}}{nI_{max}} \quad (11)$$

where $I_i^{trained}$ is the single-cell information of neuron i after training, n is the number of highest performing fourth layer neurons used to compute the measure, and I_{max} is the maximum information possible given by $\log_2(S)$ bits.

Second, we compared the performance of the trained network against the untrained network in order to check whether there is an improvement in performance due to training. To do this, we calculated the following:

$$R_r = \frac{\sum_{i=1}^{n} I_i^{trained}}{\sum_{i=1}^{n} I_i^{untrained}} \quad (12)$$

where $I_i^{untrained}$ is the single-cell information of neuron i before training.

We also compared the single-cell information measures of the n highest performing fourth layer neurons in the untrained network with the maximum possible information, with the measure:

$$R_u = \frac{\sum_{i=1}^{n} I_i^{tunrained}}{nI_{max}} \quad (13)$$

where $I_i^{untrained}$ is the single-cell information of neuron i before training.

The feedforward synaptic connections between successive layers are traced back to the retina in order to determine the specific features of a face that drive gender discrimination among the trained neurons of the output layer. Starting from the top 5% of gender specific cells in the top (fourth) layer, that is, the cells with the highest levels of gender-related information, we select the connections from the previous layer that have the highest weights, repeating this process through successive layers until the connections reach the Gabor filters in the retina. This then allows us to plot the pattern of Gabor input filters, to which a gender discriminating neuron in the output has become tuned. By combining the patterns of Gabor filters across output neurons that respond to the same gender, we can see the specific geometric features of a face used by the trained output cells for discrimination of gender.

Results

Experiment 1: Frontal View Faces

In the first experiment, we trained the network on 100 male and 100 female frontal-view faces using the Hebb rule (1). There was no rotation of these faces. The following two simulations were run for this experiment. In the first simulation, all of the training and test faces were generated using FaceGen with the gender set to the extreme values of −4 (*very female*) and 4 (*very male*). Thus, the faces were chosen for maximal dissociation between genders. We used a different set of images for training and testing in order to test the generalization of the neuronal responses to new untrained faces. The second simulation used the same test set, but the network was trained on a more natural range of faces with a more realistic distribution of gender values as shown in Figure 4. This training set included some more gender-ambiguous stimuli. For both simulations, it was found that after training the network had developed localized clusters of neurons in the fourth layer that learned to respond only when a male face was presented, while other clusters of neurons responded only when a female face was presented. The development of neuron clusters responding selectively to one particular gender in the first simulation of Experiment 1 is shown in Figure 5.

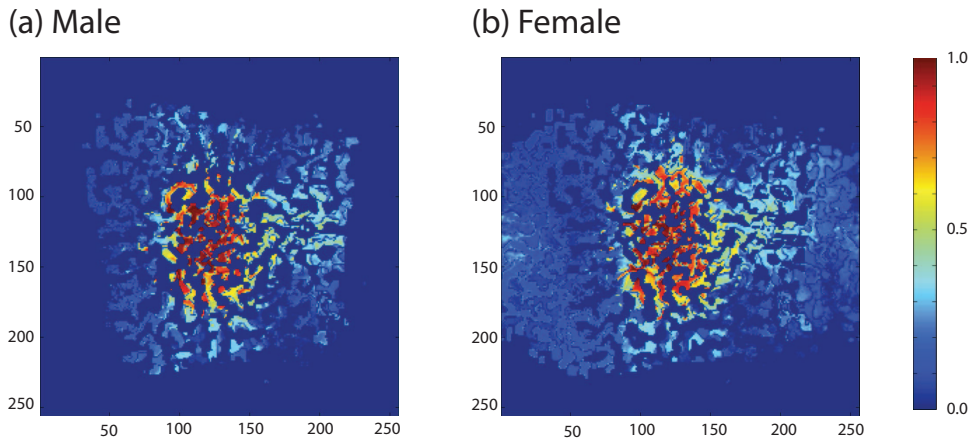

Figure 5. Responses of trained output layer neurons from the first simulation of Experiment 1, in which the network is trained and tested with frontal views of faces. (a) Firing rate responses of posttraining neurons across the fourth (output) layer of cells in response to presentation of a male face; and (b) Firing rate responses of posttraining neurons across the fourth (output) layer of cells in response to presentation of a female face. The responses of the neurons are represented on a color scale, with red indicating a high firing rate and blue a low firing rate. Because of the SOM architecture, cells with similar firing responses cluster together. This causes spatial clustering of cells that respond to either (a) male or (b) female faces, as indicated by the red area in the center of both images. See the online article for the color version of this figure.

