

Learning Innate Face Preferences

James A. Bednar and Risto Miikkulainen

Report AI-01-291 November 2001

`jbednar,risto@cs.utexas.edu`
`http://www.cs.utexas.edu/users/n/`

Artificial Intelligence Laboratory
The University of Texas at Austin
Austin, TX 78712

Abstract

Whether humans have a specific, innate perceptual ability to process faces remains controversial. Studies have found face-selective brain regions in adults and have shown that even newborns preferentially attend to face-like stimuli. On this basis researchers have proposed that there are genetically hard-wired brain regions that specifically process faces. However, other studies suggest that the face-processing hardware is general purpose and highly plastic, even at birth. We propose a solution to this apparent paradox: innate face preferences may be learned by a general-purpose self-organizing system from internally generated input patterns, such as those found in PGO waves during REM sleep. Simulating this process with the HLISSOM model, we demonstrate that such an architecture constitutes an efficient way to specify, develop, and maintain functionally appropriate perceptual organization. This preorganization can account for newborn face preferences, providing a computational explanation for how genetic influences interact with experience to construct a complex system.

1 Introduction

Do humans have an innate perceptual ability to process faces? Researchers have investigated this question repeatedly over the past several decades, but it remains unresolved and highly controversial (as reviewed in Gauthier and Nelson, 2001; Pascalis and Slater, 2001; Tovée, 1998). On the one hand, numerous studies indicate that newborns will preferentially turn their eyes or head towards face-like stimuli within minutes or hours after birth (Goren et al., 1975; Johnson et al., 1991; Johnson and Morton, 1991; Mondloch et al., 1999; Simion et al., 1998b; Valenza et al., 1996). Many researchers argue that there therefore must be a genetically hard-wired brain mechanism specific to face processing (Johnson and Morton, 1991; Mondloch et al., 1999; Simion et al., 1998a). Indeed, specific brain regions that respond preferentially to faces have been found in adult humans and in infant monkeys (Kanwisher et al., 1997; Rodman, 1994).

On the other hand, numerous studies also demonstrate rapid experience-driven development of face processing abilities, even in the first few hours and days after birth (Bushnell et al., 1989; Field et al., 1984; Walton et al., 1997; Walton and Bower, 1993). Such plasticity appears to be the rule, rather than the exception, in both the developing and the adult nervous system (as reviewed in Buonomano and Merzenich, 1998; Hirsch, 1985). Moreover, the adult face-selective regions have been found to process many stimuli other than faces (Gauthier et al., 1999; Haxby et al., 2001). Computational simulations have shown how such face-selective regions can arise without being innately specified for face perception (Dailey and Cottrell, 1999). Finally, some studies have found newborn preferences only for general image features (such as complexity or curvature), and not for face-like stimuli in particular (Easterbrook et al., 1999; Kleiner, 1993; Simion et al., 2001; Slater, 1993). From this evidence many researchers argue that the ability to process faces is learned from experience, and not genetically hard-wired.

In contrast with the above two competing hypotheses, in this paper we show that learning and hard-wiring are not only fully compatible even within a single cortical area, they can work synergetically to develop sophisticated face processing abilities. The underlying thesis is that prenatal development uses the same learning mechanisms that in the adult extract regularities from the visual environment, but it uses them to extract regularities in training inputs generated internally under genetic control (Constantine-Paton et al., 1990; Jouvett, 1999; Marks et al., 1995; Roffwarg et al., 1966; Shatz, 1990, 1996). Instead of precisely specifying the organization of the brain, the genome may simply encode a developmental process that is based on genetically determined patterns presented to a general learning mechanism. After birth, the learning mechanism can then seamlessly integrate environmental and genetic information into the same cortical hardware. This explanation can resolve many of the paradoxes underlying the controversy about face perception.

We test this idea computationally using a self-organizing model of cortical development (see Erwin et al., 1995; Swindale, 1996 for reviews of this class of models). Computational models allow the assumptions and implications of a conceptual model to be evaluated rigorously, and are particularly useful for generating testable predictions. In previous experiments using the RF-LISSOM model (Receptive-Field Laterally Interconnected Synergetically Self-Organizing Map; Miikkulainen, Bednar, Choe, and Sirosh, 1997; Sirosh and Miikkulainen, 1994; Sirosh, Miikkulainen, and Bednar, 1996) we have shown that orientation maps and specific lateral connection patterns in the primary visual cortex can form from spontaneously generated activity (Bednar and Miikkulainen, 1998). We have also shown in preliminary work how a simplified model of face pro-

cessing may account for much of the data on newborn face preferences (Bednar and Miikkulainen, 2000).

In this paper we present HLISSOM, a significant extension to RF-LISSOM that allows processing of real photographic images by modeling the retina, LGN, and higher-level visual areas in addition to the primary visual cortex. This model allows preferences for low-level features such as orientation or spatial frequency and high-level features such as faces to be compared in the same model, in order to account for the full range of data measured in experiments with newborns. Very few self-organizing models have been tested with real images, and none to our knowledge have previously modeled the self-organization of both the primary visual cortex and higher regions, nor have any simulated the large retinal and cortical area needed to process the large stimuli tested with newborns. With HLISSOM, we demonstrate how patterns found during REM sleep can explain a newborn's apparently innate predisposition for faces, within a general-purpose adaptive system for learning both low-level and high-level features of images.

In the next section, we will briefly review other relevant models of the development of face perception and experimental evidence for the internally generated brain activity upon which the model is based. Next, we introduce the HLISSOM architecture and describe the experimental setup for testing face detection. We then show how this model can account for newborn face preferences using both schematic and photographic images, and suggest future computational and psychophysical experiments to help understand the development of face processing.

2 Related work

There are three main previous explanations for how a newborn can show an initial face preference and then later develop full face processing abilities. (1) The *Linear Systems Model* (LSM) acts as a minimal baseline against which other theories can be compared. The LSM posits that newborn preferences (including those for faces) result solely from the spatial frequencies in an image, filtered by the measured sensitivity of the newborn to each frequency. These sensitivities result from general, genetically determined aspects of the visual system; adult face processing abilities are assumed to develop only later, in higher visual areas. (2) *Sensory models* are generalizations of the LSM, stating that the spatial frequency response plus some other general features of the sensory apparatus together account for newborn face preferences, still without face-specific circuitry present at birth. (3) *Multiple systems models* propose that there is a genetically fixed, face-specific system present at birth that is later replaced or augmented by a separate, plastic system that operates into adulthood. These three hypotheses will be reviewed below, and a simpler, more effective alternative will be proposed: that only a single, general-purpose system is necessary if that system is exposed to *internally generated patterns*, such as those in REM sleep.

2.1 Linear Systems Model

The Linear Systems Model (Banks and Salapatek, 1981; Kleiner, 1993) is a straightforward and effective way of explaining a wide variety of newborn pattern preferences, and is easily implemented as a computational model. It is based solely on the newborn's measured contrast sensitiv-

ity function, which is limited by the immature state of the eye and the early visual pathways. The assumption is that newborns pay attention to those patterns that activate their early visual system most strongly. For instance, low-contrast patterns and patterns with only very fine detail are not very salient to newborns, and thus newborns do not attend to them strongly (Banks and Salapatek, 1981). In turn, faces might be preferred simply because they have strong spatial-frequency components in the ranges that are most visible to newborns.

However, studies have found that the LSM fails to account for the responses to face-like stimuli. For instance, some of the face-like patterns preferred by newborns have a lower amplitude spectrum in the visible range (and thus lower expected LSM response) than patterns that are less preferred (Johnson and Morton, 1991). In particular, the LSM predicts that the newborn will respond equally well to a schematic face regardless of its orientation, because the orientation does not change the spatial frequency or the contrast. Instead, newborns prefer schematic face-like stimuli oriented right-side-up, even when those stimuli are at a less-optimal spatial frequency than another one that is upside-down (Valenza et al., 1996). Thus, matching the contrast sensitivity function does not explain face preferences, and a more complicated model is required.

2.2 Sensory models

The LSM is only a high-level abstraction of the properties of the early visual system, and it does not take into account the fact that processing in these areas is performed by neurons responding to limited portions of the visual field. Incorporating such general limitations, Acerra et al. (2001) recently developed a computational model of processing in the primary visual cortex (V1) that can account for some of the face preferences found in the Valenza et al. (1996) study. In this model, the preference for an upright face over an inverted one results from differences in the responses of some of the V1 neurons. Valenza et al. actually inverted only the internal facial features, not the entire pattern, which changed the distance between some of the features and the face outline (compare figures *5a,top* with *5d,top* in section 5). For the face-like pattern, the result of these border effects was to increase the response of some neurons preferring low spatial frequencies.

However, the Acerra et al. model was not tested with patterns from other studies, such as Johnson and Morton (1991), whose published stimuli did not have the requisite spacing between the internal features and the face outline. Moreover, Johnson and Morton used a white paddle against a light-colored ceiling, and such an edge would have a much lower contrast than the black-background patterns used in Valenza et al. (1996). Thus border effects are less likely to explain Johnson and Morton's face preference results. Most importantly, the Acerra et al. (2001) model was not tested with real images of faces, where the spacing of the internal features from the face outline varies widely depending on the way the hair falls. Thus overall the Acerra et al. (2001) model should not significantly prefer photographs of real faces over other similar images; the model introduced in this paper makes the opposite prediction.

Simion et al. (2001) recently proposed that besides the preferences suggested by the LSM, newborns merely prefer objects with a boundary of a certain size that contains denser patterns in the upper than the lower half. Nearly all of the face-like schematic patterns that have been tested in newborns do have this "top-heavy" property. Preliminary work also shows that some other patterns that are top-heavy but not face-like are preferred by newborns, relative to an inverted version of

the pattern (Simion et al., 2001). However, preferences for face-like patterns have not yet been compared directly with those for other top-heavy patterns, and thus it is not yet known whether the top-heavy preferences might be responsible for face preferences, or are instead independent of them.

