Self-Organization of Innate Face Preferences: Could Genetics Be Expressed Through Learning?

James A. Bednar and Risto Miikkulainen

Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712
jbednar, risto@cs.utexas.edu

Abstract

Self-organizing models develop realistic cortical structures when given approximations of the visual environment as input, and are an effective way to model the development of face recognition abilities. However, environment-driven self-organization alone cannot account for the fact that newborn human infants will preferentially attend to face-like stimuli even immediately after birth. Recently it has been proposed that internally generated input patterns, such as those found in the developing retina and in PGO waves during REM sleep, may have the same effect on self-organization as does the external environment. Internal pattern generators constitute an efficient way to specify, develop, and maintain functionally appropriate perceptual organization. They may help express complex structures from minimal genetic information, and retain this genetic structure within a highly plastic system. Simulations with the CRF-LISSOM model show that such preorganization can account for newborn face preferences, providing a computational framework for examining how genetic influences interact with experience to construct a complex system.

Introduction

Faces appear to be very special objects for a human newborn (see Slater and Kirby 1998 for a review). Newborns will preferentially turn their eyes or head towards face-like stimuli within a few minutes after birth (Goren, Sarty, and Wu 1975; Johnson and Morton 1991). The more attractive the face, the more readily they do this (Slater et al. 1998). Attractive faces are close to the mathematical average of the population, which suggests that newborns prefer prototypical instances of faces (Langlois and Roggman 1990). Yet newborns have not yet had any experience with faces, so it is unclear how they could have learned such a prototype. Furthermore, since newborns will even imitate facial expressions in others (Meltzoff and Moore 1997), they must have some machinery that can not only recognize facial expressions but also translate them to specific motor commands. Together, these capabilities suggest that there is a genetically-specified ability in humans (and presumably other primates) to detect and process face-like stimuli. Given that face shapes differ substantially across phylogeny, this ability must have a species-specific, genetically-determined foundation.

How can each organism encode a specification for such a specific function? Clearly there is not enough space available in the genome of a mammal to directly specify any sizable percentage of the connections in its nervous system (10^5 genes vs. 10^{15} connections; Shatz 1996.) Moreover, given the enormous post-natal flexibility of the nervous system (reviewed in Hirsch 1985), including the fact that infant face perception shows rapid learning in the first few hours and days after birth (Slater and Kirby 1998), how can the genetic specification be expressed within such a highly plastic system?

This article considers how self-organizing models, which have ordinarily focused on environment-driven development (see Swindale 1996 for a review), can also help explain genetically-driven development. Since newborn studies suggest that face detection has a strong genetic component, we will focus on that ability. The thesis is that pre-natal development uses the same learning mechanisms which extract regularities from the visual environment, but it uses them to extract regularities in training inputs generated internally under genetic control (Jouvet 1998; Marks et al. 1995; Roffwarg, Muzio, and Dement 1966; Shatz 1996). Instead of precisely specifying the organization of the brain, the genome simply encodes a developmental process that is based on genetically-determined patterns presented to a general self-organizing mechanism.

Using the RF-LISSOM self-organizing model (Receptive-Field Laterally Interconnected Synergetically Self-Organizing Map) we have previously shown how orientation maps in the primary visual cortex can form based on spontaneously-generated retinal waves (Bednar and Miikkulainen 1998). This paper introduces CRF-LISSOM, which extends RF-LISSOM with a simple model of the eye so that it can be tested with natural photographic images. We then demonstrate how patterns found during REM sleep can account for a newborn's apparently innate predisposition for faces. These simulations represent a first step towards understanding genetic expression in a highly adaptive system.

CRF-LISSOM Model

The architecture for the CRF-LISSOM model is shown in figure 1, and will be briefly reviewed below. (For the detailed equations, see Sirosh, Miikkulainen, and Bednar 1996.) The model consists of a hierarchy of two-dimensional layers of neurons whose connections adapt through Hebbian learning. The input to the model is an activity pattern on a sheet of photoreceptors or PGO generator

Copyright © 2000, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

From the *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000)*, 2000, Cambridge, MA: MIT Press, pp. 117-122.