To investigate further whether the network had developed neurons that responded selectively to either male or female faces, a single-cell information analysis was performed. Figure 6 shows the amount of single-cell information carried by individual output (fourth) layer cells in rank order, for both simulations. As there are two stimulus categories (male and female), the maximal information possible is one bit. For both simulations, before training, no neurons conveyed more than 0.3 bits of information about the gender of test faces. However, it can be seen that training the network led to a dramatic increase in the amount of information carried by the output layer cells. Moreover, training led to some output neurons developing near maximal levels of information. For the first simulation, 3 neurons were found to provide 0.97 bits of information after training. These 3 neurons responded selectively to one particular gender with almost perfect accuracy. In the second simulation, 13 neurons contained over 0.75 bits of information. The single-cell analysis confirms that after training on 100 male and 100 female faces, VisNet developed a number of output neurons that responded selectively to one or other gender, meaning that the firing of these neurons allows for almost perfect discrimination between genders, although this performance worsens slightly when more ambiguous stimuli are presented during training.

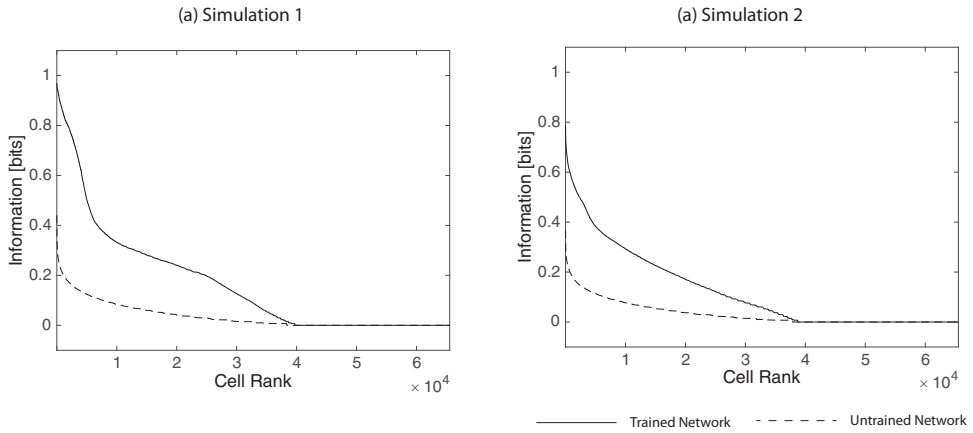

Figure 6. Single-cell information for the simulations of Experiment 1, in which the network is trained and tested with frontal views of faces. All fourth layer neurons are plotted along the abscissa in rank order according to the amount of information they convey about the male and female face categories. One bit of information reflects a cell which responds exclusively to stimuli of a particular gender, for example only male faces. The trained network is represented by the solid line, while the untrained network is represented by the dotted line.

Table 5
Information Measures for Both Simulations of Experiment 1

Information measure	Simulation 1	Simulation 2
I_{max}	.970	.786
R_r	2.472	2.333
R_a	.962	.738
R_u	.389	.316
Number of cells	18	44

Note. The following three information measures are given: I_{Max} is the maximum single-cell information reached by any of the output layer neurons after training, R_r quantifies the amount of single-cell information carried by output layer neurons in the trained network relative to the untrained network, R_a quantifies the accuracy of gender discrimination after training with a value of 1 reflecting perfect discrimination, and R_u is the performance of the untrained network again with a value of 1 reflecting perfect discrimination. The final row gives the number N of output layer cells used to compute each of the measures R_r and R_a.

Table 5 shows the maximal information reached by any of the output layer neurons after training I_{max}, the performance improvement due to training R_r, the absolute performance after training R_a, and the performance of the untrained network R_u for both simulations of Experiment 1. R_r, R_a, and R_u were each computed with different numbers n of output layer cells for the two simulations as shown in the table. For each simulation, the number n of output neurons used to compute the measures was chosen to ensure that approximately half of the cells responded to each of the two genders with a maximum deviation of no more than 5%.

For these simulations, the values of I_{max} and R_a both lie in the interval [0,1]. High performance after training is indicated by values of I_{max} and R_a close to one, with perfect gender discrimination achieved when these measures reach one. For R_r, an improvement in performance due to training is signified by a value greater than one. The higher this value, the larger the improvement after training. The values of these measures confirm that these simulations achieved excellent performance. For the first simulation, in which the network was trained on faces with extreme gender values of −4 (*very female*) and 4 (*very male*), the measures I_{max} and R_a are greater than 0.96, and the improvement over the untrained network R_r is close to 2.5. The measure R_a denotes an almost perfect response, and R_r indicates that the trained network performed over twice as well as the untrained network. For Simulation 2, in which the network was trained on a more realistic distribution of faces including more gender-ambiguous stimuli, performance worsens but remains good. I_{max} is 0.79 and R_a is 0.74, showing excellent gender discrimination. The trained network is twice improved over the untrained network, as shown by an R_r of 2.33. For both simulations, R_u is very low compared with R_a, showing a large increase in performance of the trained network over the untrained network.