To be tested computationally, the hypothesis that face preferences are due to top-heavy preferences would need to be made more explicit, with a specific mechanism for locating object boundaries and the relative locations of patterns with them. We expect that the result of such a test would be only a relatively small preference (if any) for photographs of real faces, compared to many other common stimuli. For beards, wide smiles, or wide-open mouths, control patterns might even be preferred over the face image, since such face images may no longer be top-heavy. Because the bulk of the current evidence suggests that instead the newborn preferences are more selective for faces, face-specific generated patterns will be used in the experiments in this paper.

2.3 Multiple-systems models

The most widely known conceptual model for newborn face preferences and later learning was proposed by Johnson and Morton (1991); it too has not yet been evaluated computationally. Johnson and Morton suggested that infant face preferences are mediated by two anatomically separate systems that they dubbed CONSPEC and CONLERN. CONSPEC would be a genetically fixed subcortical face-processing system, assumed to be located in the superior colliculus/pulvinar pathway. Johnson and Morton proposed that a CONSPEC responding to three blobs in a triangular configuration, one each for the eyes and one for the nose/mouth region, would account for the newborn face preferences (see figure 3a and 4d,top for examples). CONLERN would be a separate plastic cortical system, presumably the face-processing areas that have since been found in adults; this system would assume control only after about 6 weeks of age and thenceforth mediate environment-driven face preferences. This model is plausible, given that the superior colliculus is fairly mature in newborn monkeys and does seem to be controlling attention and other functions, although no face-selectivity has yet been found experimentally in that pathway in young animals (Wallace et al., 1997). The model also helps explain why infants at one month of age briefly show a reduced interest in faces, which could occur as attentional control shifts to the not-quite-mature cortical system (Johnson et al., 1991; Johnson and Morton, 1991).

However, subsequent studies showed that even newborns are capable of learning individual faces (Slater, 1993; Slater and Kirby, 1998). Thus if there are two systems, either both are plastic or both are functioning at birth, and thus there is no *a priori* reason why a single face-selective system would be insufficient. On the other hand, de Schonen et al. (1998) argue for *three* systems: a subcortical one mediating facial feature preferences at birth, another one mediating newborn learning (of objects and head/hair outlines; Slater, 1993), and a cortical system mediating older infant and adult learning of facial features. And Simion et al. (1998a) proposed that face selectivity relies on multiple systems *within* the cortex, maturing first in the dorsal stream but later supplanted by the ventral stream (which handles face-processing in the adult).

In contrast to the increasing complexity of these explanations, we propose that a single general-purpose plastic system is sufficient, if that system is first exposed to internally generated face-like patterns of neural activity.

2.4 Internally generated activity

Spontaneous neural activity has been documented in many cortical and subcortical areas as they develop, including the visual cortex, the retina, the auditory system, and the spinal cord (Feller et al., 1996; Lippe, 1994; Wong et al., 1993; Yuste et al., 1995; reviewed in O’Donovan, 1999; Wong, 1999). Such activity has previously been shown to be responsible for the segregation of the LGN into eye-specific layers before birth, indicating that internally generated activity does influence development (Shatz, 1990, 1996).

We will focus on one common type of spontaneous activity: ponto-geniculo-occipital (PGO) waves generated during rapid-eye-movement (REM) sleep. Developing embryos spend a large percentage of their time in what appears to be REM sleep, which suggests that it plays a major role in development (Roffwarg et al., 1966). During and just before REM sleep, PGO waves originate in the pons of the brain stem and travel to the LGN, visual cortex, and a variety of subcortical areas (see Callaway et al., 1987 for a review). PGO waves are strongly correlated with eye movements and with vivid visual imagery in dreams, suggesting that they activate the visual system as if they were visual inputs (Marks et al., 1995). PGO waves have been found to elicit different distributions of activity in different species (Datta, 1997), and interrupting them has been shown to increase the influence of the environment on development (Marks et al., 1995; Shaffery et al., 1999).

All of these characteristics suggest that PGO waves may be providing species-specific training patterns for development (Jouvet, 1998). However, due to limitations in experimental imaging equipment and techniques, it has not yet been possible to measure the two-dimensional shape of the activity resulting from the PGO waves (Rector et al., 1997). This paper predicts that if the PGO activity patterns have a simple configuration of three active areas surrounded by inactive areas, they can account for the measured face-detection performance of human newborns. These patterns are similar to those proposed by Johnson and Morton (1991). However, instead of genetically encoded templates implemented in a hard-wired subcortical face-detecting region, in our model they are simply training patterns for a general-purpose cortical system, the same system that later develops adult face perception abilities through training with real faces.

3 HLISSOM model

We will investigate the pattern-generation hypothesis using the HLISSOM model of the primate visual system, focusing on how sub-cortically generated patterns can drive the development of cortical areas. The HLISSOM architecture is shown in figure 1. The model consists of a hierarchy of two-dimensional sheets of neural units modeling different areas of the nervous system: two sheets of input units (the retinal photoreceptors and the PGO generator), two sheets of LGN units (ON-center and OFF-center), and two sheets of cortical units (“neurons”): the primary visual cortex (V1), and a higher-level area from the ventral processing stream with large receptive fields (here called the face-selective area, FSA¹). Because the focus is on the two-dimensional organization of

¹The FSA represents the first region in the infero-temporal portion of the ventral processing pathway that has receptive fields large enough to span a human face at close range (approximately 45° of visual arc). In newborn humans, the location of the FSA is not known, but based on its position in the ventral pathway, area V4v is a likely candidate

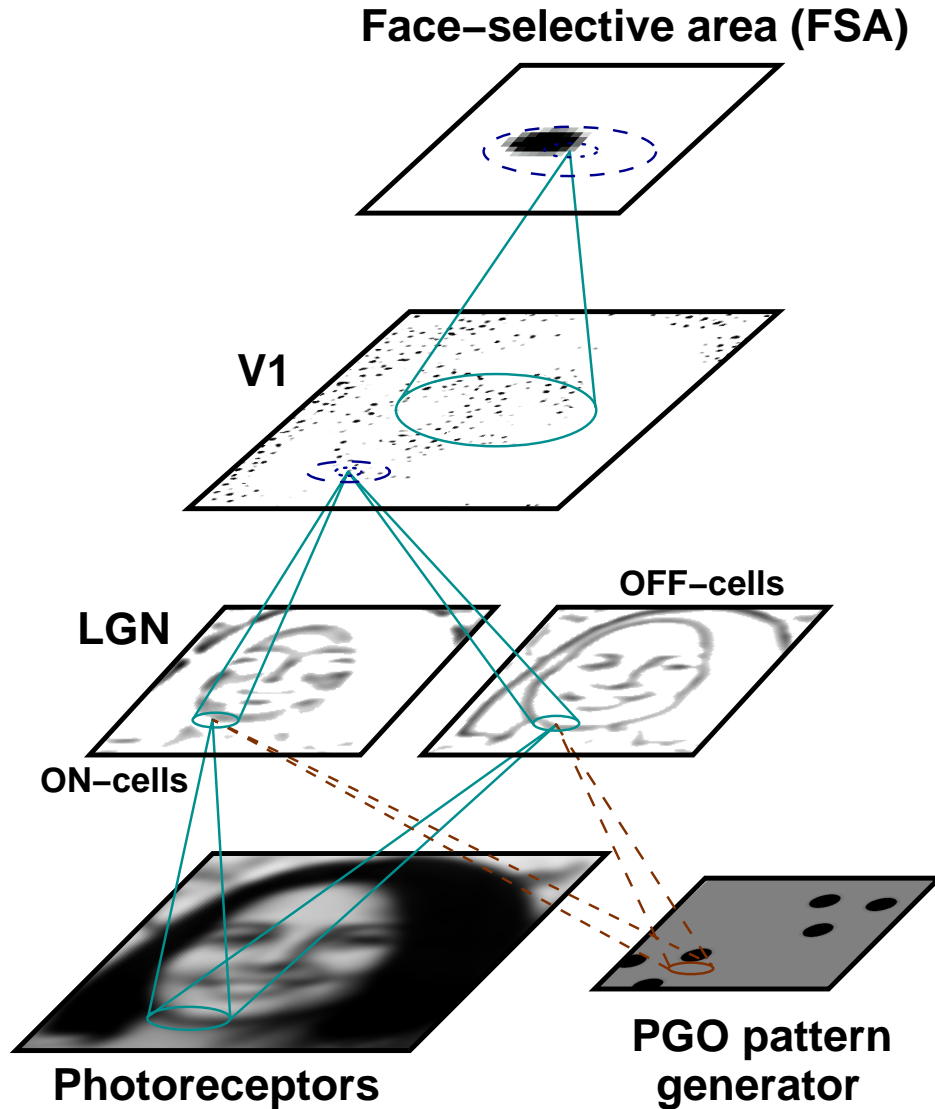


Figure 1: **Schematic diagram of the HLISSOM model.** Each sheet of units in the model visual pathway is shown with a sample activation pattern and the connections to one example unit. Grayscale visual inputs are presented on the photoreceptors, and the resulting activity propagates through afferent connections to each of the higher levels. Internally generated PGO input propagates similarly to visual input, but travels through the ponto-geniculate (PG) pathway to occipital cortex (i.e., V1), rather than the retino-geniculate pathway. As in humans and animals, activity may be generated either in the PGO area or the photoreceptors, but not both at once. In the cortical levels (V1 and FSA), activity is focused by lateral connections, which are initially excitatory between nearby neurons (dotted circles) and inhibitory between more distant neurons (dashed circles). The final patterns of lateral and afferent connections in the cortical areas develop through an unsupervised self-organizing process. After self-organization is complete, each stage in the hierarchy represents a different level of abstraction, transforming the raw image input into a more biologically relevant representation. The LGN responds best to edges and lines, suppressing areas with no information. The V1 response is further selective for the orientation of each contour; the patchiness is due to geometrical constraints on representing all orientations on a two-dimensional surface. The FSA represents the highest level of abstraction — the response of an FSA neuron signals the presence of a face at the corresponding location on the retina, which is information that an organism can use directly to control behaviors like visual fixation.

each cortical sheet, a neuron in V1 or the FSA corresponds to a vertical column of cells through the six anatomical layers of that area of the cortex.

