Learning to See:
Genetic and Environmental Influences
on Visual Development

James A. Bednar

Report AI-TR-02-294 May 2002

jbednar@cs.utexas.edu
http://www.cs.utexas.edu/users/nn/

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

The Dissertation Committee for James Albert Bednar
certifies that this is the approved version of the following dissertation:

Learning to See:
Genetic and Environmental Influences on Visual Development

Committee:

Risto Miikkulainen, Supervisor

Wilson S. Geisler

Raymond Mooney

Benjamin Kuipers

Joydeep Ghosh

Les Cohen

Learning to See:

Genetic and Environmental Influences on Visual Development

by

James Albert Bednar, B.S., B.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2002

Acknowledgments

This thesis would not have been possible without the support, advice, and encouragement of Risto Miikkulainen over the years, not to mention all his work on the draft revisions. If by some stroke of fate you have an opportunity to work with Risto, you should take it. I am also thankful for encouragement, constructive criticism, and career guidance from my incredibly wise committee members, Bill Geisler, Ray Mooney, Ben Kuipers, Joydeep Ghosh, and Les Cohen. Les Cohen in particular provided much-needed insight into the infant psychology literature. Bill Geisler has been very patient and thoughtful regarding my biological interests and with rough drafts of several theses and papers, and has provided valuable pointers to research in other fields.

I thank the members and all the hangers-on of the UTCS Neural Networks research group for invaluable feedback and many productive and entertaining discussions, especially Yoonsuck Choe, Lisa Kaczmarczyk, Marty and Coquis Mayberry, Lisa Redford, Amol Kelkar, Tal Tversky, Bobby Bryant, Tino Gomez, Adrian Agogino, Harold Chaput, Paul McQuesten, Jeff Provost, and Ken Stanley. Yoonsuck Choe in particular has long been a source of hard-hitting and constructive feedback, as well as an expert resource for operating systems and hardware maintenance, and a good friend. Lisa Kaczmarczyk provided very helpful comments on earlier paper drafts, and I have enjoyed having both of the Lisas as colleagues and friends. I have also benefitted from discussions and get-togethers with a diverse, talented, and personable set of visitors to our group, including Igor Farkas, Alex Lubberts, Nora Aguirre, Daniel Polani, Yaron Silbermann, Enrique Muro, and Alex Conradie.

I am very grateful for comments on research ideas and paper drafts from Mark H. Johnson, Harel Shouval, Francesca Acerra, Cara Cashon, Cornelius Weber, Dan Butts, and those who participated in the "Cortical Map Development" workshop at CNS*01. Joseph Sirosh provided the initial software code used in my work, and I am also grateful to Harel Shouval, Bernard Achermann, and Henry Rowley for making their face and natural scene databases available.

I am very fortunate to have a loving and supportive family, and I thank each of them for pretending to believe me each semester that I promised to finish the dissertation. In particular, my parents Eugene D. and Julia M. Bednar have been a constant source of encouragement, and my grandmothers Angelina Bednar and Julia D. Hueske have been a much-needed source of both moral and occasionally financial support. Throughout it all, the lovely and talented Tasca Shadix has kept

JAMES A. BEDNAR

The University of Texas at Austin
May 2002

Learning to See:

Genetic and Environmental Influences on Visual Development

Publication No. _____

James Albert Bednar, Ph.D.
The University of Texas at Austin, 2002

Supervisor: Risto Miikkulainen

How can a computing system as complex as the human visual system be specified and constructed? Recent discoveries of widespread spontaneous neural activity suggest a simple yet powerful explanation: genetic information may be expressed as internally generated training patterns for a general-purpose learning system. The thesis presents an implementation of this idea as a detailed, large-scale computational model of visual system development. Simulations show how newborn orientation processing and face detection can be specified in terms of training patterns, and how postnatal learning can extend these capabilities. The results explain experimental data from laboratory animals, human newborns, and older infants, and provide concrete predictions about infant behavior and neural activity for future experiments. They also suggest that combining a pattern generator with a learning algorithm is an efficient way to develop a complex adaptive system.

Contents

List of Figures

Chapter 1

Introduction

Current computing systems lag far behind humans and animals at many important information-processing tasks. One potential reason is that brains have far greater complexity (10^{15} synapses compared to e.g. fewer than 10^8 transistors; Alpert and Avnon 1993; Kandel, Schwartz, and Jessell 1991). It is unlikely that human engineers will be able to design a specific blueprint for a system with 10^{15} components. How does nature manage to do it? One clue is that the genome has fewer than 10^5 genes total, which means that any encoding scheme for the connections must be extremely compact (Lander et al. 2001; Venter et al. 2001). This thesis will examine how the human visual system might be constructed from such a compact specification by *input-driven self-organization*.

In such a system, only the largest-scale structure is specified directly. The details can then be determined by a learning algorithm driven by information in the environment. Along these lines, computational studies have shown that an initially uniform artificial neural network can develop structures like those found in the visual cortex, using simple learning rules driven by visual input (as reviewed by Erwin, Obermayer, and Schulten 1995 and Swindale 1996). However, such systems depend critically on the specific input patterns available. The system may not develop predictably if its environment is variable; what the learning algorithm discovers may not be the information most relevant to the system or organism. Such a system will also have very poor performance until learning is complete. Thus the potentially higher complexity available in a learning system comes with a cost: the system will take longer to develop and cannot be guaranteed to perform the desired task.

Recent experimental findings in neuroscience suggest that nature may have found a clever way around this tradeoff. Developing sensory systems are now known to be spontaneously active even before birth, i.e., before they could be learning from the environment (as reviewed by Wong 1999 and O'Donovan 1999). This spontaneous activity may actually guide the process of cortical development, acting as genetically specified training patterns for a learning algorithm (Constantine-Paton, Cline, and Debski 1990; Hirsch 1985; Jouvet 1998; Katz and Shatz 1996; Marks, Shaffery, Oksenberg, Speciale, and Roffwarg 1995; Roffwarg, Muzio, and Dement 1966; Shatz 1990, 1996).
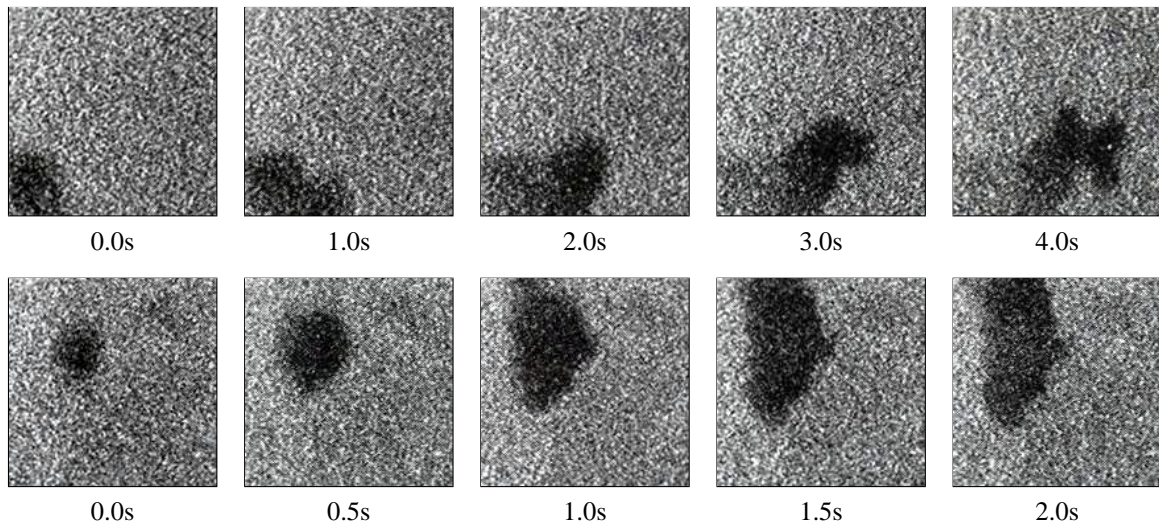
Figure 1.1: **Spontaneous waves in the ferret retina.** Each of the frames shows calcium concentration imaging of approximately 1 mm^2 of newborn ferret retina; the plots are a measure of how active the retinal cells are. Dark areas indicate increased activity. This activity is spontaneous (internally generated), because the photoreceptors have not yet developed at this time. From left to right, the frames on the top row form a 4-second sequence showing the start and expansion of a wave of activity. The bottom row shows a similar wave 30 seconds later. Later chapters will show that this type of correlated activity can explain how orientation selectivity develops before eye opening. Reprinted from Feller et al. (1996) with permission; copyright 1996, American Association for the Advancement of Science.

Figure 1.1 shows examples of spontaneous activity in the retina of a newborn ferret.

For a biological species, being able to control the training patterns can guarantee that each organism has a rudimentary level of performance from the start. Such training would also ensure that initial development does not depend on the details of the external environment. In contrast, a specific, fixed genetic blueprint could also guarantee a good starting level of performance, but performance would remain limited to that level. Thus internally generated patterns can preserve the benefits of a blueprint, within a learning system capable of much higher system complexity and performance.