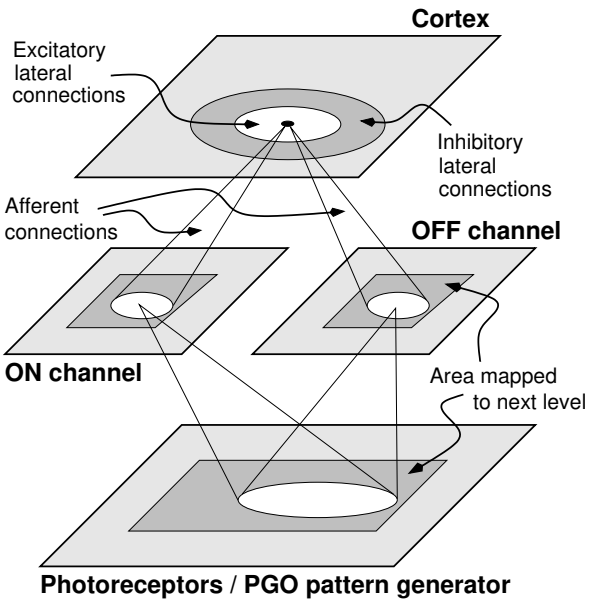
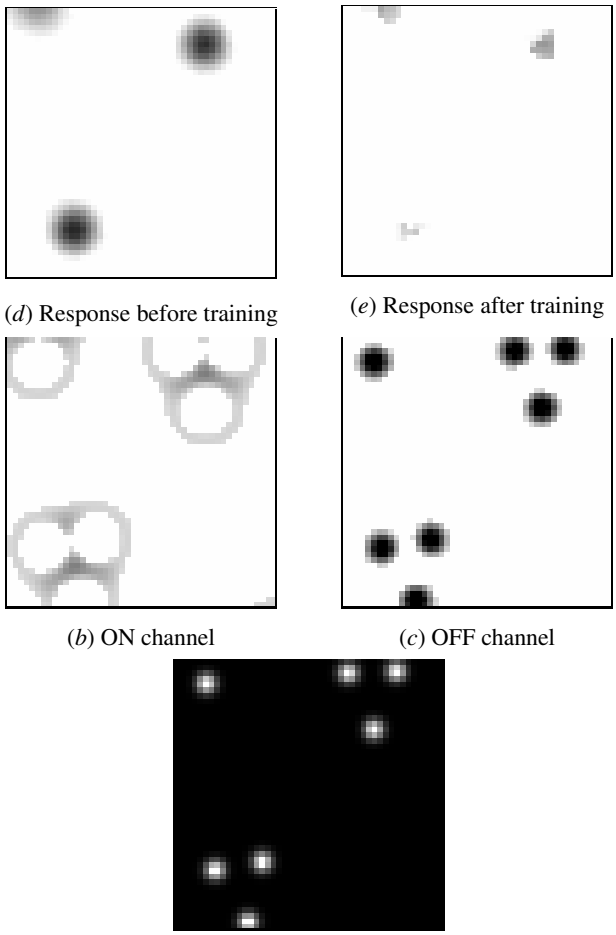


Figure 1: **The CRF-LISSOM model.** A schematic diagram of the CRF-LISSOM network, with connections shown for a single cortical neuron and for an ON-center and OFF-center ganglion cell. Activity propagates from the photoreceptor layer (bottom) up to the cortex and then spreads laterally within the cortex; weights are adapted when the activity settles.