Figure 7 shows multiple-cell information measures for both simulations of Experiment 1. While individual gender-specific cells are tuned to one specific gender, multiple-cell information measures demonstrate whether the network as a whole has learned to respond to both genders; that is, whether after training, both genders are adequately represented by different subpopulations of output neurons. In these simulations, after training a very small number of cells were required to reach the maximum multiple-cell information of one bit, confirming that the network develops some cells responsive to female stimuli and other cells responsive to male stimuli. The untrained cells can also carry information, as shown by the gradually increasing information as the number of cells included in the analysis increases. However, the trained networks differentiate between gender far more effectively.

VisNet has previously been shown to invariant to image, by using a trace learning rule to associate the different sizes of a single face as being the same face (Wallis & Rolls, 1997). Nevertheless, a number of checks were performed to ensure that the network was not relying on artefactual cues, such as image contrast

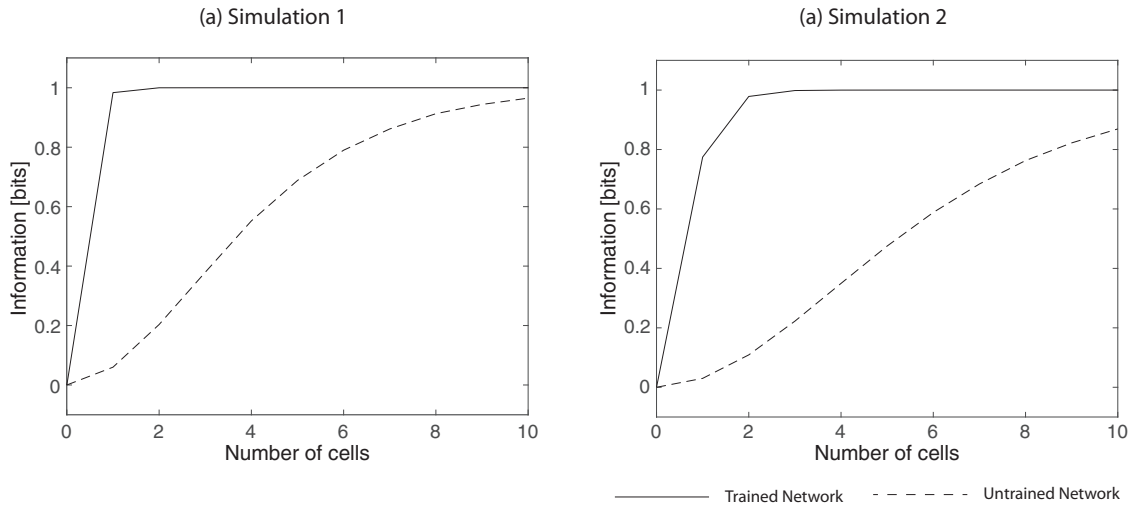

Figure 7. Multiple-cell information measures for Experiment 1. The plots show the amount of multiple-cell information when a given number of fourth layer cells are included in the analysis, and thus the number of cells required to reach maximal information of 1 bit. The trained network is represented by the solid line, while the untrained network is represented by the dotted line. We see a marked improvement of the network after training.

or size, to inform gender. To start with, Simulation 1 was repeated using contrast-normalized stimuli. The maximum single-cell information carried by an output neuron after training remained high at 0.89 showing only a minimal drop in performance (results not presented). To check that the network was not relying on overall size differences between male and female faces, Simulation 1 was repeated using a total of 600 stimuli, comprising of 100 male and 100 female faces each in three different sizes. A trace rule was used in place of a Hebb rule, so the network could associate the different sizes together. After training, maximum single-cell information was 0.93 (results not presented). Both simulations showed a negligible drop in performance, indicating that the network did not use overall size differences or contrast differences to differentiate between male and female faces.