Within each sheet, neurons receive afferent connections from broad overlapping circular patches of units in the previous sheet(s) in the hierarchy; these patches are called *receptive fields* (RFs). Neurons in the cortical sheets also have reciprocal excitatory and inhibitory lateral connections to neurons within the same sheet. Lateral excitatory connections are short range, connecting each neuron with itself and its close neighbors in a circular radius. Lateral inhibitory connections extend in a larger radius, but also include connections to the neuron itself and to its neighbors.²

Previous models have explained how the connections in the LGN could develop from internally generated activity in the retina (Eglen, 1997; Haith, 1998). Because the HLISSOM model focuses on learning at the cortical level, all connection strengths to such subcortical neurons are set to fixed values. The weights to the cells in the ON and OFF channels of the LGN are set as follows: the center of each receptive field is mapped to the location in the input sheet corresponding to the location of the LGN unit, and the weight to a receptor is a fixed function of the distance from that center. For an ON-center cell, the receptor weights are calculated from the difference of two normalized Gaussians with widths σ_c (center) and σ_s (surround). The weights for an OFF-center cell are the negative of the ON-center weights; i.e., they are the surround minus the center instead of the center minus the surround.

The connections to neurons in the cortical sheets are initially unselective (either random, for afferent weights, or Gaussian-shaped, for lateral weights), and their strengths are subsequently modified through an unsupervised learning process. The learning process is driven by input patterns. Patterns may be presented on the photoreceptor sheet, representing external visual input, or they may be generated internally, reaching V1 through the PGO pathway. The PGO-generating region of the pons and its pathway to the LGN have not yet been mapped out in detail, but since the activity that results from the PGO waves is similar to that from visual input (Marks et al., 1995), for simplicity we assume that the retino-geniculate and PGO pathways are similar. Therefore, the PGO pattern input is modeled with an area like the photoreceptor sheet, connecting to the LGN in the same way.

At each training step, the activities of all units are initialized to zero, and a grayscale pattern is drawn on the input sheet (either the PGO generator or the photoreceptors; see figures 2*a,i* and 3*a* for examples). The cells in the ON and OFF channels of the LGN compute their responses as a scalar product of their fixed weight vector and the activity of units in their receptive fields (as in figures 2*b,j* and 3*b*). The response η_{ab} of ON or OFF-center cell (a, b) is calculated from the

(Haxby et al., 1994; Rolls, 1990). The generic term “face-selective area” is used rather than V4v to emphasize that the model results do not depend on the precise location or architecture of this region, only on the fact that the region has receptive fields large enough to allow face-selective responses.

²For high-contrast inputs, long-range interactions must be inhibitory for proper self-organization to occur (Sirosh and Miikkulainen, 1994). Optical imaging and electrophysiological studies have indeed shown that long-range interactions in the cortex are inhibitory at high contrasts, even though individual lateral connections are primarily excitatory (Hirsch and Gilbert, 1991; Weliky et al., 1995). The model uses explicit inhibitory connections for simplicity since it is the high-contrast inputs that primarily drive adaptation in the Hebbian model. Long-range excitation can be included in the model as well, e.g. to demonstrate perceptual grouping phenomena (Choe, 2001; Choe and Miikkulainen, 2000).

weighted sum of the retinal activations in its receptive field as

$$\eta_{ab} = \sigma \left(\gamma_A \sum_{xy} \xi_{xy} \mu_{ab,xy} \right), \quad (1)$$

where σ is a piecewise linear approximation of the sigmoid activation function, ξ_{xy} is the activation of retina cell (x, y) , and $\mu_{ab,xy}$ is the corresponding afferent weight. An LGN neuron will respond when the input pattern is a better match to the central portion of the RF than to the surrounding portion; the response will be larger with higher contrast, subject to the minimum and maximum values enforced by the sigmoid. The constant afferent scaling factor γ_A is set so that the LGN outputs approach 1.0 in the highest-contrast regions of typical input patterns.

Cortical responses are computed from both afferent and lateral responses. The afferent response ζ_{ij} of V1 neuron (i, j) is calculated like the retinal sum in equation 1, with an additional divisive (shunting) normalization to increase response selectivity:

$$\zeta_{ij} = \frac{\gamma_A \sum_{\rho ab} \xi_{\rho ab} \mu_{ij,\rho ab}}{1 + \gamma_N \sum_{\rho ab} \xi_{\rho ab}}, \quad (2)$$

where $\xi_{\rho ab}$ is the activation of unit (a, b) in sheet ρ (either the ON channel or the OFF channel³), and $\mu_{ij,\rho ab}$ is the corresponding afferent weight. The normalization strength γ_N starts at zero and is gradually increased over training as neurons become more selective, and the scaling factor γ_A is set to compensate so that the afferent response ζ will continue to have values in the range 0 to 1.0 for typical input patterns.

The FSA computes its afferent response just as V1 does, except that sheet ρ in equation 2 is the settled response (described below) of V1, instead of the response of an ON or OFF channel of the LGN.

The initial response of a cortical neuron, both in V1 and the FSA, is computed from the afferent response only:

$$\eta_{ij}(0) = \sigma(\zeta_{ij}), \quad (3)$$

where σ is a piecewise linear approximation of the sigmoid activation function. After the initial response, the cortical activity evolves over a very short time scale through lateral interaction. At each subsequent time step, the neuron combines the afferent response ζ with lateral excitation and inhibition:

$$\eta_{ij}(t) = \sigma \left(\zeta_{ij} + \gamma_E \sum_{kl} E_{ij,kl} \eta_{kl}(t-1) - \gamma_I \sum_{kl} I_{ij,kl} \eta_{kl}(t-1) \right), \quad (4)$$

where $E_{ij,kl}$ is the excitatory lateral connection weight on the connection from neuron (k, l) to neuron (i, j) , $I_{ij,kl}$ is the inhibitory connection weight, and $\eta_{kl}(t-1)$ is the activity of neuron (k, l) during the previous time step. The scaling factors γ_E and γ_I determine the relative strengths of excitatory and inhibitory lateral interactions.

³Afferent inputs from additional ON and OFF channels with different peak spatial frequencies (i.e. different σ_c and σ_s) can be included in the same way. Only a single channel of each type was used here, so that a smaller V1 could be used and so that the V1 output would be easier to interpret.

While the cortical response in V1 and the FSA is settling, the afferent response remains constant. The cortical activity pattern in both areas starts out diffuse, but within a few iterations of equation 4, converges into a small number of stable focused patches of activity, or activity bubbles (as in figures 2*c,d,k* and 3*e,f*). After the activity has settled, the connection weights of each neuron are modified. Afferent and lateral weights in both areas adapt according to the same mechanism: the Hebb rule, normalized so that the sum of the weights is constant:

$$w_{ij,mn}(t + \delta t) = \frac{w_{ij,mn}(t) + \alpha \eta_{ij} X_{mn}}{\sum_{mn} [w_{ij,mn}(t) + \alpha \eta_{ij} X_{mn}]}, \quad (5)$$

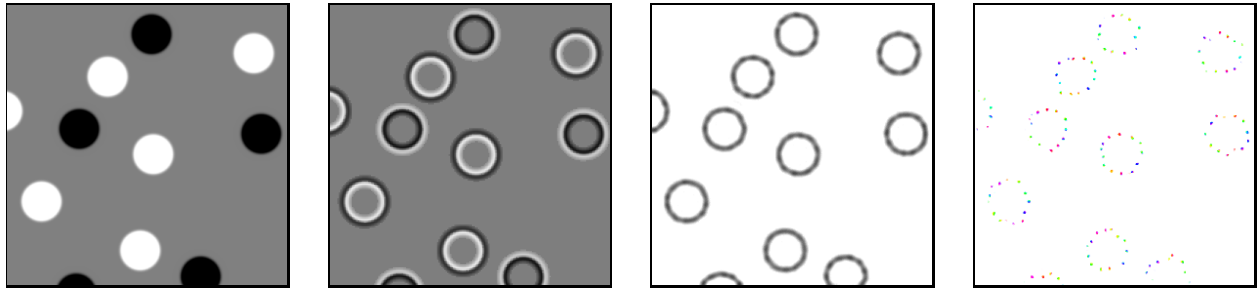
where η_{ij} stands for the activity of neuron (i, j) in the final activity bubble, $w_{ij,mn}$ is the afferent or lateral connection weight (μ , E or I), α is the learning rate for each type of connection (α_A for afferent weights, α_E for excitatory, and α_I for inhibitory) and X_{mn} is the presynaptic activity (ξ for afferent, η for lateral). The larger the product of the pre- and post-synaptic activity $\eta_{ij} X_{mn}$, the larger the weight change. At long distances, very few neurons in the cortical regions have correlated activity and therefore most long-range connections eventually become weak. The weak connections are eliminated periodically, resulting in patchy lateral connectivity similar to that observed in the visual cortex.