(*d*) Response before training (*e*) Response after training

(*b*) ON channel (*c*) OFF channel

(*a*) Generated pattern

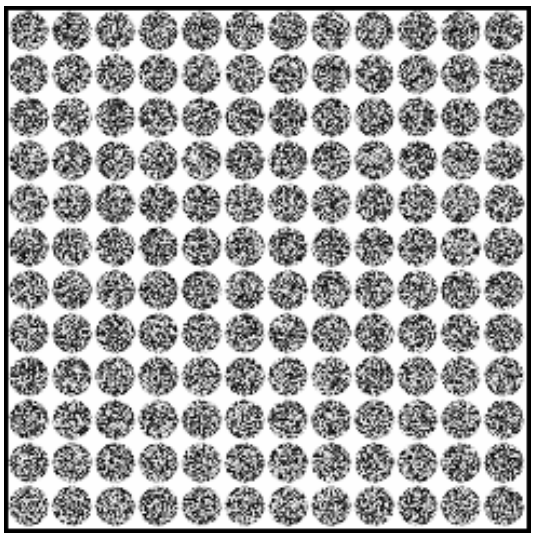
Figure 2: **Training Patterns for Face Detection.** (*a*) For training, the input consisted of simple three-dot configurations at random locations and random nearly-vertical orientations; these patterns were chosen based on the experiments of Johnson and Morton (1991). (*b-c*) The ON and OFF channels compute their responses based on this input, and (*d-e*) the cortical neurons compute their responses based on the output of the ON and OFF channels. Initially, the neurons are unselective (*d*), but after self-organization, they respond only to patterns similar to the three-dot stimuli (*e*).

output (e.g. figure 2*a*). The cells in the ON- and OFF-center layers compute their responses as a scalar product of their receptive fields and a fixed Difference-of-Gaussian weight vector (e.g. figure 2*b-c*). The ON and OFF layers, also known as input channels, were not present in RF-LISSOM, since the earlier model was designed to work directly with the output of the LGN. Like the ON and OFF cells, each cortical neuron also computes its initial response as a weighted sum of activity in its receptive field. After initial activation, the neural responses repeatedly propagate within the neural layer through the lateral connections, and evolve into coherent activity "bubbles" (figure 2*d-e*). After the activity stabilizes, weights of the active neurons are adapted by a normalized Hebb rule.

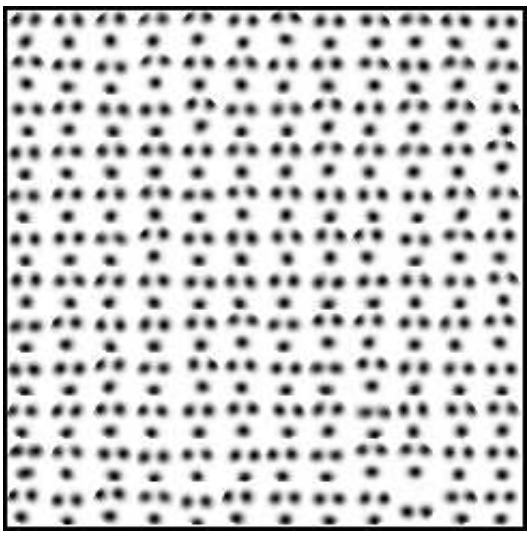
This model was initially developed to account for environment-driven self-organization in the primary visual cortex, where the input layer consists of photoreceptors in the retina, the middle layer of ON- and OFF-center ganglion cells, and the top layer of the cortical neurons (the LGN is bypassed for simplicity). Given elongated Gaussian patterns as input, the model shows how orientation columns and lateral connections between them form. The neural layer in CRF-LISSOM is very general, so the model can also be used for most of the cortical and subcortical areas that are organized as topographic maps. Different developmental phenomena can be modeled by manipulating properties of the input.

For the face-detection experiments, the input to the model is assumed to arise from 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 a major developmental role for this process (Roffwarg, Muzio, and Dement 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 as well as 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 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).

All of these characteristics suggest that PGO waves may be providing species-specific training patterns for develop-



(*a*) Initial map



(*b*) Trained map

Figure 3: **Prenatal development of the face map.** The afferent weights for the OFF-center cells for every fourth neuron in the cortex are shown by grayscale coding. Initially, the weight pattern is unselective, and is identical for both ON and OFF weights (*a*). After many iterations of the self-organizing algorithm, the OFF-center weights (*b*) have come to represent the inputs it has seen, and the neurons now respond only to patterns similar to the three-dot stimuli. The ON-center weights became approximately the photographic inverse of the OFF-center weights, and the lateral weights have a short-range, approximately Gaussian shape throughout self-organization.

ment (Jouvet 1998). However, due to limitations in experimental imaging equipment and techniques, the spatial shape of the PGO wave activity patterns has not yet been measured (Rector et al. 1997). This paper predicts that if these patterns have certain simple properties (shown in figure 2*a*), they can account for the measured face-detection performance of human newborns.