To further confirm that VisNet was indeed using the spatial form and arrangement of facial features to determine gender, we examined the specific facial features which informed gender. To do this, the synaptic connections from the top 10 performing male selective cells and top 10 female selective neurons in the output layer, in terms of single-cell information, were traced back to the input Gabor filters. This indicates the facial features to which the network has learned to respond. More precisely, for each output cell selected for this analysis, the incoming synaptic connections with high synaptic weights (top 5%) were recursively traced back through the layers to the input Gabor layer. The program then plots a corresponding oriented bar with a specific darkness that reflects a cumulative product of the synaptic weights along the layers, at the specific retinal location of that Gabor filter. In VisNet, inhibitory synaptic connections have been implicitly implemented, so this analysis did not take their effect into account. Nevertheless, this procedure provided an estimation of the collection of input features to which the output neurons learned to be sensitive.

This analysis of the synaptic connectivity through the layers reveals the facial features that are driving the highest performing cells in the output layer, showing exactly what the network has learned from the male and female face stimuli. Figure 8a and 8b show the averaged features detected for male faces and female faces, respectively. In order to analyze how these features differ between the two genders, we then subtracted one image from the other as shown in Figure 8c. This plot indicates that gender information is primarily held in the eyes, nose, and mouth. In particular, female eyes are usually closer to the nose than male eyes, and female noses are often narrower. The contours of the mouth are more distinctive in males. Similar results have been found in human eye-tracking studies, in which the eyes and mouth are most important for gender categorization (Depuis-Roy, Fortin, Fiset, & Gosselin, 2009; Gosselin & Schyns, 2001). Results are only presented for Simulation 1 for the sake of brevity; however, Simulation 2 also produced similar profiles.

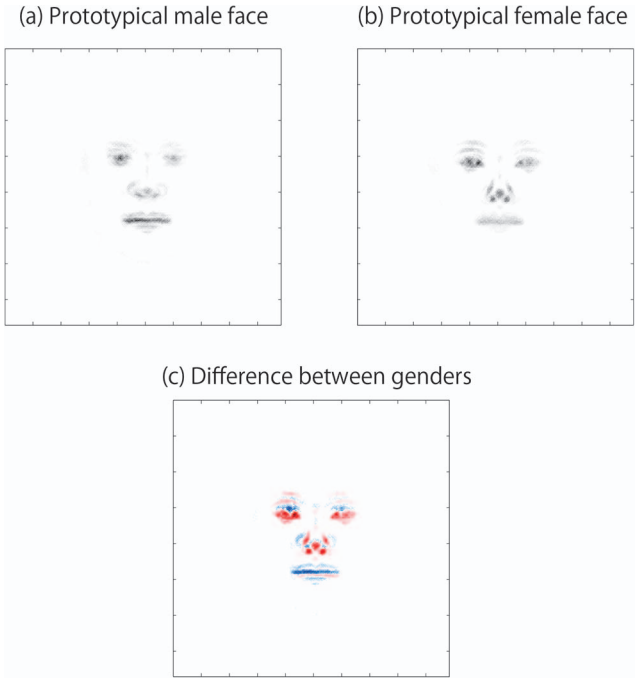

Figure 8. Analysis of the facial features represented by input Gabor filters that drive gender selective output neurons after training in the first simulation of Experiment 1. Starting from a gender (male or female) discriminating neuron in the output (fourth) layer, we select the connections from the previous layer that have the highest weights, repeating this process until the connections reach the Gabor filters in the retina. We repeated this procedure for the top 10 male selective neurons and top 10 female selective neurons. This analysis enabled us to determine the features in a face that permit gender to be discriminated. Plot (a) shows these facial features averaged over the output neurons that respond selectively to male faces. Plot (b) shows similar results averaged over the output neurons that respond to female faces. Plot (c) is the difference between plots (a) and (b), thus showing the facial features which actually distinguish gender. The red and blue colors in plot (c) represent the female and male features respectively. Areas of no color represent face regions containing no gender-specific features. See the online article for the color version of this figure.

Experiment 2: Rotated Faces

In the second experiment, we trained the network on 50 male faces and 50 female faces, where each of these faces was shown in three different views at 0° (straight ahead), turned 45°, and turned 90°. This gave a total of 300 face images. The simulations were carried out using the trace learning rule (2) to drive the development of male and female selective output neurons with view invariant responses. In order to make the training ecologically valid, the three different views of each face were shown sequentially. That is, for each face identity, first the 0° view was shown, then 45° then 90°. In order to avoid the 90° view of a particular face being associated with the following presentation of a new face at 0°, we reset the trace values $\bar{r}_i$ to zero after the three views of each face. The order of presentation of each face identity was randomized. The same two simulation conditions were run for this experiment as in Experiment 1.