Inspired by the discoveries of widespread spontaneous activity, this dissertation will test the hypothesis that a functioning sensory system can be constructed from a specification of:

1. a rough initial structure,

2. internal training pattern generators, and

3. a self-organizing algorithm.

Internal patterns drive initial development, and the external environment completes the process. The result is a compact specification of a complex, high-performance product. Using pattern generation

to guide development appears to be ubiquitous in nature, and may represent a general-purpose technique for building complex artificial systems.

1.1 Approach

The pattern generation hypothesis will be evaluated by building and testing HLISSOM, a computational model of visual system development. The visual system is the best-studied sensory system in mammals, and thus it offers the most comprehensive data to constrain and validate models. The goal of the modeling is to understand how the visual cortex is constructed, in the hope that this understanding will be useful for designing future complex information processing systems.

The simulations focus on two visual capabilities where both environmental and genetic influences appear to play a strong role: orientation processing and face detection. At birth, newborns can already discriminate between two orientations (Slater and Johnson 1998; Slater, Morison, and Somers 1988), and animals have neurons and brain regions selective for particular orientations even before their eyes open (Chapman and Stryker 1993; Crair, Gillespie, and Stryker 1998; Gödecke, Kim, Bonhoeffer, and Singer 1997). Yet orientation processing circuitry in these same areas can also be strongly affected by visual experience (Blakemore and van Sluyters 1975; Sengpiel, Stawinski, and Bonhoeffer 1999). Similarly, newborns already prefer face-like patterns soon after birth, but face processing ability takes months or years of experience to develop fully (Goren, Sarty, and Wu 1975; Johnson and Morton 1991; reviewed in de Haan 2001).

Because the orientation processing circuitry is simpler and has been mapped out in much greater detail, it will be used as a well-studied test case for the pattern generation approach. The same techniques will then be applied to face processing, in order to generate testable predictions to drive future experiments in a more complex system. The specific aims are to understand how internal activity can account for the structure present at birth in each system, and how postnatal experience can complete this developmental process. For each system, I will validate the model by comparing it to existing experimental results, and then use it to derive predictions for future experiments that will further reveal how the visual system is constructed.

1.2 Outline of the dissertation

This dissertation is organized into four main parts: background (chapters 1–3), model and methods (chapters 4–5), results (chapters 6–8), and discussion (chapters 9–10).

Chapter 2 is a survey of the experimental evidence from animal and human visual systems that forms the basis for the HLISSOM model. The chapter first describes the adult visual system, then summarizes what is known about its development, and what remains controversial.

Chapter 3 surveys previous computational and theoretical approaches to understanding the development of orientation and face processing.

Chapter 4 introduces the HLISSOM model architecture, specifies its operation mathematically, and describes the procedures for running HLISSOM simulations. As a detailed example, it gives results from a simple orientation simulation that nonetheless is a good match to a range of experimental data.

Chapter 5 introduces a set of scaling equations for topographic map simulations, and shows that they can generate similar orientation-processing circuitry in brain regions of different sizes. The equations allow each simulation to trade off computational requirements against simulation accuracy, and allow very large networks to be simulated when needed. This capability will be crucial for the experiments in chapters 6–8.

Chapter 6 shows that together internally generated and visually evoked activation can explain how orientation preferences develop prenatally and postnatally. The resulting orientation processing circuitry is a good match to experimental findings in newborn and adult animals. These orientation simulations also provide a foundation for the face processing experiments in chapter 7.

Chapter 7 presents results from a combined model of newborn orientation processing and face preferences. When trained on proposed types of internally generated activity, the model replicates the face preferences found in studies of human infants, and provides a concrete explanation for how those face preferences occur.

Chapter 8 presents simulations of postnatal experience with real faces, objects, and visual scenes. Visual experience gradually drives the coarse neural representations at birth to become better tuned to real faces. The postnatal learning process replicates several surprising findings from studies of newborn face learning, such as a decrease in response to schematic drawings of faces, and provides novel predictions that can drive future experiments.

Chapter 9 discusses implications of the results presented here, and proposes future experimental, computational, and engineering studies based on this approach.

Chapter 10 summarizes and evaluates the contributions of the thesis.

Appendix A lists the parameter values used in each simulation. It also discusses how those parameters were set, and how to set them to perform other types of simulations.

Chapter 2

Background

This thesis presents computational simulations of how the human visual system develops. So that the simulations are a meaningful tool for understanding natural systems, they are based on detailed anatomical, neurophysiological, and psychological evidence from animals and human infants. In this chapter I will review this evidence for adult humans and for animals that have similar visual systems, focusing on the orientation and face-processing capabilities that will be modeled in later chapters. I will then summarize what is known about the state of these systems at birth and their prenatal and postnatal development, as well as what remains unclear. Throughout, I will emphasize the important role that neural activity plays in this development, and that this activity can be either visually evoked or internally generated.

2.1   The adult visual system

The adult visual system has been studied experimentally in a number of mammalian species, including human, monkey, cat, ferret, and tree shrew. For a variety of reasons, many of the important results have been measured in only one or a subset of these species, but they are expected to apply to the others as well. This thesis focuses on the human visual system, but also relies on data from these animals where human data is not available.

Figure 2.1 shows a diagram of the main feedforward pathways in the human visual system (see e.g. Wandell 1995, Daw 1995, or Kandel et al. 1991 for an overview). Other mammalian species have a similar organization. During visual perception, light entering the eye is detected by the *retina*, an array of photoreceptors and related cells on the inside of the rear surface of the eye. The cells in the retina encode the light levels at a given location as patterns of electrical activity in neurons called *retinal ganglion cells*. This activity is called *visually evoked activity*. Retinal ganglion cells are densest in a central region called the *fovea*, corresponding to the center of gaze; they are much less dense in the *periphery*. Output from the ganglion cells travels through neural connections to the *lateral geniculate nucleus* (LGN) of the thalamus, at the base of each side of the brain. From the
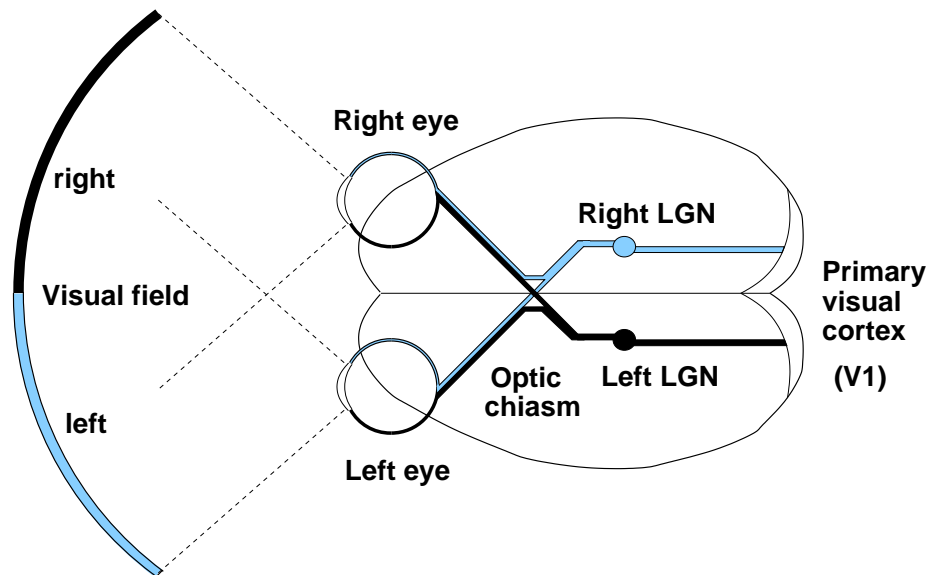
Figure 2.1: **Human visual sensory pathways (*top view*).** Visual information travels in separate pathways for each half of the visual field. For example, light entering the eye from the right hemifield reaches the left half of the retina, on the rear surface of each eye. The right hemifield inputs from each eye join at the optic chiasm, and travel to the left lateral geniculate nucleus (LGN) of the thalamus, then to the primary visual cortex (V1) of the left hemisphere. Signals from each eye are kept segregated into different neural layers in the LGN, and are combined in V1. There are also smaller pathways from the optic chiasm and LGN to other subcortical structures, such as the superior colliculus (not shown). For simplicity, the model in this thesis will focus on the pathway from a single eye to the LGN and visual cortex, although it can also be expanded to include both eyes (Miikkulainen et al. 1997; Sirosh 1995).

LGN, the signals continue to the *primary visual cortex* (V1; also called *striate* cortex or area 17) at the rear of the brain. V1 is the first *cortical* site of visual processing; the previous areas are termed *subcortical*. The output from V1 goes on to many different higher cortical areas, including areas that appear to underlie object and face processing (as reviewed by Merigan and Maunsell 1993; Van Essen, Anderson, and Felleman 1992). Much smaller pathways also go from the optic nerve and LGN to subcortical structures such as the *superior colliculus* and *pulvinar*, but these areas are not thought to be involved in orientation-specific processing (see e.g. Van Essen et al. 1992).