Three assumptions and extensions were made to the CRF-LISSOM model to account for the development of face detection. First, in line with recent neurobiological evidence that shows that much of the adult cortical face processing circuitry is present in the infant (Rodman 1994), newborn face detection is assumed to require only cortical circuitry, most likely in human area V4v and nearby areas (Haxby et al. 1994). Even though multiple cortical areas are likely to be involved, for simplicity a single functional layer is used in this model. Second, until more details are learned about how PGO waves affect the cortex, it is reasonable to assume that this pathway has similar properties as the pathway from the retina. Therefore, the same structure of topographic input and ON/OFF middle layer is used in the model. Third, to keep the size of the model manageable, the PGO patterns are assumed to vary little in size and orientation, and are presented statically. With larger maps, moving patterns and larger variations can be taken into account (Sirosh and Miikkulainen 1996; Sirosh, Miikkulainen, and Bednar 1996).

The input patterns consisted of multiple (3-4) triples of 2-D Gaussian dots on a 132×132 array representing the PGO waves. These stimuli are an implementation of the genetically-encoded template postulated by Johnson and Morton (1991), but they are used as a cortical training stimulus rather than as hard-wired sub-cortical function. Each triple was placed at a random location at least 30 units away from others, with a random angle from a narrow ($\sigma = \pi/36$ radians) normal distribution around vertical. A pair of 66×66 ON-center and OFF-center cell layers received input from the PGO array. Each ON/OFF cell had a fixed Difference of Gaussians (DoG) receptive field (RF) within the PGO array (center σ=1.0, surround σ=1.6; ratio is from Marr 1982). The cortical layer consisted of 48×48 neurons that received input from the ON/OFF cell layers and from their neighbors within the cortical layer. The receptive field for each neuron was placed at the location in the central 48×48 portion of the ON/OFF cell layer corresponding to the location of the neuron in the cortex, giving every neuron a complete set of afferent (input) connections. Initially, the afferent weights were random and the lateral weights had a smooth Gaussian profile; all weights adapted through Hebbian learning as patterns were presented on the PGO layer. The learning rates and other parameters were similar to those of Sirosh, Miikkulainen, and Bednar (1996), scaled for this cortex size.

Results

In 20,000 presentations, the ordered configuration shown in figure 3*b* emerged. As one would expect, the neurons in the cortical layer developed receptive fields that closely resemble the training inputs used (figure 2*a*). The cortical map was then tested with the same schematic stimuli on which human newborns have been tested (Goren, Sarty, and Wu 1975; Johnson and Morton 1991). Assuming that the newborn attends most strongly to those stimuli that are most effective at activating his visual processing system (including face-selective areas), the activations in the network can be