Figure 9 shows the single-cell information carried by all output layer cells for both simulations of Experiment 2. As in Experiment 1, the results for Simulation 1 show that the network has learned to develop neurons in the output layer that successfully discriminate between genders. One hundred sixty-three neurons in the output layer carried a maximum information of 0.99 bits. The right plot of Figure 9 shows that in the second simulation training led to a substantial increase in the information carried by the output layer neurons. In this case, 181 neurons have at least 0.6 bits of information.

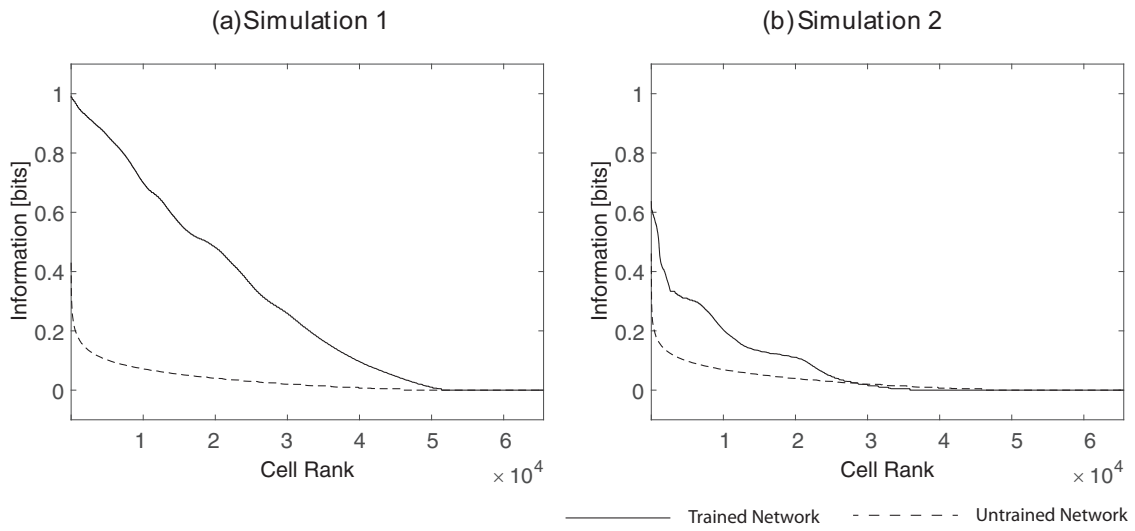

Figure 9. Single-cell information results for both simulations of Experiment 2, in which the network is trained and tested with three different views (0° [straight ahead], turned 45°, and turned 90° of rotating faces). Conventions as for Figure 6. One bit of information reflects a cell that responds exclusively to stimuli of a particular gender, such as male faces, regardless of facial orientation. For both simulations, it can be seen that training has led to a large increase in the information carried by the output cells about facial gender. Although performance is somewhat reduced for the second simulation, in which the network was trained on a more realistic distribution of faces that included more gender-ambiguous stimuli.

Table 6 shows the maximal information reached by any output neuron I_{max}, the performance improvement due to training R_r, the absolute performance of the trained network R_a, and performance of the untrained network R_u for both simulations of Experiment 2. Values close to 1 for I_{max} and R_a in the first simulation show that the trace learning rule has successfully caused the network to develop neurons, which discriminate between genders across different views. A low R_u value shows that the untrained network did not successfully discriminate gender. The R_r value indicates improvement of the trained network over the untrained network. In Simulation 2, all indicators show a large improvement in the performance of the trained network over the untrained network, albeit with a lower overall performance compared with the first simulation.

Table 6
The Four Information Measures I_{Max}, R_r, R_a, and R_u Computed for Both Simulations of Experiment 2

Information measure	Simulation 1	Simulation 2
I_{max}	.993	.637
R_r	2.630	1.853
R_a	.992	.618
R_u	.377	.333
Number of cells	20	1,000

Note. The following three information measures are given: I_{Max} is the maximum single-cell information reached by any of the output layer neurons after training, R_r quantifies the amount of single-cell information carried by output layer neurons in the trained network relative to the untrained network, R_a quantifies the accuracy of gender discrimination after training with a value of 1 reflecting perfect discrimination, and R_u is the performance of the untrained network again with a value of 1 reflecting perfect discrimination. The final row gives the number N of output layer cells used to compute each of the measures R_r and R_a.

Figure 10 shows multiple-cell information measures for both simulations of Experiment 2. In the first simulation, only one or two cells were required to reach the maximum information of one bit, confirming that the trace learning rule enabled the network to develop neurons that represent both male and female faces. The second simulation required a larger number of cells to reach the maximum information as it was a more difficult problem for the network to solve. However, as shown in the right plot of Figure 10, the trained network still outperforms the untrained network.