4 Experiments

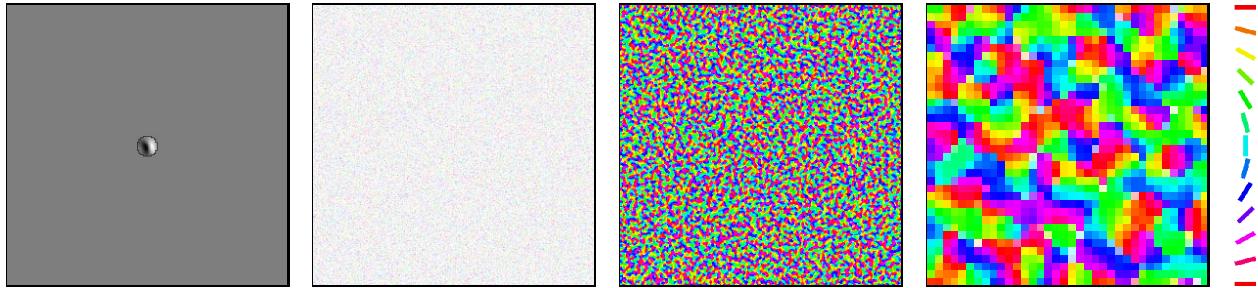
To model the presentation of large stimuli at close range (as in typical experiments with newborns), V1 parameters were generated using the scaling equations from Kelkar et al. (2000) to model a large V1 area (approximately 1600 mm² total) at a relatively low sampling density per mm² (approximately 50 neurons/mm²). The total area was chosen to be just large enough for the corresponding photoreceptor area to cover the required range, and the density was chosen to be the minimum that would still provide a V1 organization that matched animal maps. The input pattern scale was then set so that the spatial frequencies for which V1 response was maximal would match the frequency range to which newborns are most sensitive, as cited by Valenza et al. (1996).

The FSA parameters were generated using the scaling equations from Kelkar et al. (2000) to model a small region of units that have receptive fields large enough to span the central portion of a face. The final FSA activation threshold (lower bound on the sigmoid function in equations 3 and 4) was set to a high value so that the presence of a blob of activity in the FSA would be a criterion for the presence of a face at the corresponding location in the retina. This high threshold allows the FSA output to be interpreted unambiguously as a behavioral response, without the need for feedback from an additional higher level set of units modeling newborn attention and fixation processes.

The model consisted of 438×438 photoreceptors, 220×220 PGO generator units, 204×204 ON-center LGN units, 204×204 OFF-center LGN units, 288×288 V1 units, and 36×36 FSA units, for a total of 408,000 distinct units. Area FSA was mapped to the central 160×160 region of V1 such that even the units near the borders of the FSA had a complete set of afferent connections on V1, with no RF extending over the outside edges of V1. V1 was similarly mapped to the central 192×192 region of the LGN channels, and the LGN channels to the central 384×384 region of the photoreceptors and the central 192×192 region of the PGO generators. In the figures below,



(a) Generated pattern (b) LGN activity before training (c) V1 activity before training (d) V1 activity after training



(e) Afferent weights of one V1 neuron (f) Initial V1 orientation map (g) Final V1 orientation map (h) Detail of orientation map (g)



(i) Sample visual image (j) LGN response (k) V1 response

only the area mapped directly to V1 is shown, to ensure that all plots have the same size scale. The model requires 300 megabytes of physical memory to represent the 80 million connections in the two cortical sheets. The remaining parameters are listed in appendix A.

For simplicity, V1 and the FSA were self-organized in separate training phases. V1 was self-organized for 10,000 iterations on approximately 11 randomly located circular discs 25 PGO units wide (figure 2a). The background activity level was 0.5, and the brightness of each disc relative to this surround (either +0.3 or -0.3) was chosen randomly. The borders of each disc were smoothed into the background level following a Gaussian of half-width $\sigma = 1.5$.⁴

⁴The V1 training patterns were selected as well-defined approximations of internally generated spatial patterns

Figure 2: Training the V1 orientation map (color figure; on previous page). The input consisted of randomly-located circular patterns (*a*) approximating intrinsic neural activity waves. The ON and OFF LGN channels compute their responses based on this input; in (*b*), the ON response minus the OFF response is shown, with medium gray indicating regions with no response. The V1 neurons compute their responses based on the output of the ON and OFF channels. Initially V1 neurons have random weights in their receptive fields, and thus all neurons with RFs overlapping the active areas of the LGN respond (*c*). After training, the afferent weights of each neuron show a clear preference for a particular orientation (*e*); the plot shows the weights to the ON-center LGN sheet minus those to the OFF-center sheet, drawn on the full LGN sheet. The orientation maps (*f,g*) summarize the development of such preferences. Each neuron is color-coded according to its preferred orientation (using the color key at far right), brighter colors indicate a stronger selectivity. Initially the neurons are not selective for orientation (*f*), but after training, V1 develops an orientation map similar to those measured in experimental animals (*g*; Blasdel, 1992). The map covers a large area of V1; previous work corresponded to only the central 36×36 portion (*h*) of the 288×288 map in (*g*). For the same input (*a*), the self-organized network response is patchy (*d*), because only those neurons whose preferred orientation matches the orientations of local features in the image respond. In (*d*) each neuron is colored with the orientation it prefers, and the brightness of the color indicates how active that neuron is. The trained orientation map works well with natural images; each V1 neuron responds to the orientation of lines and edges in the corresponding region of the image (*i-k*).

After 10,000 iterations (10 hours on a 600MHz Pentium III workstation), the orientation map shown in figure 2*g-h* emerged, consisting of neurons that respond strongly to oriented edges and thick lines. The map organization is similar to maps measured in experimental animals (Blasdel, 1992).

After V1 training, the V1 weights were fixed to their self-organized values, and the FSA was allowed to activate and learn from the V1 responses. The FSA was self-organized for 10,000 iterations using 1–2 triples of circular dots (brightness -0.3) arranged in a triangular face-like configuration (as was proposed for a hardwired model by Johnson and Morton, 1991; figure 3*a*). Each dot was 20 PGO units wide. Each triple was placed at a random location at least 118 PGO units away from the centers of others (to avoid overlap), with a random angle drawn from a narrow ($\sigma = \pi/36$ radians) normal distribution around vertical.⁵

After 10,000 additional iterations with the three-dot patterns (13 hours on a 600MHz Pentium III workstation), the face-selective map shown in figure 3*h* emerged, consisting of neurons that respond most strongly to patterns similar to the training patterns.

such as spontaneous retinal waves (Feller et al., 1996). Miller (1994) has argued that stimuli that (like retinal waves) are much larger than the V1 receptive field size and are not themselves oriented could not account for the development of orientation selectivity, but the results here demonstrate otherwise. Simulations with other spatially extended patterns show that the precise shape is not important as long as it is smooth and large relative to the RF size of the orientation map.

⁵Because the face preferences found in newborns have all been for faces of approximately the same size (life-sized at a distance of around 20 cm), only a single training pattern size was used. As a result, the model will only be able to detect face-like patterns at one particular size scale. If response to multiple face sizes (i.e., distances) is desired, the spatial scale of the training patterns can be varied during self-organization (Sirosh and Miikkulainen, 1996). However, a larger map would be needed to represent the different sizes, and the resulting patchy responses would be less straightforward to interpret.

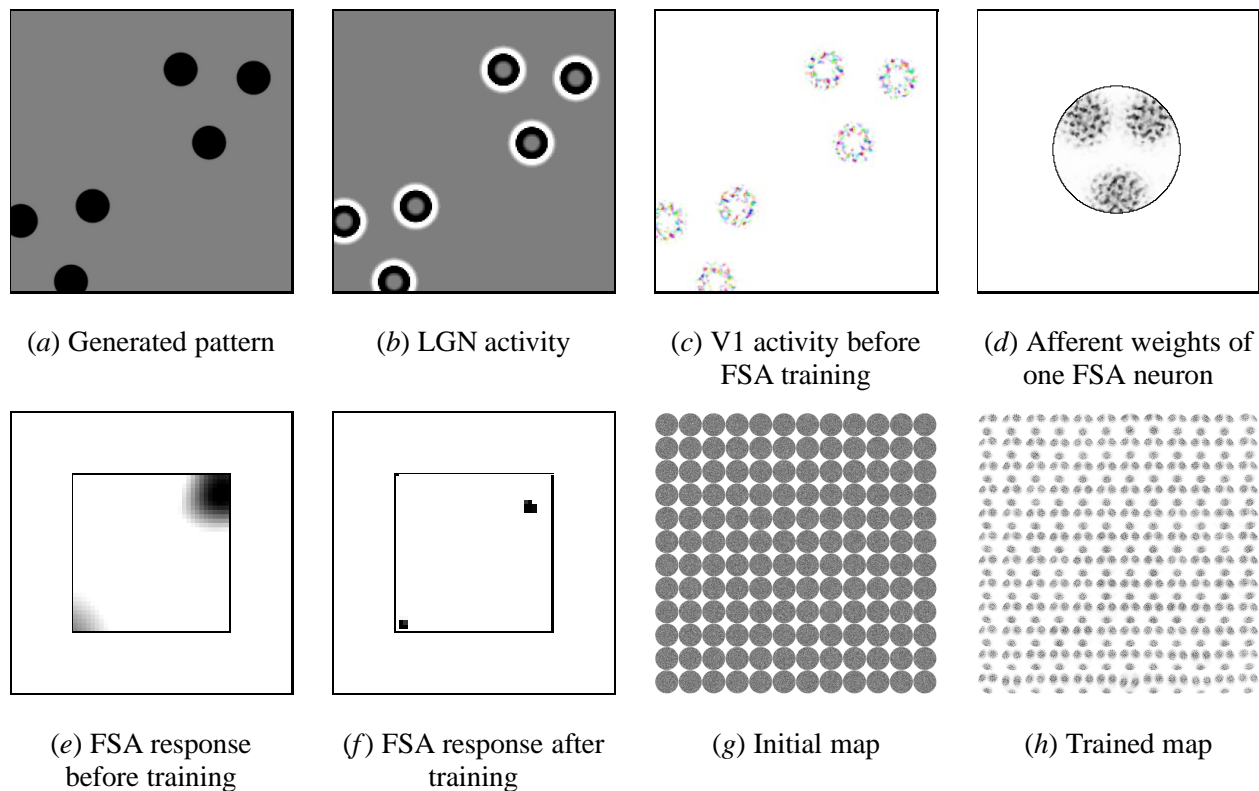


Figure 3: Training the FSA face map. (a) For training, the input consisted of simple three-dot configurations at random locations in the PGO layer and random nearly-vertical orientations; these patterns were chosen based on the experiments of Johnson and Morton (1991). (b-c) The LGN and V1 sheets compute their responses based on this input. Through self-organization, the FSA neurons develop RF weight profiles (d) that are selective for a range of V1 activity patterns like those resulting from the three-dot stimuli. Initially, the FSA neurons respond to any activity in their receptive fields (e), but after training only neurons with closely matching RFs respond (f). Plots (g) and (h) show the afferent weights for every third neuron in the face area. All neurons develop similar weight profiles, differing primarily by the position of their preferred stimuli on the retina, although the selectivity of those along the outside borders is slightly lower due to border effects. Area FSA thus develops into a face detection map, signaling the presence and location of face-like stimuli.