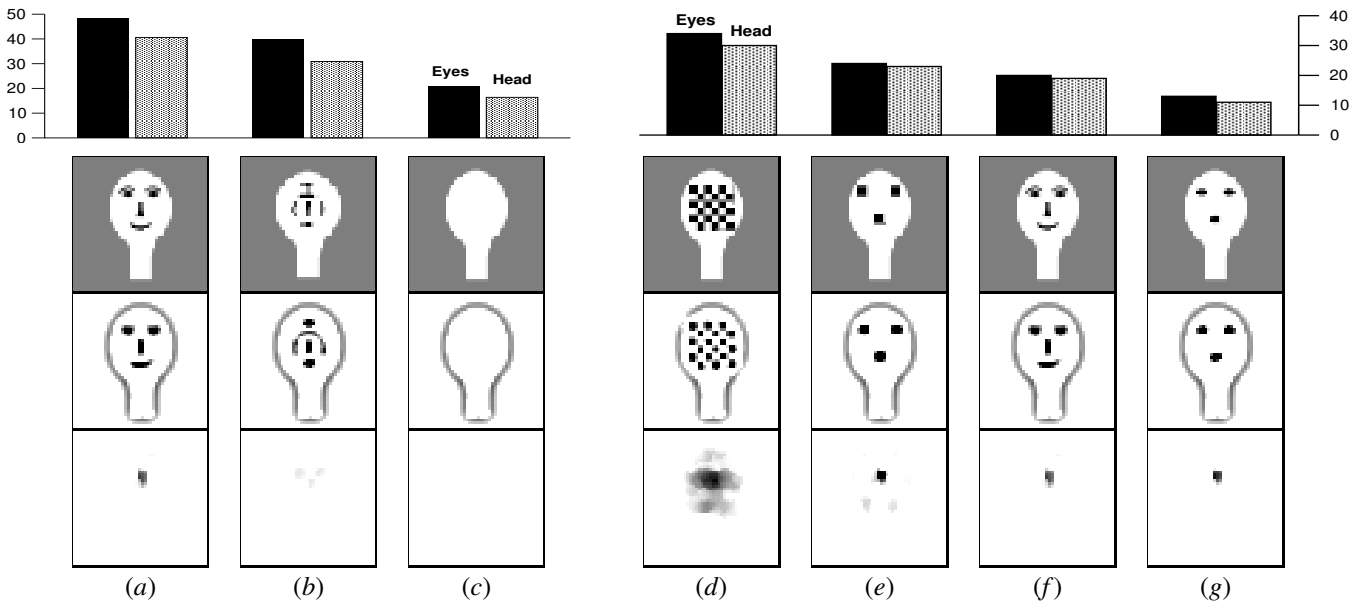


Figure 4: **Human newborn vs. model response to schematic images.** The graph at left shows the result of a study of human newborns tested with two-dimensional moving patterns within one hour of birth; the one at right shows the results from a separate study of newborns an average of 21 hours old (Johnson and Morton 1991). Each pair of bars shows the average newborn eye and head tracking, in degrees, for the image pictured below it; eye and head tracking showed similar trends here and thus either may be used as a measure. Since the procedures and conditions differed between the two studies, only the relative magnitudes should be compared. The second row of images shows the OFF-center activations in the CRF-LISSOM model; the ON-center cells responded as suggested by figure 2*b*. The bottom row shows the settled response of the cortical layer. Remarkably, the settled response in the model reflects the same relative preferences for all stimuli as in the newborn. Overall, the study at left demonstrates that the newborn and the network both respond to face-like stimuli more strongly than to simple control conditions. The study at right shows what features of the face-like pattern are essential. In all cases, the CRF-LISSOM model shows behavior remarkably similar to that of the newborns, and gives detailed computational insight into why these behaviors occur.

compared directly with the infant's looking preferences.

Remarkably, the responses of the pre-trained network turned out to match to the measured stimulus preference of newborns very well, with the same relative ranking in each case (figure 4). For both newborns and the model, a checkerboard pattern (4*d*) is the most effective, because it activates a very large region of neurons that mutually reinforce each other, despite the pattern being a poor match for a face. (Of course, in a newborn, other object processing areas are presumably even more effectively stimulated by a checkerboard, and thus such a pattern would be salient even if it did not activate face detection areas.) A simple three-square-dot pattern (4*e*) is also more effective than the more face-like pattern (4*f*), suggesting that the details of the input features are irrelevant. An oval-dot pattern evokes somewhat lesser activity than the other face-like stimuli since it has less energy in the spatial frequency range tested here (figure 4*g*); the difference is slight for the model due to the limited resolution of the model retina. These results provide 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.

The model also shows some interesting spurious responses. Figure 4*e* shows two extra responses (the lighter blobs) caused by fortuitous matches with the three-dot receptive fields – e.g. the "mouth" dot can also match the right eye blob of a receptive field, while the face border matches the left eye and the mouth regions. These false matches actually serve to make the pattern even more salient. This prediction could be tested in newborns using stimuli with a single dot in one of the three positions; the newborn should respond more strongly to a pattern with a dot in the mouth position than to one in either eye position (assuming the training pattern has equal-sized dots and is processed using ON/OFF channels like those in the model).

Presumably, an ability to detect faces in natural images underlies the newborn's preference for schematic faces, so the CRF-LISSOM network was also tested with natural stimuli. As shown in figure 5, the pre-trained map also performs remarkably well on natural images. The map responds strongly to most human faces of the right size, and 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 brain regions if they are using a template with three equal-sized dots processed with ON- and OFF-center cells. Conversely, if human newborns show different responses, then it is likely that they are using a more sophisticated template, the shape of which can be recovered from the responses. Such experiments could constitute an important validation of the model, and further our understanding of the development of face detection in general.