Figure 11 shows the features of a face that are being used to determine gender by tracing back the strengthened connections from the top performing male and female selective cells in the output layer to the input Gabor filters. This procedure was carried out for the first simulation of Experiment 2. It can be seen that tracing back the connections from top-performing cells in the output layer results in similar findings to Experiment 1; gender information is held in the eyes, nose, and mouth. However, the rotated faces also contain gender information in the chin and supraorbital arch. Images are less clear than in Experiment 1 due to using three rotated views of each face. Results are only presented for Simulation 1 of Experiment 2 for the sake of brevity; however, Simulation 2 also produced similar profiles.

We also reran the first simulation of Experiment 2 using the Hebb rule (1), in order to check the necessity of using a trace learning rule for learning view invariant representations of male and female faces. As expected, it resulted in a failure to associate the three different views to the same face gender, meaning that

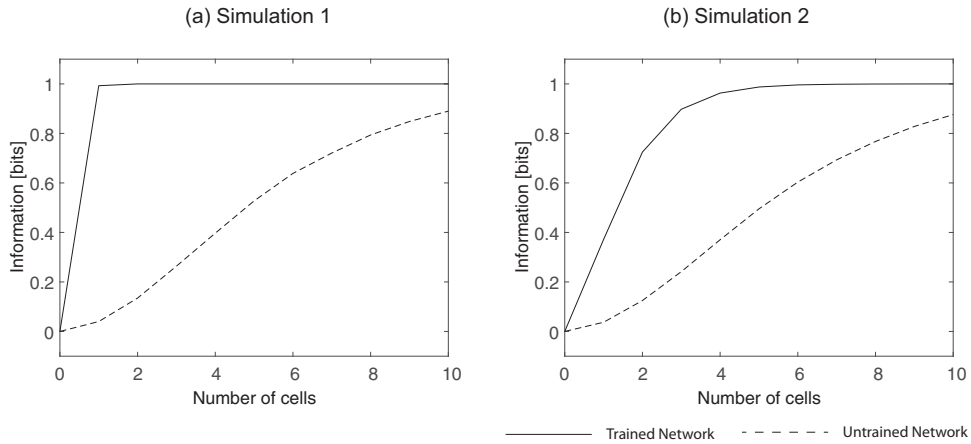

Figure 10. Multiple-cell information measures for both simulations of Experiment 2, in which the network is trained and tested with three different views (0° [straight ahead], turned 45°, and turned 90° of rotating faces). Conventions as for Figure 7. There is a large increase in the multiple-cell information carried by the fourth layer cells after training. After training, the multiple-cell information rapidly asymptotes to the maximum of 1 bit for both simulations with only a few (1 to 4) cells included in the analysis.

the network failed to develop neurons that could correctly categorize gender across different views. It did however develop gender-selective neurons within each view, leading to a total of six representations of gender. Figures for this simulation are not shown.

Discussion

There is evidence that facial genders are represented by separate subpopulations of neurons in the FFA (Contreras et al., 2013; Kaul et al., 2011; Ng et al., 2006; Podrebarac et al., 2013). In this paper, we have presented a biologically plausible, unsupervised learning model of how gender-specific subpopulations might develop. By exploiting the natural statistics of face images, the proposed network learns the distinct patterns of facial features that distinguish between genders in an unsupervised manner, that is, without ever needing to be explicitly informed about the gender of faces during training. Freeman and colleagues found activity in the fusiform gyrus and FFA in response to objective differences between genders of FaceGen faces. These objective differences can be considered to be the facial patterns that VisNet learns, which in turn suggests that the fourth layer of VisNet may capture neural processing within the fusiform area. By implementing a SOM, neurons responding to a particular gender clustered together; a computational analogue to the subpopulations in the brain. Many high-performing gender recognition models already exist, but to the authors' knowledge, this is the first example of a biologically plausible, unsupervised learning model. The success of the network in discriminating between genders is shown by examining the firing rate responses of cells and the information carried by these cells.