After self-organization, the cortical maps were tested on natural images and with the same schematic stimuli on which human newborns have been tested (Goren et al., 1975; Johnson and Morton, 1991; Simion et al., 1998b; Valenza et al., 1996). For all of the tests, the same set of model parameter settings described in appendix A were used. Schematic images were scaled to a brightness range (difference between the darkest and lightest pixels in the image) of 1.0. Natural images were scaled to a brightness range of 2.5, so that facial features in images with faces would have a contrast comparable to that of the schematic images.

The model provides detailed responses for each neural region; these can be compared with electrophysiological and imaging measurements in animals and humans. In addition, the model neural responses can be used to predict behavioral responses. To make such a prediction, we assume that the newborn attends most strongly to the stimuli that are most effective at activating his or her visual processing system, focusing on the highest level activated (as suggested by Cohen,

1998). That is, when two stimuli both activate the FSA in the model, we will assume that the one with the higher total FSA activation will be preferred. Similarly, with two stimuli activating only V1, the higher total V1 activation will be preferred.

When one stimulus activates only V1 and another activates both V1 and the FSA, the model's preference is strictly undefined. Such a preference would depend on the numerical value of the relative ability of each region to control the infant's looking behavior, which can only be determined empirically. We will assume in such cases that the pattern that produces FSA activity will be preferred by the newborn, unless the V1 activity is vastly greater than for typical patterns (as it is for e.g. a checkerboard pattern). Using these guidelines, the computed model preferences can be compared with the newborn's looking preferences.

5 Results

Figures 4 and 5 show that the responses of the self-organized network match the measured stimulus preference of newborns remarkably well, with the same relative ranking in each case where it was defined. For both newborns and the model, a checkerboard pattern (4a) is by far the most effective at activating V1, because it contains a large number of highly visible edges. Due to the magnitude of this V1 response, it is not surprising that the checkerboard pattern was also the most preferred by newborns.

Despite the large V1 response, the checkerboard did not activate the FSA strongly enough to reach the activation threshold. Face-like schematic patterns do activate the FSA (figure 4b-d; 5a,b,f,h), while most other patterns activate only V1 (figure 4e-i; 5c-e,g). In each case the preference rankings computed from the model response match those measured in newborns (Goren et al., 1975; Johnson et al., 1991; Johnson and Morton, 1991; Simion et al., 1998a; Valenza et al., 1996).

Interestingly, the model also showed clear preference for one case in which no significant preference was found in newborns: between the images in figure 5f and 5g. The model shows similar V1 responses to both patterns but an FSA response to the upright three-blob configuration only, and thus the upright configuration would be predicted to be preferred. As discussed below, newborns may not show a preference between these patterns because without a border they fail to attract sufficient newborn interest.

Overall, these results provide strong computational support for the speculation of Johnson and Morton (1991) that the newborn could simply be responding to a three-dot face-like configuration, rather than performing sophisticated face detection. Internally generated patterns provide an account for how such "innate" machinery can be constructed during prenatal development.

Presumably, an ability to detect faces in natural images underlies the newborn's preference for schematic faces. However, because natural images exhibit wide extremes of contrast and other types of variability that are difficult to handle in a computational model, no previous model of newborn face preferences has been tested with actual face images. HLISSOM was designed to allow such tests by incorporating nonlinear contrast normalization mechanisms at multiple levels of the visual system. The resulting network works remarkably well as a face detector for natural images, as shown in the examples in figure 6.

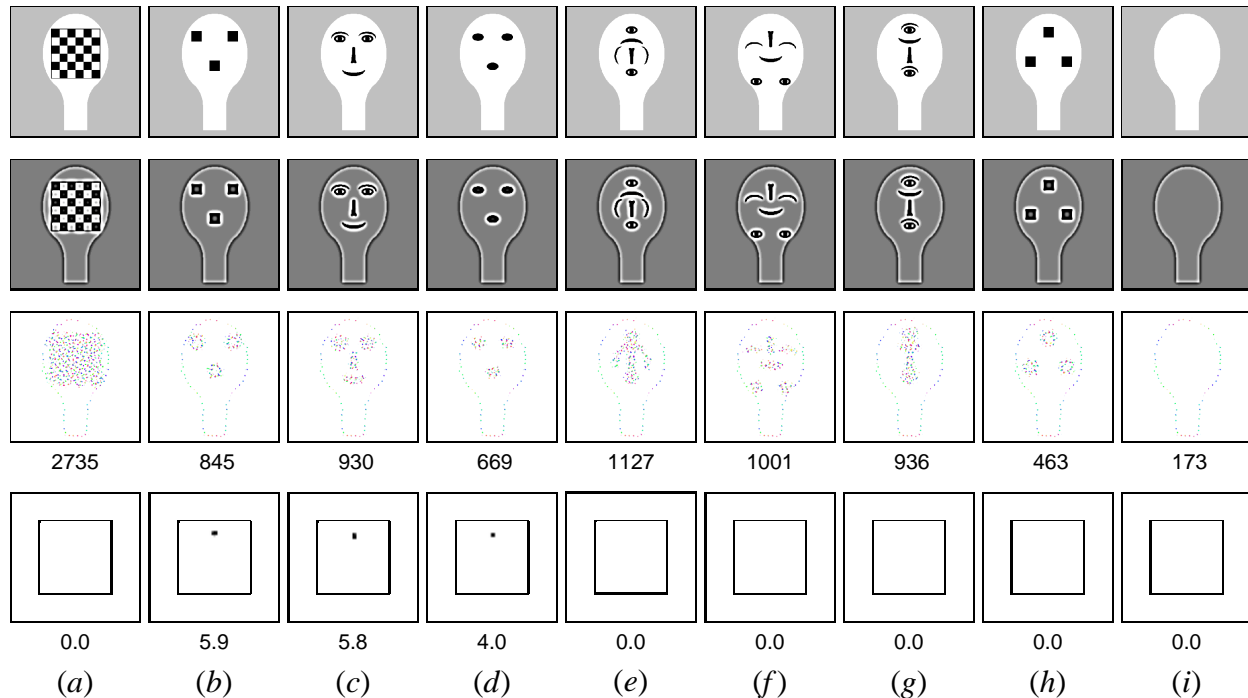


Figure 4: **Human newborn and model response to Goren et al.'s and Johnson et al.'s schematic images.** The schematic patterns in the top row have been presented to newborn human infants on head-shaped paddles moving at a short distance from the newborn's eyes (about 20 cm). The newborn's preference was determined by measuring the average distance his or her eyes or head will track each pattern, compared to other patterns. Goren et al. (1975) measured infants between 3 and 27 minutes after birth. They found that $c > f > i$ and $c > e > i$, where $x > y$ indicates that image (x) was preferred over image (y). Similarly, Johnson et al. (1991), in one experiment measuring within one hour of birth, found $c > e > i$. In another, measuring at an average of 43 minutes, they found $c > e$, and $c > h$. Finally, Johnson and Morton (1991), measuring newborns an average of 21 hours old, found that $a > (b, c, d)$, $b > d$, and $c > d$. The HLISSOM model has the same preference for each of these patterns, as shown in the images above. The top row shows the activation in the photoreceptor sheet. The second row shows the model LGN activations (ON minus OFF). The third row shows the V1 activations, with the numerical sum of the activities shown underneath. The bottom row shows the settled response of the FSA with the FSA activity sum underneath; activity at a given location signals the presence of a face-like stimulus at that location on the retina. The strongest V1 response by far is to the checkerboard pattern (a), which explains why the newborn would prefer that pattern over the others. The face-like patterns (b - d) are preferred over patterns (e - i) because of activation in the FSA. The details of the face-like patterns do not significantly affect the results — (b) is a better match to the training stimulus, but (c) has a higher V1 activation, and thus they both have similar FSA responses. Both (b) and (c) result in higher FSA activity than (d), and are thus preferred, because their greater edge length causes higher V1 activation. The remaining patterns are ranked by their V1 activity alone, because they do not activate the FSA. In all conditions tested, the HLISSOM model shows behavior remarkably similar to that of the newborns, and gives detailed computational insight into why these behaviors occur.