2.1.1 Early visual processing

At the photoreceptor level, the representation of the visual field is much like an image, but significant processing of this information occurs in the subsequent subcortical and early cortical stages (reviewed by e.g. Daw 1995; Kandel et al. 1991). First, retinal ganglion cells perform a type of edge detection on the input, responding most strongly to borders between bright and dark areas. Figure 2.2*a–b* illustrates the two typical response patterns of these neurons, ON-center and OFF-

(a) ON cell　　　　(b) OFF cell　　　　(c) Two-lobe V1　　　　(d) Three-lobe V1
　　　　　　　　　　　　　　　　　　　　　　　simple cell　　　　　　simple cell
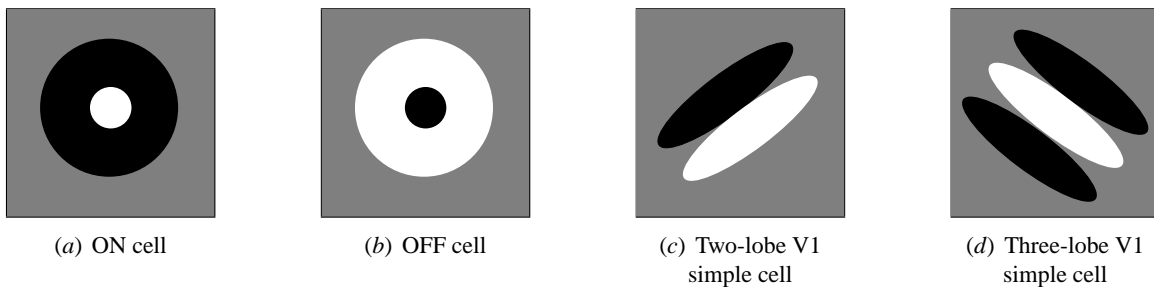
Figure 2.2: **Receptive field (RF) types in retina, LGN and V1.** Each diagram shows an RF on the retina for one neuron. Areas of the retina where light spots excite this neuron are plotted in white (ON areas), areas where dark spots excite it are plotted in black (OFF areas), and areas with little effect are plotted in gray. All RFs are spatially localized, i.e. have ON and OFF areas only in a small portion of the retina. (*a*) ON cells are found in the retina and LGN, and prefer light areas surrounded by darker areas. (*b*) OFF cells have the opposite preferences, responding most strongly to a dark area surrounded by light areas. RFs for both ON and OFF cells are isotropic, i.e. have no preferred orientation. Starting in V1, most cells in primates have orientation-selective RFs instead. The V1 RFs can be classed into a few basic types, of which the most common are shown here. Figure (*c*) shows a two-lobe arrangement, favoring a 45° edge with dark in the upper left and light in the lower right. Figure (*d*) shows one with three lobes, favoring a 135° white line against a darker background. RFs of all orientations are found in V1, but those representing the cardinal axes (horizontal and vertical) are more common. Adapted from Hubel and Wiesel (1968); Jones and Palmer (1987). Chapter 4 will introduce a model for the ON and OFF cells, and will show how simple cells like those in (*c*–*d*) can develop.

center. An ON-center retinal ganglion cell responds most strongly to a spot of light located in a certain region of the photoreceptors, called its *receptive field* (RF). An OFF-center ganglion instead prefers a dark area surrounded by light. Neurons in the LGN have properties similar to retinal ganglion cells, and are also arranged retinotopically, so that nearby LGN cells respond to nearby portions of the retina. The ON-center cells in the retina connect to the ON cells in the LGN, and the OFF cells in the retina connect to the OFF cells in the LGN. Because of this independence, the ON and OFF cells are often described as separate processing *channels*, the ON channel and the OFF channel.

Like LGN neurons, nearby neurons in V1 also respond to nearby portions of the retina. However, they prefer edges and lines of a particular range of orientations, and do not respond to unoriented stimuli or orientations far from their preferred orientation (Hubel and Wiesel 1962, 1968). Because V1 neurons are the first to have significant orientation preferences, theories of orientation processing focus on areas V1 and above. See figure 2.2*c*–*d* for examples of typical receptive fields of V1 neurons. The neurons illustrated are what is known as *simple* cells, i.e. neurons whose ON and OFF patches are located at specific areas of the retinal field. Other neurons (*complex* cells) respond to the same configuration of light and dark over a range of positions (Hubel and Wiesel 1968). HLISSOM models the simple cells only, which are thought to be the first in V1 to show orientation selectivity (Hubel and Wiesel 1968).

V1, like the other parts of the cortex, is composed of a two-dimensional, slightly folded sheet of neurons and other cells. If flattened, human V1 would cover an area of nearly four square inches (Wandell 1995). It contains at least 150 million neurons, each making hundreds or thousands of specific connections with other neurons in the cortex and in subcortical areas like the LGN (Wandell 1995). The neurons are arranged in six layers with different anatomical characteristics (using Brodmann's scheme for numbering laminations in human V1, as described by Henry 1989). Input from the thalamus goes through *afferent* connections to V1, typically terminating in layer 4 (Casagrande and Norton 1989; Henry 1989). Neurons in the other layers form local connections within V1 or connect to higher visual processing areas. For instance, many neurons in layers 2 and 3 have long-range *lateral connections* to the surrounding neurons in V1 (Gilbert, Hirsch, and Wiesel 1990; Gilbert and Wiesel 1983; Hirsch and Gilbert 1991). There are also extensive feedback connections from higher areas (Van Essen et al. 1992).

At a given location on the cortical sheet, the neurons in a vertical section through the cortex generally respond most strongly to the same eye of origin, stimulus orientation, stimulus size, etc. It is customary to refer to such a section as a *column* (Gilbert and Wiesel 1989). The HLISSOM model discussed in this thesis will treat each column as a single unit, thus representing the cortex as a purely two-dimensional surface. This model is only an approximation, but it is a valuable one because it greatly simplifies the analysis while retaining the basic functional features of the cortex.

Nearby columns generally have similar, but not identical, preferences; slightly more distant columns generally have more dissimilar preferences. Preferences repeat at regular intervals (approximately 1–2 mm) in every direction, which ensures that each type of preference is represented for every location on the retina. For orientation preferences, this arrangement of neurons forms a smoothly varying *orientation map* of the retinal input (Blasdel 1992a; Blasdel and Salama 1986; Grinvald, Lieke, Frostig, and Hildesheim 1994; Ts'o, Frostig, Lieke, and Grinvald 1990). See figure 2.3 for an explanation of how the orientation map can be measured, and figure 2.4 for an example orientation map from monkey cortex. Each location on the retina is mapped to a region on the orientation map, with each possible orientation at that retinal location represented by different but nearby orientation-selective cells. Other mammalian species have largely similar orientation maps, although there are differences in some of the details (Müller, Stetter, Hubener, Sengpiel, Bonhoeffer, Gödecke, Chapman, Löwel, and Obermayer 2000; Rao, Toth, and Sur 1997). Maps of preferences for other stimulus features are also present, including direction of motion, spatial frequency, and ocular dominance (left or right eye preference; Issa, Trepel, and Stryker 2001; Obermayer and Blasdel 1993; Shatz and Stryker 1978; Shmuel and Grinvald 1996; Weliky, Bosking, and Fitzpatrick 1996).

Within V1, the lateral connections correlate with stimulus preferences, particularly for orientation. For instance, the long-range lateral connections of a given neuron target neurons in other patches that have similar orientation preferences, aligned along the preferred orientation of the neuron (Bosking, Zhang, Schofield, and Fitzpatrick 1997; Schmidt, Kim, Singer, Bonhoeffer, and Lowel 1997; Sincich and Blasdel 2001; Weliky et al. 1995). Figure 2.5 shows examples of these
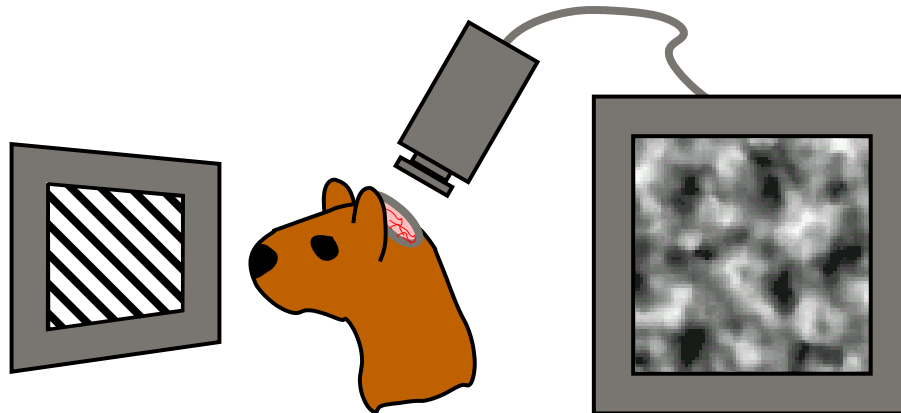
Figure 2.3: **Measuring orientation maps.** Optical imaging techniques allow orientation preferences to be measured for large numbers of neurons at once (Blasdel and Salama 1986). In such experiments, part of the skull of a laboratory animal is removed by surgery, exposing the surface of the visual cortex. Visual patterns are then presented to the eyes, and a video camera records either light absorbed by the cortex or light given off by fluorescent chemicals that have been applied to it. Both methods allow the two-dimensional patterns of neural activity to be measured, albeit indirectly. Measurements can then be compared between different stimulus conditions, e.g. different orientations, determining which stimulus is most effective at activating each small patch of neurons. Figure 2.4 and later figures in this chapter will show maps of orientation preference computed using these techniques. Adapted from Weliky et al. (1995).

connections. Anatomically, individual long-range connections are usually excitatory, but for high-contrast inputs their net effects are inhibitory due to contacts on local inhibitory neurons (Hirsch and Gilbert 1991; Weliky et al. 1995; Hata, Tsumoto, Sato, Hagihara, and Tamura 1993; Grinvald et al. 1994; see discussion in Bednar 1997). Thus for modeling purposes the long-range connections are usually treated as inhibitory, as they will be in HLISSOM.