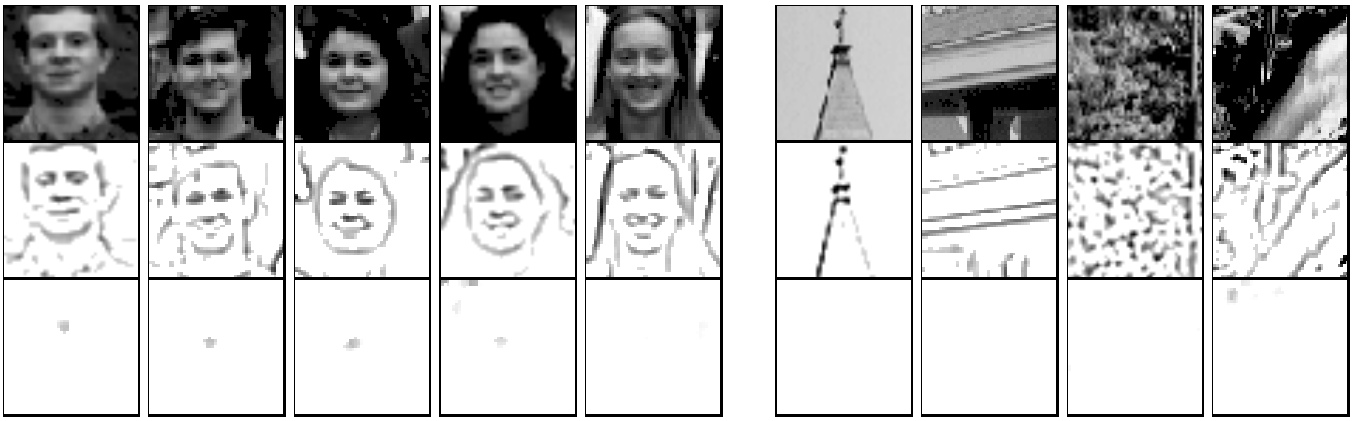


Figure 5: **Model response to natural images.** The top row shows the central 96×96 portion of a 132×132 image presented to the input layer (data from Rowley, Baluja, and Kanade 1998). As in figure 4, the second row shows the activation of the OFF-center cells, while the bottom row shows the response of the pre-natally trained network. The position of the network activation corresponds to the retinal position of the object to which it is responding. The pre-natally trained network is indeed activated for most real faces (*a-d*) and for some non-face patterns present in nature scenes (*i*) and complicated backgrounds (*d*). Just as important, the network is not activated for most man-made objects (*f-g*) and natural scenes (*h*); it is also inactive for faces with broad smiles (*e*). The model predicts that newborns would show similar responses.

Discussion and future work

The CRF-LISSOM simulations show that internally-generated patterns and self-organization can together account for newborn face detection. Pattern-generation represents a synthesis of the strengths of genetic specification and environment-driven learning. Once the genetically-specified component has ensured that the newborn pays attention to faces, experience with real faces drives the development of more sophisticated abilities. Preliminary further work has shown that the pretrained map does learn from natural images more quickly and more effectively than does a map with a random or uniform initial state.

The results reported in this paper suggest constraints on the types of patterns that may be used by the PGO pattern-generating system. For instance, if the patterns use the three equal-sized dots proposed by Johnson and Morton (1991), the system will not perform well on broad, fixed smiles; this prediction could be tested in the human newborn. More direct confirmation of the pattern-generation approach would come from imaging measurements of 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 is planned for the near future (R. M. Harper, personal communication). So it may soon be possible to test the assumptions and the predictions of the model directly in the developing animal.

Johnson and Morton (1991) originally proposed that newborn face detection occurred subcortically, and a face-selective subcortical pathway has recently been found in the adult (Morris, Ohman, and Dolan 1999). However, it is not known whether this pathway is present in infants, while face-selective inferior temporal cortex cells have been found in the youngest primates yet studied (6 weeks old; Rodman 1994). Thus since the current evidence points to a cortical mechanism, only cortical face-selective regions were modeled here. Incorporating subcortical influences would complicate the model, but the same principles of pattern generation and self-organization would apply.