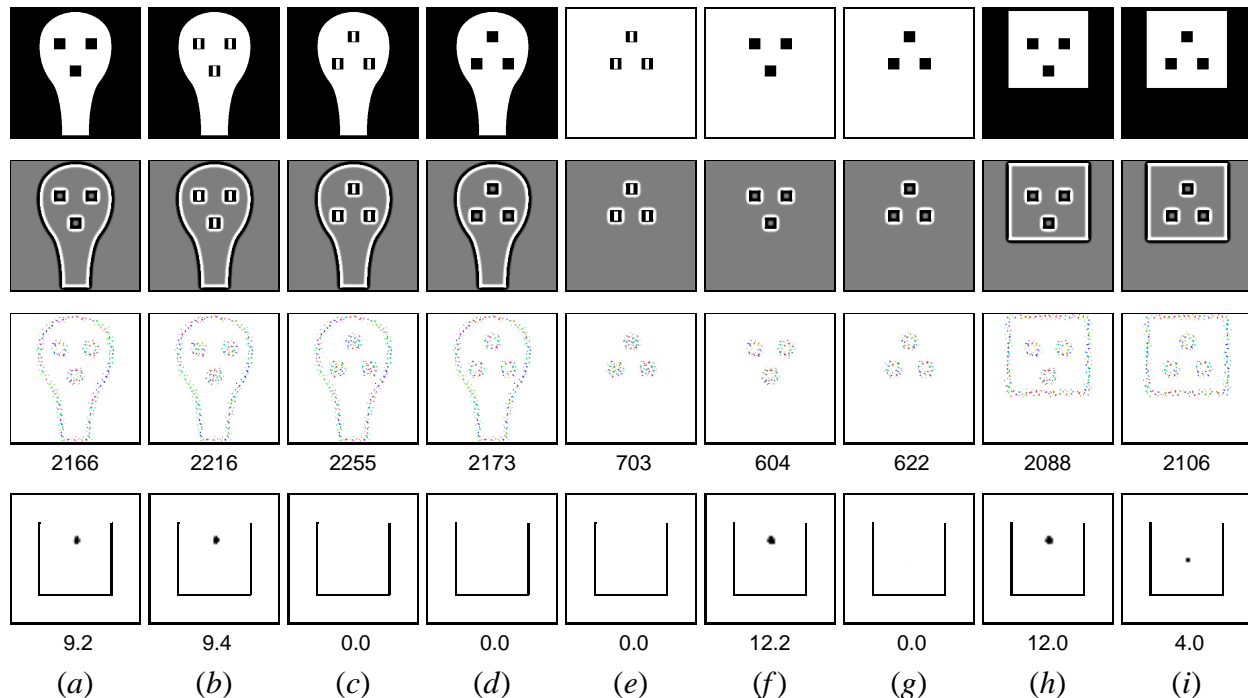


Figure 5: **Response to schematic images from Valenza et al. (1996) and Simion et al. (1998a).** Valenza et al. (1996) measured preference between static, projected versions of pairs of the schematic images in the top row, using newborns ranging from 24 to 155 hours after birth. They found the following preferences: $a > c$, $a > d$, $c > d$, and $e > g$. Simion et al. (1998a) similarly found a preference for $a > d$ and $h > i$. The LGN, V1, and FSA responses of the model to these images are displayed here as in figure 4. In all cases where the newborn showed a preference, the model preference matched that of the newborn. For instance, the model FSA responds to the face-like pattern (a) but not to the inverted version (d). Patterns that closely match the newborn’s preferred spatial frequencies (c,e) caused a greater V1 response than their corresponding lower-frequency patterns (d,g). Some non-face-like patterns with high-contrast borders can cause FSA activation (i); such spurious responses did not affect the predicted preferences. Simion et al. (1998a) found no preference between (f) and (g), but the model predicts that (f) would be preferred due to the FSA response.

In addition to such qualitative tests, the face detection performance of the map was tested quantitatively using two image databases: a set of images of 15 adult males without glasses photographed at the same distance against blank backgrounds (Achermann, 1995), and a set of 58 non-face images of various natural scenes (National Park Service, 1995). The face image set contained two views of each person facing forwards, upwards, downwards, left, and right, for a total of 150 face images; figure 6a shows an example of one frontal view. Each natural scene was presented at 6 different size scales, for a total of 348 non-face presentations. Overall, the results indicated very high face detection performance: the FSA responded to 91% (137/150) of the face images, but to only 4.3% (15/348) of the natural scenes.

Because the two sets of real images were not closely matched in terms of lighting, backgrounds, and distances, it is important to consider the actual response patterns to be sure that the differences in the overall totals are genuine. The FSA responded with activation in the location corresponding to the center of the face in 88% (132/150) of the face images. At the same time, the FSA had spurious responses in 27% (40/150) of the face images, i.e. responses in locations other than the

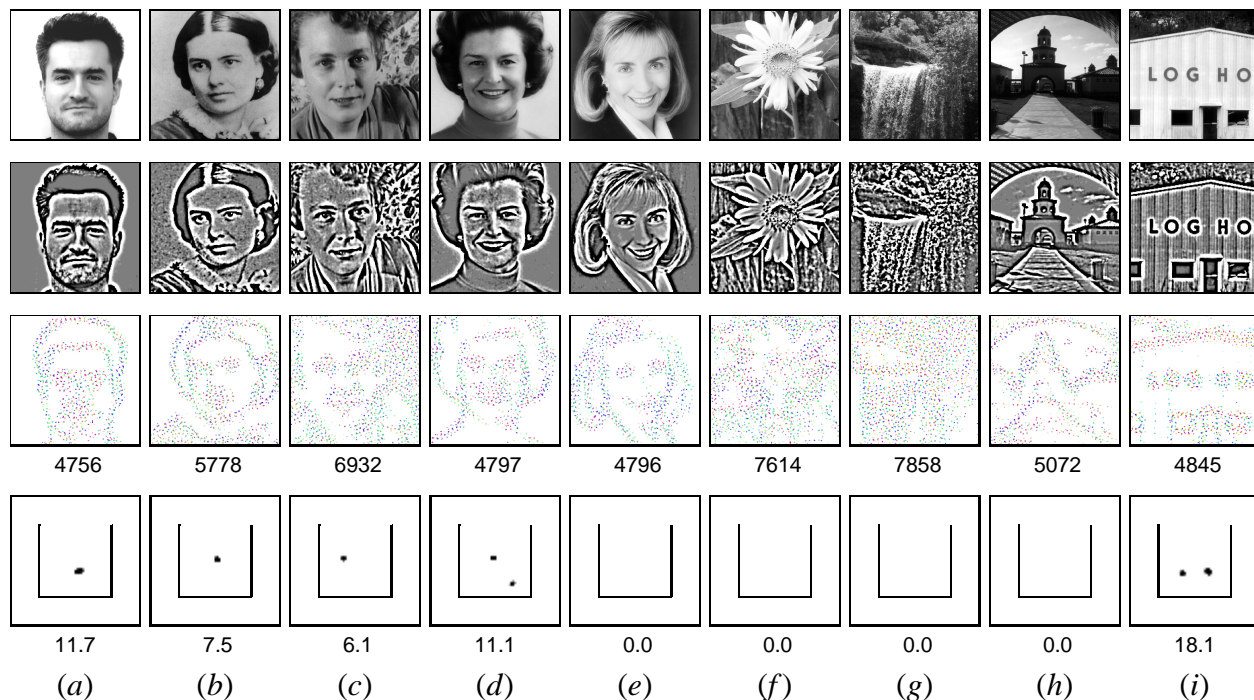


Figure 6: **Model response to natural images.** The top row shows a sample set of photographic images. The LGN, V1, and FSA responses of the model to these images are displayed as in figures 4 and 5. The FSA is indeed activated at the correct location for most real faces of the correct size and orientation (*a-d*). Just as important, the network is not activated for most natural scenes (*f-g*) and man-made objects (*h*). Besides human faces, the FSA responds to patterns causing a V1 response similar to that of a three-dot arrangement of contours (*d,i*), including related patterns such as dog and monkey faces (not shown). Response is low to images where hair or glasses obscures the borders of the eyes, nose, or mouth, and to front-lit downward-looking faces, which have low V1 responses from nose or mouth contours (*e*). The model predicts that newborns would show similar responses if tested. Credits: (*a*) copyright 1995 Bernard Achermand, (*b-e*) public domain; (*f-i*) copyright 1999-2001, James A. Bednar.

center of the face. Nearly half of the spurious responses were from less-selective neurons along the outer border of the model FSA (see figure 3*h*); these can be ignored. Most of the remaining spurious responses resulted from a genuine V1 eye or mouth response plus V1 responses to the hair or jaw outlines. Such responses would actually serve to direct attention to the general region of the face, although they would not pinpoint the precise center. For the natural scenes, most of the responses were from the less-selective neurons along the border of the FSA, and those responses can again be ignored. The remainder of the natural scene responses were to image regions that coincidentally had a triangular arrangement of three contour-rich areas, surrounded by areas with a more uniform shading.

In summary, the FSA responds to most human faces of the right size, signaling their location in the visual field. It does not respond to most other stimuli, except when they contain accidental three-dot patterns. The model predicts that human newborns will have a similar pattern of responses in the face-selective cortical regions.

6 Discussion and future work

The HLISSOM simulations show that internally generated patterns and self-organization can together account for newborn face detection. The novel contributions of the model are: (1) providing a rigorous, computational test of the Johnson and Morton (1991) three-dot conceptual model for newborn schematic face preferences, (2) demonstrating that preferences for three-dot patterns result in significant preference for faces in natural images, (3) proposing a specific, plausible way that a system for face detection could be constructed in the developing brain, (4) demonstrating how a single system can both exhibit genetically encoded preferences in the newborn and later learn from real faces, (5) demonstrating at a detailed, realistic neural level how a hierarchy of brain areas can self-organize and function, (6) showing how the reported newborn pattern preferences can result from specific feature preferences at the different levels of the hierarchy, and (7) showing that the effects of internally generated patterns should be considered when explaining genetically influenced behavior. The fundamental thesis is that pattern generation can combine the strengths of both genetic pre-specification and input-driven learning into a system more powerful than the sum of its parts. The following subsections relate the HLISSOM results to those from other models, and discuss future research directions.