Lateral connections are thought to underlie a variety of psychophysical phenomena, including contour integration and the effects of context on visual perception (Bednar and Miikkulainen 2000b; Choe 2001; Choe and Miikkulainen 1998; Gilbert 1998; Gilbert et al. 1990). For computational efficiency, most models treat the lateral connections as a simple isotropic function, but orientation-specific connections are important for several theories of orientation map development (as described in chapter 3). For this reason, HLISSOM will simulate the development of the patchy pattern of lateral connectivity.

2.1.2 Face and object processing

Beyond V1 in primates are dozens of less-understood *extrastriate* visual areas that can be arranged into a rough hierarchy (Van Essen et al. 1992). The relative locations of the areas in this hierarchy are largely consistent across individuals of the same species. Non-primate species have fewer higher areas, and in at least one mammal (the least shrew, a tiny rodent-like creature) V1 is the only visual

9

(*a*) Orientation map　　　　　　　(*b*) Orientation selectivity
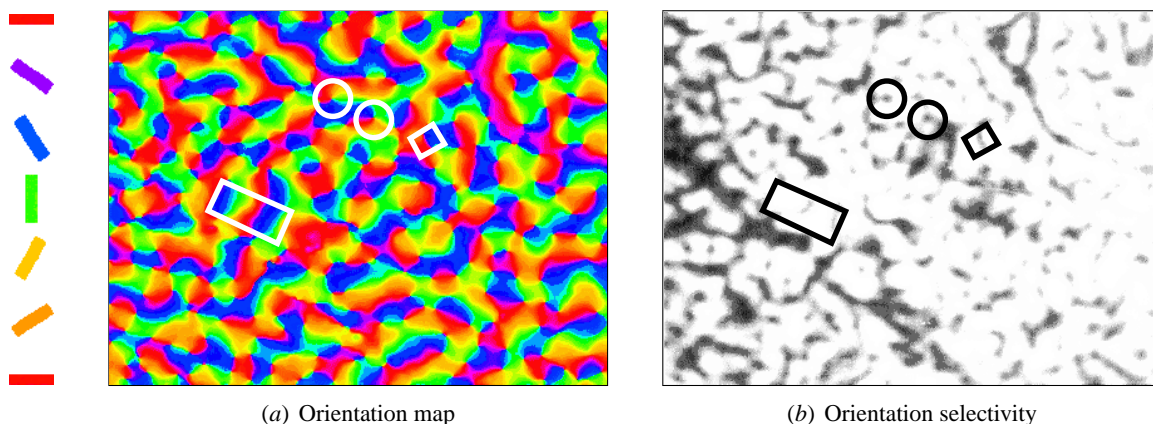
Figure 2.4: **Adult monkey orientation map (*color figure*).** Figures (*a*) and (*b*) show the preferred orientation and orientation selectivity of each neuron in a 7.5×5.5mm area of adult macaque monkey V1, measured by optical imaging techniques (reprinted with permission from Blasdel 1992b, copyright 1992 by the Society for Neuroscience; annotations added.) Each neuron in (*a*) is colored according to the orientation it prefers, using the color key at the left. Nearby neurons in the map generally prefer similar orientations, forming groups of the same color called *iso-orientation blobs*. Other qualitative features are also found: *Pinwheels* are points around which orientation preference changes continuously; a pair of pinwheels is circled in white. *Linear zones* are straight lines along which the orientations change continuously, like a rainbow; a linear zone is marked with a long white rectangle. *Fractures* are sharp transitions from one orientation to a very different one; a fracture between red and blue (without purple in between) is marked with a white square. As shown in (*b*), pinwheel centers and fractures tend to have lower selectivity (dark areas) in the optical imaging response, while linear zones tend to have high selectivity (light areas). Chapters 4 and 6 will model the development of similar orientation and selectivity maps.

area (Catania, Lyon, Mock, and Kaas 1999). Although the higher levels have not been studied as thoroughly as V1, the basic circuitry within each region is thought to be largely similar to V1. Even so, the functional properties differ, in part due to differences in connectivity with other regions (Kandel et al. 1991). For instance, neurons in higher areas tend to have larger retinal receptive fields, respond to stimuli at a greater range of positions, and process more complex visual features (Ghose and Ts'o 1997; Haxby, Horwitz, Ungerleider, Maisog, Pietrini, and Grady 1994; Rolls 2000). In particular, extrastriate cortical regions that respond preferentially to faces have been found in both adult monkeys (using single-neuron studies; Gross, Rocha-Miranda, and Bender 1972; Rolls 1992) and adult humans (using imaging techniques like fMRI; Halgren, Dale, Sereno, Tootell, Marinkovic, and Rosen 1999; Kanwisher, McDermott, and Chun 1997; Puce, Allison, Gore, and McCarthy 1995).[1]

These face-selective areas receive visual input via the V1 orientation map. They appear

---

[1] I will use the term *face selective* to refer to any cell or region that shows a higher response to faces than to other similar stimuli. Some studies also show more specific types of face selectivity, such as face recognition or face detection.

(a) Single orientation   (b) Orientation map
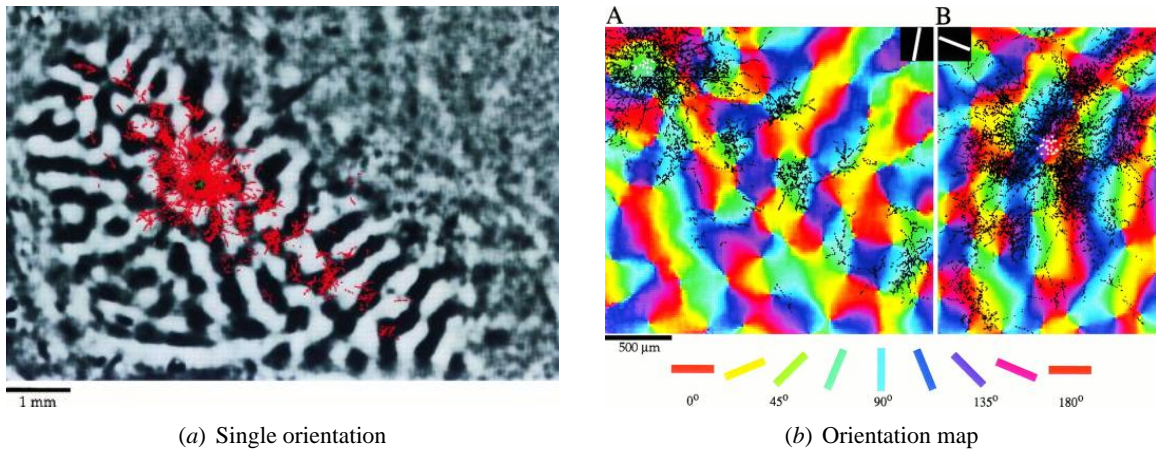
Figure 2.5: **Lateral connections in the tree shrew align with the orientation map (*color figure*).** Figure (*a*) shows the orientation preferences of a section of adult tree shrew V1 measured using optical imaging. In the figure, vertical in the visual field (90°) corresponds to a diagonal line pointing towards 10 o'clock (135°). Areas responding to vertical stimuli are plotted in black, and horizontal in white. Overlaid on the map is a small green dot marking the site where a patch of nearby vertical-selective neurons were injected with a tracer chemical. In red are plotted neurons to which that chemical propagated through lateral connections. Short-range lateral connections target all orientations equally, but long-range connections target neurons that have similar orientation preferences and are extended along the orientation preference of this neuron. Image A in figure (*b*) shows a detailed view of the information in figure (*a*) plotted on the full orientation map. The injected neurons are colored greenish cyan (80°), and connect to other neurons with similar preferences. Image B in figure (*b*) shows similar results from a different location. These neurons prefer reddish purple (160°), and more densely connect to other red or purple neurons. Measurements in monkeys show similar patchiness, but in monkey the connections do not usually extend as far along the orientation axis of the neuron (Sincich and Blasdel 2001). These results, theoretical analysis, and computational models suggest that the lateral connections play a significant role in orientation processing (Bednar and Miikkulainen 2000b; Gilbert 1998; Sirosh 1995). Chapter 4 will show how these lateral connection patterns can develop. Reprinted from Bosking et al. (1997) with permission; copyright 1997 by the Society for Neuroscience.

to be loosely segregated into different regions that process faces in different ways. For instance, some areas appear to perform face detection, i.e. respond unspecifically to many face-like stimuli (de Gelder and Rouw 2000, 2001). Others selectively respond to facial expressions, gaze directions, or prefer specific faces (i.e., perform face recognition; Perrett 1992; Rolls 1992; Sergent 1989). Whether these regions are exclusively devoted to face processing, or also process other common objects, remains controversial (Haxby, Gobbini, Furey, Ishai, Schouten, and Pietrini 2001; Kanwisher 2000; Tarr and Gauthier 2000). HLISSOM will model areas involved in face detection (and not face recognition or other types of face processing), but does not assume that the areas modeled will process faces exclusively.