Since PGO waves continue into adulthood, internal pattern generation may also have a lifelong role in maintaining brain function while the system adapts to environmental stimuli, particularly when subcortical pathways are considered. For instance, the patterns may help sustain connections between the phylogenetically older regions that provide basic visual functionality (such as the superior colliculus, the pulvinar, and the amygdala) and the more comprehensive object processing performed in the various sensory and motor areas of the neocortex. Presenting similar patterns to separate areas simultaneously would strengthen the connections between them, ensuring that despite pervasive adaptation they continue to work together in their genetically-specified roles. Thus ongoing internal pattern generation and self-organization may help explain how multiple highly-plastic processing streams are integrated into a coherent system.

From a larger information-processing perspective, pattern generation combined with self-organization may represent a general way to solve difficult problems like face detection and recognition. Simple inputs, easy to specify and generate, combined with simple learning mechanisms that also learn from the environment, together allow a complex system to be specified and developed. (Here "complex" is used in the sense of Gell-Mann's effective complexity, measuring the total amount of regularity of all types; Gell-Mann and Lloyd 1996). The specification for the system need only encode a process for constructing an individual along with interaction from its environment, rather than meticulously specifying the individual itself. Yet even so the result can incorporate

the full complexity of the environment as well as *a priori* information about the desired function of the system, with seamless integration between the two. This approach could be used for engineering complex artificial systems, e.g. for handwritten character recognition, speech recognition, and language processing.

Conclusion

We propose that internally-generated patterns represent a general mechanism that allows an organism to specify, develop, and maintain complex functional structures. CRF-LISSOM simulations have shown how initial orientation maps and face-selective maps could develop from non-visual inputs. Future work will examine how these genetic factors interact with learning in the adult, as well as how distinct cortical areas can develop simultaneously. Such experiments with self-organizing models should greatly improve our understanding of the balance between environmental and genetic determinants of individuality, and should increase our ability to construct complex artificial systems.

Acknowledgments This research was supported in part by the National Science Foundation under grant #IIS-9811478. Computer time for the simulations was provided by the Texas Advanced Computing Center at the University of Texas at Austin. Software, demonstrations, and related publications for CRF-LISSOM and RF-LISSOM are available at `http://www.cs.utexas.edu/users/nn/`.

References

Bednar, J. A., and Miikkulainen, R. 1998. Pattern-generator-driven development in self-organizing models. In Bower, J. M., ed., *Computational Neuroscience: Trends in Research, 1998*, 317–323. New York: Plenum.

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.

Datta, S. 1997. Cellular basis of pontine ponto-geniculo-occipital wave generation and modulation. *Cellular and Molecular Neurobiology* 17(3):341–365.

Gell-Mann, M., and Lloyd, S. 1996. Information measures, effective complexity, and total information. *Complexity* 2(1):44–52.

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.

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.

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.

Langlois, J. H., and Roggman, L. A. 1990. Attractive faces are only average. *Psychological Science* 1:115–121.

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.

Marr, D. 1982. *Vision*. New York: Freeman.

Meltzoff, A. N., and Moore, A. K. 1997. Explaining facial imitation: A theoretical model. *Early Developmental Parenting* 6:179–192.

Morris, J. S.; Ohman, A.; and Dolan, R. J. 1999. A subcortical pathway to the right amygdala mediating "unseen" fear. *Proceedings of the National Academy of Sciences, USA* 96(4):1680–1685.

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.

Rowley, H. A.; Baluja, S.; and Kanade, T. 1998. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20:23–38.

Shatz, C. J. 1996. Emergence of order in visual system development. *Proceedings of the National Academy of Sciences, USA* 93:602–608.

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., eds., *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., and Kirby, R. 1998. Innate and learned perceptual abilities in the newborn infant. *Experimental Brain Research* 123(1-2):90–94.

Slater, A.; Von der Schulenburg, C.; Brown, E.; Badenoch, M.; Butterworth, G.; Parsons, S.; and Samuels, C. 1998. Newborn infants prefer attractive faces. *Infant Behavior and Development* 21(2):345–354.

Swindale, N. V. 1996. The development of topography in the visual cortex: A review of models. *Network – Computation in Neural Systems* 7:161–247.