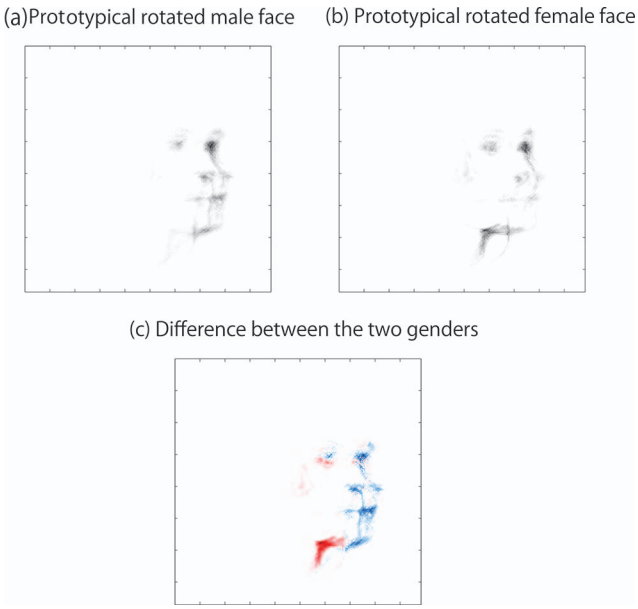

Figure 11. Analysis of the facial features represented by input Gabor filters that drive gender selective output neurons after training in the first simulation of Experiment 2. Conventions as for Figure 8. See the online article for the color version of this figure.

Simulations were run using VisNet, an established model of the primate ventral visual pathway (e.g., Tromans et al., 2011; Wallis & Rolls, 1997). In Experiment 1, the network developed neurons in the output layer that contained almost perfect levels of gender information, thereby allowing for accurate discrimination of gender. Experiment 2 explored the ability of the network to identify gender under more realistic circumstances; with the faces seen from different viewpoints. Again, the network developed neurons in the output layer that responded selectively to a particular gender, regardless of whether the face was presented at 0°, 45°, or 90°.

In the second simulation of each experiment, a more realistic distribution of faces along the gender continuum was used. Accuracy was reduced in these simulations, reflecting the more

difficult problem of learning to determine gender from more gender-ambiguous stimuli. In both simulations however, the trained network still outperformed the untrained network. Additionally, the numbers of training epochs were optimized for Simulation 1. It is possible that the second simulation, being a more difficult problem for the network to solve, would have benefitted from more training epochs. Similarly, the second experiment used three rotations because this was found to be the minimum number of rotations between frontal and profile views required for the network to develop view invariant gender-selective neurons. A larger number of rotations may have helped the network to more effectively associate together the different facial views by trace learning, potentially improving the accuracy of the network overall. It may also be possible to further improve performance by combining the trace learning with another invariance learning mechanism known as continuous transformation learning (Spoerer, Eguchi, & Stringer, 2016) that utilizes spatially continuous stimulus transformations.

We also aimed to investigate that features of a face determine gender. Tracing back the synaptic connections from the gender selective output layer cells to the input Gabor filters showed the patterns of facial features that those output neurons had become tuned to. This allowed us to determine the facial features that distinguish between genders. The eyes, eyebrows, nose, mouth, and chin were found to differentiate between male and female faces. Behavioral studies have previously shown the eyes and mouth to be the most important facial features for gender discrimination, in accordance with the findings of this study (Depuis-Roy et al., 2009; Gosselin & Schyns, 2001). Furthermore, as the output neurons were not shown to be cueing off the outline of the face, it implies that gender was not identified by overall size of the face, but rather by individual features or spatial relationships between these features.

For simplicity, this study used faces created by the FaceGen software package. While the images it produces are quite lifelike, it results in some loss of detail that could be important in face processing, such as color, hair, and wrinkles. However, in order to fully represent the richer information present in real faces, the retina of VisNet would need to be increased in size. To run such a simulation would be very computationally expensive, and so was not attempted in the present study.

In this paper, we have developed a biologically plausible, unsupervised learning neural network model that develops neurons that have learned to respond selectively to either male or female faces when presented with lots of face images during training. The naturally occurring differences in male and female faces are exploited, in order to build separate representations of gender to which all newly presented faces are compared. We find that this can still occur even when the faces are rotated through different views, which is an important further test for ecological validity. These simulations predict the development of similar gender discriminating neurons in the primate brain, which would provide a neural basis for the perception of the gender of faces. Furthermore, the network model develops localized clusters of neurons in the output layer that fire to either male or female faces, suggesting the presence of similar localized subpopulations of gender-specific neurons in the brain (Contreras et al., 2013; Kaul et al., 2011; Ng et al., 2006; Podrebarac et al., 2013). The simulation results reported here provide important predictions for future experimental studies investigating the neural basis of face perception in the brain.

References

Baudouin, J.-Y., & Brochard, R. (2011). Gender-based prototype formation in face recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 888–898.

Bestelmeyer, P., Jones, B., DeBruine, L., Little, A., Perrett, D., Schneider, A., . . . Conway, C. (2008). Sex-contingent face aftereffects depend on perceptual category rather than structural encoding. *Cognition, 107,* 353–365.