2.2 Development of early visual processing

Despite the progress made in understanding the structure and function of the adult visual system, much less is known about how this circuitry is constructed. Two extreme alternative theories state that: (a) the visual system develops through general-purpose learning of patterns seen in the environment, or (b) the visual system is constructed from a specific blueprint encoded somehow in the genome. The conflict between these two positions is generally known as the Nature–Nurture debate, which has been raging for centuries in various forms (Diamond 1974).

The idea of a specific blueprint does seem to apply to the largest scale organization of the visual system, at the level of areas and their interconnections. These patterns are largely similar across individuals of the same species, and their development does not generally depend on neural activity, visually evoked or otherwise (Miyashita-Lin, Hevner, Wassarman, Martinez, and Rubenstein 1999; Rakic 1988; Shatz 1996). But at smaller scales, such as orientation maps and lateral connections within them, there is considerable evidence for both environmental and internally controlled development. Thus debates center on how this seemingly conflicting evidence can be reconciled. In the subsections below I will summarize the evidence for environmental and genetic influences, focusing first on orientation processing (for which the evidence of each type is substantial), and then on the less well-studied topic of face processing.

2.2.1 Environmental influences on early visual processing

Experiments since the 1960s have shown that the environment can have a large effect on the structure and function of the early visual areas (as reviewed by Movshon and van Sluyters 1981). For instance, Blakemore and Cooper (1970) found that if kittens are raised in environments consisting of only vertical contours during a critical period, most of their V1 neurons become responsive to vertical orientations. Similarly, orientation maps from kittens with such rearing devote a larger area to the orientation that was overrepresented during development (Sengpiel et al. 1999). Even in normal adult animals, the distribution of orientation preferences is slightly biased towards horizontal and vertical contours (Chapman and Bonhoeffer 1998; Coppola, White, Fitzpatrick, and Purves 1998). Such a bias would be expected if the neurons learned orientation selectivity from typical environments, which have a similar orientation bias (Switkes, Mayer, and Sloan 1978). Conversely, kittens who were raised without patterned visual experience at all, e.g. by suturing their eyelids shut, have few orientation-selective neurons in V1 as an adult (Blakemore and van Sluyters 1975; Crair et al. 1998). Thus visual experience can clearly influence how orientation selectivity and orientation maps develop.

The lateral connectivity patterns within the map are also clearly affected by visual experience. For instance, kittens raised without patterned visual experience in one eye (by monocular lid suture) develop non-specific lateral interactions for that eye (Kasamatsu, Kitano, Sutter, and Norcia 1998). Conversely, lateral connections become patchier when inputs from each eye are decorrelated

during development (by artificially inducing strabismus, i.e. squint; Gilbert et al. 1990; Löwel and Singer 1992).

Finally, in ferrets it is possible to reroute the connections from the eye that normally go to V1 via the LGN, so that instead they reach auditory cortex (as reviewed in Sur, Angelucci, and Sharma 1999; Sur and Leamey 2001). The result is that the auditory cortex develops orientation-selective neurons, orientation maps, and patchy lateral connections, although the orientation maps do show some differences from normal maps. Furthermore, the ferret can use the rewired neurons to make visual distinctions, such as to discriminate between two grating stimuli (von Melchner, Pallas, and Sur 2000). Thus the structure and function of the cortex can be profoundly affected by its inputs. Together with the results from altered environments, this evidence suggests that the structure and function of V1 could simply be learned from experience with oriented contours in the environment. More specifically, the neural activity patterns in the LGN might be sufficient to direct the development of V1 and other cortical areas.

2.2.2   Genetic influences on early visual processing

Yet despite the clear environmental effects on orientation processing and the role of visual activity, individual orientation-selective cells have long been known to exist in newborn kittens and ferrets even before they open their eyes (Blakemore and van Sluyters 1975; Chapman and Stryker 1993).[2]. Recent advances in experimental imaging technologies have even allowed the full map of orientation preferences to be measured in young animals. Such experiments show that large-scale orientation maps exist prior to visual experience, and that these maps have many of the same features found in adults (Chapman, Stryker, and Bonhoeffer 1996; Crair et al. 1998; Gödecke et al. 1997). Figure 2.6 shows an example of such a map from a kitten. The lateral connections within the orientation map are also already patchy before eye opening (Gödecke et al. 1997; Luhmann, Martínez Millán, and Singer 1986; Ruthazer and Stryker 1996).

Furthermore, the global patterns of blobs in the maps appear to change very little with normal visual experience, even as the individual neurons gradually become more selective for orientation, and lateral connections become more selective (Chapman and Stryker 1993; Crair et al. 1998; Gödecke et al. 1997). Although the actual map patterns have so far only been measured in animals, not newborn or adult humans, psychological studies suggest that human newborns can already discriminate between patterns based on orientation (Slater and Johnson 1998; Slater et al. 1988). Thus despite the clear influence of long-term rearing in abnormal conditions, normal visual experience appears primarily to preserve and fine-tune the existing structure of the V1 orientation map, rather than drive its development.

---

[2]Note that although human newborns open their eyes soon after birth, visual experience begins much later in other species. For instance, ferrets and cats open their eyes only days or weeks after birth, which makes them convenient for developmental studies. This section focuses on eye opening as the start of visual experience, because it discusses animal experiments. Elsewhere this thesis will simply use "prenatal" and "postnatal" to refer to phases before and after visual experience, because the primary focus is on human infants.
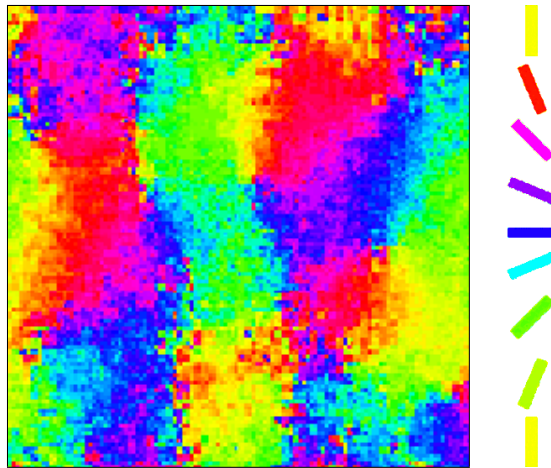
Figure 2.6: **Neonatal orientation maps (*color figure*).** A 1.9×1.9mm section of an orientation map from a two-week-old binocularly deprived kitten, i.e. a kitten without prior visual experience. The map is not as smooth as in the adult, and many of the neurons are not as selective (not shown), but the map already has iso-orientation blobs, pinwheel centers, fractures, and linear zones. Reprinted with permission from Crair et al. (1998), copyright 1998, American Association for the Advancement of Science.

2.2.3   Internally generated activity

The preceding sections show that initial cortical development does not require visual experience, yet postnatal visual experience can change how the cortex develops. Moreover, for orientation processing it is clear that the regions that are already organized at birth are precisely the same regions that are later affected by visual experience. (For face processing it is not yet known whether the underlying circuitry is the same, but the behavioral effects are similar to those for the well-studied case of orientation processing.) Thus one important question is, how could the same circuitry be both genetically hardwired, yet also capable of learning from the start? New experiments are finally starting to shed light on this longstanding mystery: many of the structures present at birth could result from learning of spontaneous, internally generated neural activity. The same activity-dependent learning mechanisms that can explain postnatal learning may simply be functioning before birth, driven by activity from internal instead of external sources. Thus the "hardwiring" may actually be learned.