6.1 Comparison to other models

Predictions from the HLISSOM model differ strongly from those of other recent models, and these predictions can be tested by experiment. HLISSOM predicts that under suitable conditions, newborns will show a significant preference for both real faces and face schematics over similar images without faces, independent of an external border. The Acerra et al. (2001) computational model makes the opposite predictions, namely that the preference for real images would be relatively weak, if any, and that any preference would depend on the spacing with the hair outline. Experimental evidence to date cannot yet decide between these two explanations. For instance, Simion et al. (1998a) found no face preference without an external contour, but also found that the shape of the contour “did not seem to affect the preference” for the patterns. The Acerra et al. model provides an explanation for why a border would be required, but predicts a large difference in response depending on the border shape and the spacing of features relative to the border. The HLISSOM model predicts that if future experiments do confirm a dependence on the border, the preference will not depend on such precise spacing. Instead the border would just satisfy a general requirement such as a minimum threshold of complexity, size, or contrast needed for responses to be measurable. Thus future experiments could distinguish between the two models by presenting photographic image stimuli and schematic stimuli with a variety of border shapes and spacings, to newborns and to each model.

When trained on a three-dot pattern, the HLISSOM model also makes predictions that differ from those of the Simion et al. (2001) “top-heavy preference” explanation. HLISSOM predicts that a pattern with three dots in the typical symmetrical arrangement would be preferred over the same pattern with both eye dots pushed to one side, despite both patterns being equally top-heavy. Finding instead a general preference for top-heavy patterns would imply that the internally generated patterns are closer in shape to a triangle than to three dots, i.e. are less face-specific.

However, because such a pattern would cause a lower preference for real faces than found in this study, such a revised model would predict that newborns also have a lower preference for real faces.

Many of the predictions of the fully trained HLISSOM model are similar to those of the genetically fixed CONSPEC model proposed by Johnson and Morton (1991). In fact, our earlier face preference model (which did not include V1) can be seen as the first model of the newborn's CONSPEC system to be implemented computationally, along with a concrete proposal for how such a system could be constructed during prenatal development (Bednar and Miikkulainen, 2000). The primary functional difference between the trained HLISSOM network and CONSPEC/CONLERN is that only cortical regions are responsible for the detection of faces in HLISSOM. Whether newborn face detection is mediated cortically or subcortically has been debated extensively, yet no clear consensus has emerged from behavioral studies (Simion et al., 1998a). Brain imaging may eventually settle this issue; in either case, the principles of pattern generation and self-organization can explain the prenatal construction of the face-processing system.

6.2 Future directions

Although the work reported in this paper covers self-organization up to the day of birth only, postnatal learning of faces is also very important for the development of adult face processing abilities. Once the genetically specified component has ensured that the newborn pays attention to faces, experience with real faces should drive the development of more sophisticated abilities. Preliminary further work has shown that a prenatally trained face map does learn from natural images more quickly and more effectively than does a map with a random or uniform initial state (Bednar and Miikkulainen, 2000). In future work we plan to extend these results, examining how postnatal visual experience and prenatal training combine to develop the adult face processing system.

The pattern generation hypothesis itself can potentially be verified directly by imaging the actual activity patterns produced during REM sleep. Very recent advances in imaging hardware have made limited measurements of this type possible (Rector et al., 1997), and imaging of the pons of experimental animals is planned for the near future (R. M. Harper, personal communication). It may soon be possible to test the assumptions and predictions of the HLISSOM model directly in developing animals.

Why would evolution have favored a pattern-generation approach over fixed hard-wiring or strictly general learning? One important reason is that the genome can encode the desired outcome, i.e. three-blob responses, independently of the actual architecture of the face processing hardware. As shown by Dailey and Cottrell (1999), even when the underlying cortical architecture is specified only in domain-general terms, specific regions can still become primarily devoted to face processing. The divide-and-conquer strategy of pattern generation would allow such an architecture to evolve and develop independently of the function, which could potentially enable the rapid evolution and development of more complex adaptive systems.

In terms of information processing, pattern generation combined with self-organization may represent a general way to solve difficult problems like face detection and recognition. Rather than meticulously specifying the final, desired individual, the specification need only encode a process

for constructing an individual through interaction with its environment. The result can seamlessly combine the full complexity of the environment with *a priori* information about the desired function of the system. In the future, this approach could be used for engineering complex artificial systems for real-world tasks, e.g. handwriting recognition, speech recognition, and language processing.

7 Conclusion

Internally generated patterns explain how genetic influences can interact with general adaptation mechanisms to specify, develop, and maintain a complex system, such as that responsible for face perception. The HLISSOM model of the visual system incorporates this idea, and is the first to self-organize both low-level and high-level cortical regions at the scale and detail needed to model such behavior realistically. The results match experimental data from newborns remarkably well, and for the first time demonstrate preferences for faces in real images. These results and future experiments with postnatal learning should help our understanding of the balance between environmental and genetic determinants of individuality, and increase our ability to construct complex artificial systems.

A Parameter values

Most of the model parameters were calculated using the scaling equations from Kelkar et al. (2000). Of those below, only the γ parameters and the sigmoid thresholds were set empirically for this study. For the LGN and V1, the values for these free parameters were set so that the output from each neural sheet would be in a useful range for the following sheet over the course of training. For the highest level (FSA), the free parameters were set so that the response would be nearly binary, and therefore an unambiguous criterion for the presence of a face-like pattern on the input. Apart from such threshold parameters, small parameter variations produce roughly equivalent results. The specific parameter values for each subcortical and cortical region follow.

For the LGN, σ_c was 0.4 and σ_s was 1.6. The afferent scale γ_A was 10.6, and the upper and lower thresholds of the sigmoid were 0.14 and 1.0, respectively.

For V1, the initial lateral excitation radius was 3.6 and was gradually decreased to 1.5. The lateral inhibitory radius of each neuron was 8, and inhibitory connections whose strength was below 0.01 were pruned away at 10,000 iterations. The lateral inhibitory connections were initialized to a Gaussian profile with $\sigma = 17$, and the lateral excitatory connections to a Gaussian with $\sigma = 2.8$, with no connections outside the nominal circular radius. Initially, the divisive normalization strength γ_N was zero, the afferent scale γ_A was 1.0, and the lateral excitation γ_E and inhibition strength γ_I were both 0.9. The γ values were constant during V1 training and were gradually increased over the course of FSA training to $\gamma_n = 4$, $\gamma_a = 3.25$, $\gamma_e = 1.2$, and $\gamma_i = 1.4$. The learning rate α_A was gradually decreased from 0.0035 to 0.00075, α_E from 0.059 to 0.0029, and α_I was a constant 0.00088. The lower and upper thresholds of the sigmoid were increased from 0.08 to 0.5 and from 0.63 to 0.86, respectively over the course of V1 training; the lower threshold was then decreased to 0.22 over the course of FSA training. The number of iterations for which the lateral connections were allowed to settle at each training iteration was initially 9, and was increased to 13.

For the FSA, the initial lateral excitation radius was 6.3 and was gradually decreased to 1.5. The lateral inhibitory radius of each neuron was 15.8, and inhibitory connections whose strength was below 0.0027 were pruned away at 20,000 iterations. The lateral inhibitory connections were initialized to a Gaussian profile with $\sigma = 33$, and the lateral excitatory connections to a Gaussian with $\sigma = 4.9$. The lateral excitation γ_E was 0.9 and the inhibition strength γ_I was 0.9. Initially, the divisive normalization strength γ_N was zero, the afferent scale γ_A was 3.0, and the lateral excitation γ_E and inhibition strength γ_I were both 0.9. These values were gradually changed over the course of FSA training to $\gamma_n = 9$, $\gamma_a = 10.6$, $\gamma_e = 0.4$, and $\gamma_i = 0.6$. The learning rate α_A was gradually decreased from 0.0001 to 0.000022, α_E from 0.025 to 0.013 and α_I was a constant 0.003. The lower and upper thresholds of the sigmoid were increased from 0.1 to 0.81 and from 0.65 to 0.88, respectively over the course of training. The number of iterations for which the lateral connections were allowed to settle at each training iteration was initially 9, and was increased to 13.

Acknowledgments

This research was supported in part by the National Science Foundation under grant #IIS-9811478. We thank Lisa Kaczmarczyk for critical comments on an earlier draft. Software, demonstrations, and related publications are available at <http://www.cs.utexas.edu/users/nn/>.

References

- Acerra, F., Burnod, Y., and de Schonen, S. (2001). Modelling aspects of face processing in early infancy. *Developmental Science*. In press.
- Achermann, B. (1995). Full-faces database. Copyright 1995, University of Bern, all rights reserved. <http://iamwww.unibe.ch/fkiwww/Personen/achermann.html>.
- Banks, M. S., and Salapatek, P. (1981). Infant pattern vision: A new approach based on the contrast sensitivity function. *Journal of Experimental Child Psychology*, 31(1):1–45.
- Bednar, J. A., and Miikkulainen, R. (1998). Pattern-generator-driven development in self-organizing models. In Bower, J. M., editor, *Computational Neuroscience: Trends in Research, 1998*, 317–323. New York: Plenum.
- Bednar, J. A., and Miikkulainen, R. (2000). Self-organization of innate face preferences: Could genetics be expressed through learning?. In *Proceedings of the 17th National Conference on Artificial Intelligence*, 117–122. Cambridge, MA: MIT Press.
- Blasdel, G. G. (1992). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12:3139–3161.
- Buonomano, D. V., and Merzenich, M. M. (1998). Cortical plasticity: From synapses to maps. *Annual Review of Neuroscience*, 21:149–86.
- Bushnell, I. W. R., Sai, F., and Mullin, J. T. (1989). Neonatal recognition of the mother's face. *The British Journal of Developmental Psychology*, 7:3–15.
- Callaway, C. W., Lydic, R., Baghdoyan, H. A., and Hobson, J. A. (1987). Pontogeniculooccipital waves: Spontaneous visual system activity during rapid eye movement sleep. *Cellular and Molecular Neurobiology*, 7(2):105–49.
- Choe, Y. (2001). *Perceptual Grouping in a Self-Organizing Map of Spiking Neurons*. PhD thesis, Department of Computer Sciences, The University of Texas at Austin, Austin, TX. Technical Report AI01-292.
- Choe, Y., and Miikkulainen, R. (2000). Contour integration and segmentation with self-organized lateral connections. Technical Report AI2000-286, Department of Computer Sciences, The University of Texas at Austin.
- Cohen, L. B. (1998). An information-processing approach to infant perception and cognition. In (Simion and Butterworth, 1998), 277–300.
- Constantine-Paton, M., Cline, H. T., and Debski, E. (1990). Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways. *Annual Review of Neuroscience*, 13:129–154.
- Dailey, M. N., and Cottrell, G. W. (1999). Organization of face and object recognition in modular neural network models. *Neural Networks*, 12(7):1053–1074.