Bestelmeyer, P., Jones, B., DeBruine, L., Little, A., & Welling, L. (2010). Face aftereffects suggest interdependent processing of expression and sex and of expression and race. *Visual Cognition, 18,* 255–274.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel patterns in Fusiform Face Area differentiate face by sex and race. *PLOS One, 8,* 1–6.

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,* 5602–5618.

Depuis-Roy, N., Fortin, I., Fiset, D., & Gosselin, F. (2009). Uncovering gender discrimination cues in a realistic setting. *Journal of Vision, 9,* 1–8.

Eguchi, A., Humphreys, G. W., & Stringer, S. M. (2016). The visually-guided development of facial representations in the primate ventral visual pathway: A computer modeling study. *Psychological Review, 123,* 696–739.

Fang, F., & He, S. (2005). Viewer-centred object representation in the human visual system revealed by viewpoint aftereffects. *Neuron, 45,* 793–800.

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

Freeman, J., Rule, N. O., Adams, R. B., Jr., & Ambady, N. (2010). The neural basis of categorical face perception: Graded representations of face gender in fusiform and orbitofrontal cortices. *Cerebral Cortex, 20,* 1314–1322.

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

Frisby, J. (1979). *Seeing*. New York, NY: Oxford University Press.

Gosselin, F., & Schyns, P. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41,* 2261–2271.

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research, 32,* 203–218.

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

Kaul, C., Rees, G., & Ishai, A. (2011). The gender of face stimuli is represented in multiple regions in the human brain. *Frontiers in Human Neuroscience, 4,* 1–12.

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

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science, 32,* 1502–1507.

Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of The Royal Society B, 272,* 2283–2287.

Moghaddam, B., & Yang, M.-H. (2000). Gender classification with support vector machines. In *4th IEEE International Conference on Automatic Face and Gesture Recognition,* 306–311.

Ng, M., Ciaramitaro, V. M., Anstis, S., Boynton, G. M., & Fine, I. (2006). Selectivity for the configural cues that identify the gender, ethnicity, and identity of faces. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 19552–19557.

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

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences, 335,* 23–30.

Perry, G., Rolls, E. T., & Stringer, S. M. (2006). Spatial vs temporal continuity in view invariant visual object recognition learning. *Vision Research, 46,* 3994–4006.

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,* 83–96.

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 of the United States of America, 89,* 8366–8370.

Podrebarac, S. K., Goodale, M. A., van der Zwan, R., & Snow, J. C. (2013). Gender-selective neural populations: Evidence from event-related fMRI repetition suppression. *Experimental Brain Research, 226,* 241–252.

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W., & Nakayama, K. (2003). Fitting the mind to the World: Face adaptation and attractiveness aftereffects. *Psychological Science, 14,* 558–566.

Rhodes, G., Jeffery, L., Watson, T. L., Winkler, C., & Clifford, C. W. (2004). Orientation-contingent face aftereffects and implications for face-coding mechanisms. *Current Biology, 14,* 2119–2123.

Rolls, E. T., Cowey, A., & Bruce, V. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas [and discussion]. *Philosophical Transactions: Biological Sciences, 335,* 11–21.

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,* 2547–2572.

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

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

Sergent, J., Ohta, S., & Macdonald, B. (1992). Functional neuroanatomy of face and object processing. *Brain, 115,* 15–36.

Spoerer, C. J., Eguchi, A., & Stringer, S. M. (2016). A computational exploration of complementary learning mechanisms in the primate ventral visual pathway. *Vision Research, 119,* 16–28.

Stringer, S. M., Perry, G., Rolls, E. T., & Proske, J. H. (2006). Learning invariant object recognition in the visual system with continuous transformations. *Biological Cybernetics, 94,* 128–142.

Stringer, S. M., & Rolls, E. T. (2008). Learning transform invariant object recognition in the visual system with multiple stimuli present during training. *Neural Networks, 21,* 888–903.

Stringer, S. M., Rolls, E. T., & Tromans, J. M. (2007). Invariant object recognition with trace learning and multiple stimuli present during training. *Network, 18,* 161–187.

Sun, Z., Bebis, G., Yuan, X., & Louis, S. J. (2002). Genetic feature subset selection for gender classification: A comparison study. In *IEEE Workshop on Applications of Computer Vision*.

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,* e25616.

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

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

Zhao, W., Chellappa, R., Phillips, P., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys, 35,* 399–458.

Received November 2, 2015
Revision received August 12, 2016
Accepted October 6, 2016 ■