Spontaneous neural activity has recently been discovered 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, Meister, and Shatz 1993; Yuste, Nelson, Rubin, and Katz 1995; reviewed by O'Donovan 1999; Wong 1999). Figure 1.1 on page 2 showed an example of one type of spontaneous activity, retinal waves. In several cases, experiments have shown that interfering with spontaneous activity can change the outcome of development. For instance, when the retinal waves are abolished, the LGN fails to develop normally (e.g., inputs from the two eyes are no longer segregated; Chapman 2000; Shatz 1990, 1996; Stellwagen and Shatz 2002). Similarly,

14

when activity is silenced at the V1 level during early development, neurons in mature animals have much lower orientation selectivity (Chapman and Stryker 1993).

The debate now centers on whether such spontaneous activity is merely *permissive* for development, perhaps by keeping newly formed connections alive until visual input occurs, or whether it is *instructive*, i.e. whether patterns of activity specifically determine how the structures develop (as reviewed by Chapman, Gödecke, and Bonhoeffer 1999; Crair 1999; Katz and Shatz 1996; Miller, Erwin, and Kayser 1999; Penn and Shatz 1999; Sur et al. 1999; Sur and Leamey 2001; Thompson 1997). Several recent experiments have shown that spontaneous activity can clearly be instructive. For instance, Weliky and Katz (1997) artificially activated a large number of axons in the optic nerve of ferrets, thereby disrupting the pattern of spontaneous activity. Even though this manipulation increased the total amount of activity, leaving any permissive aspects of activity unchanged, the result was a reduction in orientation selectivity in V1. Thus spontaneous activity cannot simply be permissive. Similarly, pharmacologically increasing the number of retinal waves in one eye has very recently been shown to disrupt LGN development (Stellwagen and Shatz 2002; but see Crowley and Katz 2000). Yet increasing the waves in *both* eyes restores normal development, which again shows that it is not simply the presence of the activity that is important (Stellwagen and Shatz 2002). However, it is not yet known what specific features of the internally generated activity are instructive for development in each region, because it has not yet been possible to manipulate the activity precisely.

Retinal waves are the most well-studied source of spontaneous activity, because they are easily accessible to experimenters. However, other internally generated patterns also appear to be important for visual cortex development. One example is the ponto-geniculo-occipital (PGO) waves that are the hallmark of rapid-eye-movement (REM) sleep (figure 2.7). During and just before REM sleep, PGO waves originate in the brainstem and travel to the LGN, visual cortex, and a variety of subcortical areas (see Callaway, Lydic, Baghdoyan, and Hobson 1987 for a review).

In adults, 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). Studies also suggest that PGO wave activity is under genetic control: PGO waves elicit different activity patterns in different species (Datta 1997), and the eye movement patterns that are associated with PGO waves are more similar in identical twins than in unrelated age-matched subjects (Chouvet, Blois, Debilly, and Jouvet 1983). Thus PGO waves are a good candidate for genetically controlled visual system training patterns.

In 1966, Roffwarg et al. proposed that REM sleep must be important for development. Their reasoning was that (1) developing mammalian embryos spend a large percentage of their time in states that look much like adult REM sleep, and (2) the duration of REM sleep is strongly correlated with the degree of neural plasticity, both during development and between species (also see the more recent review by Siegel 1999, as well as Jouvet 1980). also Consistent with Roffwarg et al.'s hypothesis, it has recently been found that blocking REM sleep and/or the PGO waves

15
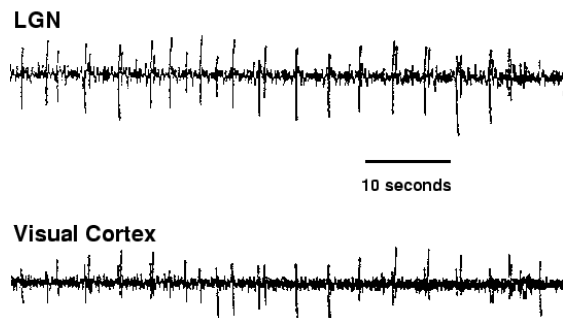
Figure 2.7: **PGO waves.** Each line shows an electrical recording from a cell in the indicated area during REM sleep in the cat. Spontaneous REM sleep activation in the pons of the brain stem is relayed to the LGN of the thalamus (top), to the primary visual cortex (bottom), and to many other regions in the cortex. It is not yet known what spatial patterns of visual cortex activation are associated with this temporal activity, or with other types of spontaneous activity during sleep. Reprinted from *Behavioural Brain Research*, 69, Marks et al., "A functional role for REM sleep in brain maturation", 1–11, copyright 1995, with permission from Elsevier Science.

alone *heightens* the effect of visual experience during development (Marks et al. 1995; Oksenberg, Shaffery, Marks, Speciale, Mihailoff, and Roffwarg 1996; Pompeiano, Pompeiano, and Corvaja 1995). In kittens with normal REM sleep, when the visual input to one eye of a kitten is blocked for a short time during a critical period, the cortical and LGN area devoted to signals from the other eye increases (Blakemore and van Sluyters 1975). When REM sleep (or just the PGO waves) is interrupted as well, the effect of blocking one eye's visual input is even stronger (Marks et al. 1995). This result suggests REM sleep, and PGO waves in particular, ordinarily limits or counteracts the effects of visual experience.

All of these characteristics suggest that PGO waves and other REM-sleep activity may be instructing development, like the retinal waves do (Jouvet 1980, 1998; Marks et al. 1995). However, due to limitations in experimental imaging equipment and techniques, it has not yet been possible to measure the two-dimensional spatial shape of the activity associated with the PGO waves (Rector, Poe, Redgrave, and Harper 1997). This thesis will evaluate different candidates for internally generated activity, including retinal and PGO waves, and show how this activity can explain how maps and their connections develop in the visual cortex.

2.3 Development of face detection

In previous sections I have focused on the early visual processing pathways, up to V1, because recent anatomical and physiological evidence from cats and ferrets has begun to clarify how those areas develop. These detailed studies have been made possible by the fact that much of the cat and ferret visual system develops postnatally but before the eyes open. However, face selective neurons

16

or regions have not yet been documented in cats or ferrets, either adult or newborn. Thus studies of the neural basis of face selectivity focus on primates.

The youngest primates that have been tested are six week old monkeys, which do have face selective neurons (Rodman 1994; Rodman, Skelly, and Gross 1991). Six weeks is a significant amount of visual experience, and it has not yet been possible to measure neurons or regions in younger monkeys. Thus it is unknown whether the cortical regions that are face-selective in adult primates are also face-selective in newborns, or whether they are even fully functional at birth (Bronson 1974; Rodman 1994). As a result, how these regions develop remains highly controversial (for review see de Haan 2001; Gauthier and Nelson 2001; Nachson 1995; Slater and Kirby 1998; Tovée 1998).

Although measurements at the neuron or region level are not available, behavioral tests with human infants suggest that the postnatal development of face detection is similar to the well-studied case of orientation map development. In particular, internal, genetically determined factors also appear to be important for face detection. The main evidence for a genetic basis is a series of studies showing that human newborns turn their eyes or head towards facelike stimuli in the visual periphery, longer or more often than they do so for other stimuli (Goren et al. 1975; Johnson, Dziurawiec, Ellis, and Morton 1991; Johnson and Morton 1991; Mondloch, Lewis, Budreau, Maurer, Dannemiller, Stephens, and Kleiner-Gathercoal 1999; Simion, Valenza, Umiltà, and Dalla Barba 1998b; Valenza, Simion, Cassia, and Umiltà 1996). These effects have been found within minutes or hours after birth. Figure 2.8 shows how several of these studies have measured the face preferences, and figure 2.9 shows a typical set of results. Whether these preferences represent genuine preference for faces has been very controversial, in part because of the difficulties in measuring pattern preferences in newborns (Easterbrook, Kisilevsky, Hains, and Muir 1999; Hershenson, Kessen, and Munsinger 1967; Kleiner 1993, 1987; Maurer and Barrera 1981; Simion, Macchi Cassia, Turati, and Valenza 2001; Slater 1993; Thomas 1965). Newborn preferences for additional patterns will be shown in chapter 7, which also shows that HLISSOM exhibits similar face preferences when trained on internally generated patterns.

Early postnatal visual experience also clearly affects face preferences, as for orientation map development. For instance, an infant only a few days old will prefer to look at its mother's face, relative to the face of a female stranger with "similar hair coloring and length" (Bushnell 2001) or "broadly similar in terms of complexion, hair color, and general hair style" (Pascalis, de Schonen, Morton, Deruelle, and Fabre-Grenet 1995). A significant mother preference is found even when non-visual cues such as smell and touch are controlled (Bushnell 2001; Bushnell, Sai, and Mullin 1989; Field, Cohen, Garcia, and Greenberg 1984; Pascalis et al. 1995). The mother preference presumably results from postnatal learning of the mother's appearance. Indeed, Bushnell (2001) found that newborns look at their mother's face for an average of 23% of their time awake over the first few days, which provides ample time for learning.