- Datta, S. (1997). Cellular basis of pontine ponto-geniculo-occipital wave generation and modulation. *Cellular and Molecular Neurobiology*, 17(3):341–365.
- de Boysson-Bardies, B., editor (1993). *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht; Boston: Kluwer.
- de Schonen, S., Mancini, J., and Leigeois, F. (1998). About functional cortical specialization: The development of face recognition. In (Simion and Butterworth, 1998), 103–120.
- Easterbrook, M. A., Kisilevsky, B. S., Hains, S. M. J., and Muir, D. W. (1999). Faceness or complexity: Evidence from newborn visual tracking of facelike stimuli. *Infant Behavior and Development*, 22(1):17–35.
- Eglen, S. (1997). *Modeling the Development of the Retinogeniculate Pathway*. PhD thesis, University of Sussex at Brighton, Brighton, UK. Technical Report CSRP 467.
- Erwin, E., Obermayer, K., and Schulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: A critical comparison. *Neural Computation*, 7(3):425–468.
- Feller, M. B., Wellis, D. P., Stellwagen, D., Werblin, F. S., and Shatz, C. J. (1996). Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves. *Science*, 272:1182–1187.
- Field, T. M., Cohen, D., Garcia, R., and Greenberg, R. (1984). Mother–stranger face discrimination by the newborn. *Infant Behavior and Development*, 7:19–25.
- Gauthier, I., and Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology*, 11(2):219–224.
- Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nature Neuroscience*, 2(6):568–73.
- Goren, C. C., Sarty, M., and Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics*, 56(4):544–549.
- Haith, G. L. (1998). *Modeling Activity-Dependent Development in the Retinogeniculate Projection*. PhD thesis, Department of Psychology, Stanford University, Stanford, CA.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–30.
- Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J. M., Pietrini, P., and Grady, C. L. (1994). The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience*, 14:6336–6353.
- Hirsch, H. V. B. (1985). The role of visual experience in the development of cat striate cortex. *Cellular and Molecular Neurobiology*, 5:103–121.

- Hirsch, J. A., and Gilbert, C. D. (1991). Synaptic physiology of horizontal connections in the cat's visual cortex. *Journal of Neuroscience*, 11:1800–1809.
- Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40:1–19.
- Johnson, M. H., and Morton, J. (1991). *Biology and Cognitive Development: The Case of Face Recognition*. Oxford, UK; New York: Blackwell.
- Jouvet, M. (1998). Paradoxical sleep as a programming system. *Journal of Sleep Research*, 7(Suppl 1):1–5.
- Jouvet, M. (1999). *The Paradox of Sleep: The Story of Dreaming*. Cambridge, MA: MIT Press.
- Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11):4302–4311.
- Kelkar, A., Bednar, J. A., and Miikkulainen, R. (2000). Modeling large cortical networks with growing self-organizing maps. Technical report, Department of Computer Sciences, The University of Texas at Austin. Technical Report AI-2000-285.
- Kleiner, K. (1993). Specific vs. non-specific face-recognition device. In (de Boysson-Bardies, 1993), 103–108.
- Lippe, W. R. (1994). Rhythmic spontaneous activity in the developing avian auditory system. *Journal of Neuroscience*, 14(3):1486–1495.
- Marks, G. A., Shaffery, J. P., Oksenberg, A., Speciale, S. G., and Roffwarg, H. P. (1995). A functional role for REM sleep in brain maturation. *Behavioural Brain Research*, 69:1–11.
- Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (1997). Self-organization, plasticity, and low-level visual phenomena in a laterally connected map model of the primary visual cortex. In Goldstone, R. L., Schyns, P. G., and Medin, D. L., editors, *Perceptual Learning*, vol. 36 of *Psychology of Learning and Motivation*, 257–308. San Diego, CA: Academic Press.
- Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *Journal of Neuroscience*, 14:409–441.
- Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., and Kleiner-Gathercoal, K. A. (1999). Face perception during early infancy. *Psychological Science*, 10(5):419–422.
- National Park Service (1995). Image database. <http://www.freestockphotos.com/NPS>.
- O'Donovan, M. J. (1999). The origin of spontaneous activity in developing networks of the vertebrate nervous system. *Current Opinion in Neurobiology*, 9:94–104.

- Pascalis, O., and Slater, A. (2001). The development of face processing in infancy and early childhood: Current perspectives. *Infant and Child Development*, 10(1/2):1–2.
- Rector, D. M., Poe, G. R., Redgrave, P., and Harper, R. M. (1997). A miniature CCD video camera for high-sensitivity light measurements in freely behaving animals. *Journal of Neuroscience Methods*, 78(1-2):85–91.
- Rodman, H. R. (1994). Development of inferior temporal cortex in the monkey. *Cerebral Cortex*, 4(5):484–98.
- Roffwarg, H. P., Muzio, J. N., and Dement, W. C. (1966). Ontogenetic development of the human sleep-dream cycle. *Science*, 152:604–619.
- Rolls, E. T. (1990). The representation of information in the temporal lobe visual cortical areas of macaques. In Eckmiller, R., editor, *Advanced Neural Computers*, 69–78. New York: Elsevier.
- Shaffery, J. P., Roffwarg, H. P., Speciale, S. G., and Marks, G. A. (1999). Ponto-geniculo-occipital-wave suppression amplifies lateral geniculate nucleus cell-size changes in monocularly deprived kittens. *Brain Research. Developmental Brain Research*, 114(1):109–119.
- Shatz, C. J. (1990). Impulse activity and the patterning of connections during CNS development. *Neuron*, 5:745–756.
- Shatz, C. J. (1996). Emergence of order in visual system development. *Proceedings of the National Academy of Sciences, USA*, 93:602–608.
- Simion, F., and Butterworth, G., editors (1998). *The Development of Sensory, Motor and Cognitive Capacities in Early Infancy: From Perception to Cognition*. East Sussex, UK: Psychology Press.
- Simion, F., Macchi Cassia, V., Turati, C., and Valenza, E. (2001). The origins of face perception: Specific versus non-specific mechanisms. *Infant and Child Development*, 10(1/2):59–66.
- Simion, F., Valenza, E., and Umiltà, C. (1998a). Mechanisms underlying face preference at birth. In (Simion and Butterworth, 1998), 87–102.
- Simion, F., Valenza, E., Umiltà, C., and Dalla Barba, B. (1998b). Preferential orienting to faces in newborns: A temporal-nasal asymmetry. *Journal of Experimental Psychology: Human Perception and Performance*, 24(5):1399–1405.
- Sirosh, J., and Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics*, 71:66–78.
- Sirosh, J., and Miikkulainen, R. (1996). Self-organization and functional role of lateral connections and multisize receptive fields in the primary visual cortex. *Neural Processing Letters*, 3:39–48.

- Sirosh, J., Miikkulainen, R., and Bednar, J. A. (1996). Self-organization of orientation maps, lateral connections, and dynamic receptive fields in the primary visual cortex. In Sirosh, J., Miikkulainen, R., and Choe, Y., editors, *Lateral Interactions in the Cortex: Structure and Function*. Austin, TX: The UTCS Neural Networks Research Group. Electronic book, ISBN 0-9647060-0-8, <http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96>.
- Slater, A. (1993). Visual perceptual abilities at birth: Implications for face perception. In (de Boysson-Bardies, 1993), 125–134.
- Slater, A., and Kirby, R. (1998). Innate and learned perceptual abilities in the newborn infant. *Experimental Brain Research*, 123(1-2):90–94.
- Swindale, N. V. (1996). The development of topography in the visual cortex: A review of models. *Network – Computation in Neural Systems*, 7:161–247.
- Tovée, M. J. (1998). Face processing: Getting by with a little help from its friends. *Current Biology*, 8:R317–R320.
- Valenza, E., Simion, F., Cassia, V. M., and Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4):892–903.
- Wallace, M. T., McHaffie, J. G., and Stein, B. E. (1997). Visual response properties and visuotopic representation in the newborn monkey superior colliculus. *Journal of Neurophysiology*, 78(5):2732–2741.
- Walton, G. E., Armstrong, E. S., and Bower, T. G. R. (1997). Faces as forms in the world of the newborn. *Infant Behavior and Development*, 20(4):537–543.
- Walton, G. E., and Bower, T. G. R. (1993). Newborns form “prototypes” in less than 1 minute. *Psychological Science*, 4:203–205.
- Weliky, M., Kandler, K., Fitzpatrick, D., and Katz, L. C. (1995). Patterns of excitation and inhibition evoked by horizontal connections in visual cortex share a common relationship to orientation columns. *Neuron*, 15:541–552.
- Wong, R. O. L. (1999). Retinal waves and visual system development. *Annual Review of Neuroscience*, 22:29–47.
- Wong, R. O. L., Meister, M., and Shatz, C. J. (1993). Transient period of correlated bursting activity during development of the mammalian retina. *Neuron*, 11(5):923–938.
- Yuste, R., Nelson, D. A., Rubin, W. W., and Katz, L. C. (1995). Neuronal domains in developing neocortex: Mechanisms of coactivation. *Neuron*, 14(1):7–17.