Pascalis et al. (1995) found that the mother preference disappears when the external outline

Figure 2.8: **Measuring newborn face preferences.** Newborn face preferences have been measured by presenting schematic stimuli to human infants within a few minutes or hours of birth, and measuring how far the babies' eyes or head track the stimulus. The experimenter is blind to the specific pattern shown, and an observer also blind to it measures the baby's responses. Face preferences have been found even when the experimenter's face and all other faces seen by the baby were covered by surgical masks. Reprinted from Johnson and Morton (1991) with permission; copyright 1991 Blackwell Publishing.

of the face is masked, and argued that newborns are learning only face outlines, not faces. They concluded that newborn mother learning might differ qualitatively from adult face learning. However, HLISSOM simulation results in chapter 8 will show that learning of the whole face (internal features and outlines) can also result in mother preferences. Importantly, masking the outline can still erase these preferences, even though outlines were not the only parts of the face that were learned. Thus whether mother preferences are due to outlines alone remains an open question. Newborns may instead learn faces holistically, as in HLISSOM.

Experiments with infants as they develop over the first few months reveal a surprisingly complex pattern of face preferences. Newborns up to one month of age continue to track facelike schematic patterns in the periphery, but older infants do not (Johnson et al. 1991). Curiously, in central vision, schematic face preferences are not measurable until about two months of age (Maurer and Barrera 1981), and they decline by five months of age (Johnson and Morton 1991). Chapter 8 will show that in each case the decline in response to schematic faces can be a natural result of learning real faces.
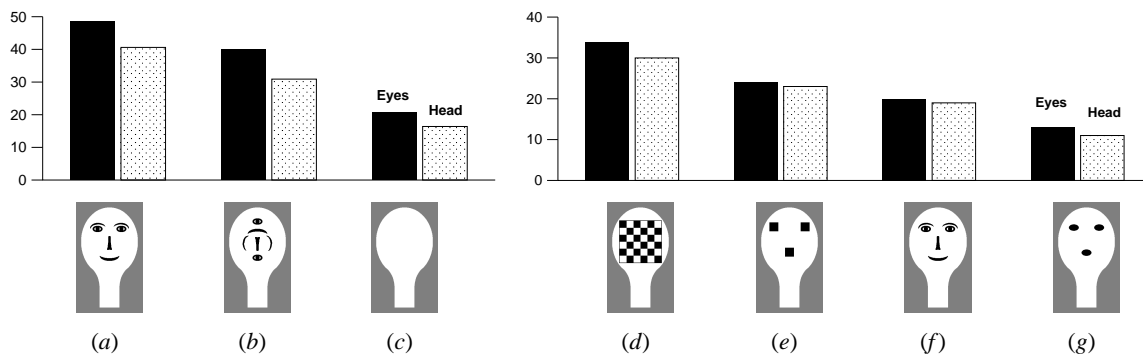
Figure 2.9: **Face preferences at birth.** Using the procedure from figure 2.8, Johnson et al. (1991) measured responses to a set of head-sized schematic patterns. The graph at left gives the result of a study of human newborns tested with two-dimensional moving patterns within one hour of birth (Johnson et al. 1991); the one at right gives results from a separate study of newborns an average of 21 hours old (also published in Johnson et al. 1991). Each pair of bars represents the average newborn eye and head tracking, in degrees, for the image pictured below it; eye and head tracking had similar trends here and thus either may be used as a measure. Because the procedures and conditions differed between the two studies, only the relative magnitudes should be compared. Overall, the study at left shows that newborns respond to face-like stimuli more strongly than to simple control conditions; all comparisons were statistically significant. This result suggests that there is some genetic basis to face processing abilities. In the study at right, the checkerboard pattern (*d*) was tracked significantly farther than the other stimuli, and pattern (*g*) was tracked significantly less far. The checkerboard was chosen to be a good match to the low-level visual preferences of the newborn, and shows that such preferences can outweigh face preferences. No significant difference was found between the responses to (*e*) and (*f*). The results from this second study suggest that face preferences are broad, perhaps as simple as a preference for two dark blobs above a third one. Adapted from Johnson et al. (1991).

2.4 Conclusion

Both internal and environmental factors strongly influence the development of the human visual system. For orientation, which is processed similarly by many species of experimental animals, these influences have been studied at the neural level. Recent evidence suggests that, before eye opening, spontaneously generated activity leads to a noisy version of the orientation map seen in adults. Subsequent normal visual experience increases orientation selectivity, but does not dramatically change the overall shape of the map. Yet abnormal experience can have large effects. These experimental results make orientation processing a well-constrained test case for studying how internal and external sources of activity can affect development.

Much less is known about the neural basis of face processing, in the adult and especially in newborns. However, behavioral experiments with human newborns and infants suggest that there is some capacity for face detection at birth, and that it further develops postnatally from experience with real faces. Thus the development of face-selective neurons appears to involve both prenatal and postnatal factors, as for the neurons in the orientation map. In each case, internally generated neural activity offers a simple explanation for how the same system could organize before and after birth, learning from patterns of activity.

Chapter 3

Related work

In the previous chapter I outlined the experimental evidence for how orientation and face processing develop in infants. Later in the thesis I will use this evidence to build the HLISSOM computational model of visual development. In this chapter, I review other computational models and theoretical explanations of visual development, and show how they relate to HLISSOM. I will focus first on models of orientation processing, then on models of face processing.

3.1 Computational models of orientation maps

As reviewed in the previous chapter, visual system development is a complex process, and it is difficult to integrate the scattered experimental results into a specific, coherent understanding of how the system is constructed. Computational models provide a crucial tool for such integration. Because each part of a model must be implemented for it to work, computational models require that often-unstated assumptions be made explicit. The model can then show what types of structure and behavior follow from those assumptions. In a sense, computational models are a concrete implementation of a theory: they can be tested just like animals or humans can, either to validate the theory or to provide predictions for future experimental tests.

In this section I will review computational models for how orientation selectivity and orientation maps can develop. The discussion is roughly chronological, focusing on the model properties that will be needed for the simulations in this thesis: developing realistic orientation maps, receptive fields, and patchy lateral connections; self-organizing based on internally generated activity and/or grayscale natural images, and providing output from the orientation map suitable as input for a higher area (for the face processing experiments). As reviewed below, no model has yet brought all these elements together, but many previous models have had some of these properties.

3.1.1   von der Malsburg's model

Over the years, computational models have been limited by the computational power available, but even very early models were able to show how orientation selectivity can develop computationally (Barrow 1987; Bienenstock, Cooper, and Munro 1982; Linsker 1986a,b,c; von der Malsburg 1973; see Swindale 1996 and Erwin et al. 1995 for critical reviews). Pioneering studies by von der Malsburg (1973) using a 1MHz UNIVAC first demonstrated that columns of orientation-selective neurons could develop from unsupervised learning of oriented bitmap patterns (with each bit on or off). This model already had many of the features of later ones, such as treating the retina and cortex as two-dimensional arrays of units, using one unit for each column of neurons, using a number to represent the firing rate of a unit, using a number to represent the strength of the connection between two neurons, assuming fixed-strength isotropic lateral interactions within V1, and assuming lateral inhibitory connections have a wider radius than lateral excitatory connections. Figure 3.1 outlines the basic architecture of models of this type.

Like HLISSOM and most other models, the von der Malsburg (1973) model was based on incremental Hebbian learning, in which connections are strengthened between two neurons if those neurons are activated at the same time (Hebb 1949). To prevent strengths from increasing without bound, the total connection strength to a neuron was normalized to have a constant sum (as in Rochester, Holland, Haibt, and Duda 1956). Given a series of input patterns, the model self-organized into iso-orientation domains, i.e. patches of neurons preferring similar orientations. The model also exhibited pinwheels, which had not yet been discovered in the visual cortex of animals.

3.1.2   SOM-based models

Because of computational constraints, von der Malsburg's (1973) model included only a small number of neurons, all with overlapping RFs, and thus represented only orientation and not position on the retina. Later work showed how such a topographic map of the retina could develop, e.g. using a more abstract but computationally efficient architecture called the Self-Organizing Map (SOM; Kohonen 1982). Durbin and Mitchison (1990) and Obermayer, Ritter, and Schulten (1990) extended the SOM results to the orientation domain, showing that realistic retinotopic, orientation, and orientation selectivity maps could develop. These SOM orientation maps replicate many of the properties of experimental maps like the one in figure 2.4, including pinwheels, fractures, and linear zones.

The SOM models are driven by oriented features in their input patterns, using a different randomly chosen orientation and position at each iteration. Durbin and Mitchison (1990) provided the orientation and position as numbers directly to the model, while Obermayer et al. (1990) used them to draw a grayscale oriented Gaussian bitmap pattern on a model retina. Either way, these oriented features can be seen as an abstract representation of small parts of visual images. Later work discussed below showed how real images could also be used as input.

21
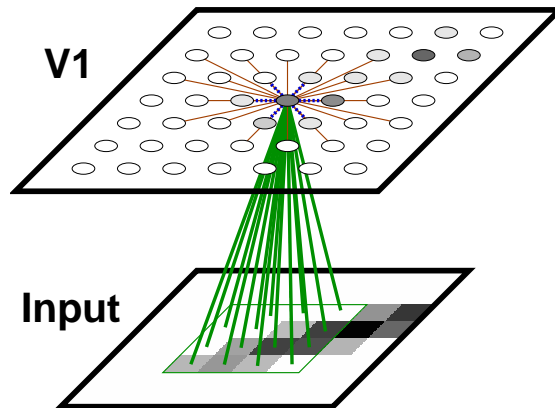
Figure 3.1: **General architecture of orientation map models.** Models of this type typically have a two-dimensional bitmap input sheet where an abstract pattern or grayscale image is drawn. The sheet is usually either a hexagonal or a rectangular grid; a rectangular 8×8 bitmap grid is shown here. Instead of a bitmap input, some models provide the orientation and x, y position directly to input units (Durbin and Mitchison 1990). Others dispense with individual presentations of input stimuli altogether, abstracting them into functions that describe how they correlate with each other over time (Miller 1994). Neurons in the V1 sheet have afferent (incoming) connections from neurons in a receptive field on the input sheet. Sample afferent connections are shown as thick lines for a neuron in the center of the 7×7 V1 sheet. In some models the RF is the entire input sheet (e.g. von der Malsburg 1973). In addition to the afferent input, neurons in V1 generally have short-range excitatory connections to their neighbors (short dotted lines) and long-range inhibitory connections (long thin lines). Most models save computation time and memory by assuming that the values of these lateral connections are fixed, isotropic, and the same for every neuron, but specific connections are needed for many phenomena. Neurons generally compute their activation level as a scalar product of their weights and the units in their receptive fields. Sample V1 activation levels are shown in grayscale for each unit. Weights that are modifiable are updated after an input is presented, using an unsupervised learning rule. In SOM models, only the most active unit and its neighbors are adapted; others adapt all active neurons. After many input patterns are presented, the afferent weights for each neuron become stronger to units lying along one particular line of orientation, and thus become orientation selective.

3.1.3 Correlation-based learning (CBL) models

Miller (1994) used a different approach to making simulations practical from that used by SOM. Miller showed that if the visual system is essentially linear, then the sequence of input patterns can be replaced with a simple function representing its long-term correlations. This approximation speeds up the calculations considerably compared to incremental Hebbian learning, and makes theoretical analysis much simpler. Miller's approach is sometimes called correlation-based learning (CBL; Erwin et al. 1995), but that term is confusing because most other models are also driven by some type of correlations.

Unlike most previous models, Miller's included the ON and OFF cell layers for the LGN; he assumed that they would compete for total synaptic strength to each neuron in V1. With this framework, he found that multi-lobed receptive fields like those in figure 2.2*c-d* can develop, and

that the neurons form orientation maps with pinwheels and fractures. However, the orientation maps lack several key features seen in animal maps, including linear zones and periodically repeating blobs of iso-orientation patches (Erwin et al. 1995; Swindale 1996). The final RFs are also nearly binary in strength, with sharp edges between ON and OFF subregions; animal RFs instead have smooth edges and transitions (Jones and Palmer 1987). Thus the more realistic incremental Hebbian approaches appear to be required, even though these are more difficult to simulate and analyze.

3.1.4   RF-LISSOM

Although SOM models have been more successful than CBL models at reproducing the animal orientation maps, they too do not explain how the patchy lateral connections develop along with the orientation map. Instead SOM uses a fixed "neighborhood function" to represent lateral interactions, which requires much less computation time and memory than storing specific, plastic connections. However, several authors have recently proposed that prenatally organized lateral connections drive the development of the orientation map (e.g. Adorján, Levitt, Lund, and Obermayer 1999; Bartsch and van Hemmen 2001; Ernst, Pawelzik, Sahar-Pikielny, and Tsodyks 2001; Shouval, Goldberg, Jones, Beckerman, and Cooper 2000). Thus a complete account of orientation map development needs to include an explanation for how the patchy lateral connections develop prenatally.

The first cortical map model to include patchy lateral connections was RF-LISSOM (the Receptive-Field Laterally Interconnected Synergetically Self-Organizing Map; Miikkulainen, Bednar, Choe, and Sirosh 1997; Sirosh 1995; Sirosh and Miikkulainen 1994, 1997; Sirosh, Miikkulainen, and Bednar 1996). RF-LISSOM was based on the von der Malsburg (1973) and SOM models. In part thanks to advances in supercomputing hardware, many of the approximations from the earlier models were no longer necessary in RF-LISSOM. Like SOM, RF-LISSOM produced orientation maps with the qualitative features found in animal maps. At the same time, it showed how patchy connections like those in figure 2.5 could develop, preferentially targeting neurons of similar orientation and extending along the axis of the neuron's preferred orientation. In this model, afferent and lateral connections develop synergetically through incremental Hebbian learning. Simulations of adult visual performance have demonstrated that the specific lateral connections allow the network to better suppress redundancy in the visual input (Sirosh 1995), and have shown that adapting lateral connections can explain orientation-specific visual artifacts and may underlie perceptual grouping abilities (Bednar and Miikkulainen 2000b; Choe 2001; Choe and Miikkulainen 2000b). The HLIS-SOM model introduced in the next chapter is based on RF-LISSOM, but extended to work with natural images as described below.

3.1.5   Models based on natural images

Like the SOM models before it, the RF-LISSOM orientation map develops only single-lobed RFs, because it does not include models of the ON and OFF cell layers of the LGN. In addition to being

less realistic, chapter 4 (figure 4.9) will show that this limitation prevents RF-LISSOM from being used with natural images. Natural images will be crucial in later chapters for testing how visual inputs interact with the prenatally organized orientation map, as well as for testing face perception. To make such tests possible, HLISSOM will include models for the ON and OFF cells.

Using ON and OFF cells or filtering that approximates them, a number of previous models have shown how orientation selectivity can develop from natural images, without studying how they can organize into an orientation map (Barrow 1987; Barrow and Bray 1992; Blais, Shouval, and Cooper 1999; Einhauser, Kayser, Konig, and Kording 2002; Lee, Blais, Shouval, and Cooper 2000; Olshausen and Field 1996). A few models have also shown how orientation maps can develop from natural images (Barrow and Bray 1993; Burger and Lang 1999; Hyvärinen and Hoyer 2001; Shouval, Intrator, and Cooper 1997; Weber 2001). The performance of these networks on individual natural images has not yet been reported, so it is not known whether their output would be suitable as input for a higher cortical level (as will be shown for HLISSOM in chapter 7). For instance, I expect that most models will respond only to the highest-contrast contours in an image, such as face outlines, while having no response to lower-contrast areas like the internal features of the face.

3.1.6   Models with lateral connections

Several models with modifiable lateral connections have also been introduced very recently, developed concurrently with HLISSOM but after RF-LISSOM (Alexander, Bourke, Sheridan, Konstandatos, and Wright 1999; Bartsch and van Hemmen 2001; Bray and Barrow 1996; Burger and Lang 1999; Weber 2001). Of these, the Alexander et al. (1999) and Burger and Lang (1999) models develop patchy long-range lateral connections, but the rest do not. As mentioned earlier, explaining how such connections develop is a crucial part of explaining the prenatal development of the orientation map. The Alexander et al. (1999) model relies on abstract binary input patterns like those in von der Malsburg (1973), and it is not clear how to extend it to support grayscale and natural image inputs, or to develop multi-lobed receptive fields. Thus apart from Burger and Lang (1999), which will be discussed below, these other laterally connected models are not appropriate platforms for the phenomena to be studied in this thesis.

3.1.7   The Burger and Lang model

Like HLISSOM, the Burger and Lang (1999) model includes ON and OFF cells and lateral connections, and can organize maps using natural images and random noise. Despite having been developed independently, the model is very similar to the first few visual processing stages of HLISSOM (specifically, the photoreceptors, LGN, and V1). Chapter 6 will show that HLISSOM can also develop orientation maps based on random noise, but that more realistic maps develop from structured noise, like retinal waves. Mathematically, the only significant difference between their model and HLISSOM is in the activation function. The Burger and Lang model uses a linear activation func-

tion (sum of all inputs), whereas HLISSOM's activation function is nonlinear, and also iteratively incorporates lateral influences. The nonlinearity ensures that neurons remain orientation-selective over a wide range of contrasts. Such contrast-invariance is crucial when using the model as input to another cortical sheet, because different regions of large natural images have very different contrasts. Although the Burger and Lang model also uses a different method for normalizing the afferent weights from that used in HLISSOM (and RF-LISSOM), either of these methods should have similar results. Thus if extended with an activation function that will work with large natural images, the Burger and Lang model could be used for orientation map simulations similar to those in this thesis. Previous orientation map models would need significant extensions or modifications.

3.1.8 Models combining spontaneous activity and natural images

In part for the reasons discussed above, none of the map models have yet demonstrated how the orientation map can smoothly integrate information from internally generated patterns and from the environment. A number of the models have simulated spontaneously generated activity (e.g. Burger and Lang 1999; Linsker 1986a,b,c; Miller 1994), and a few models have shown self-organization based on natural images (section 3.1.5). Yet to my knowledge, the only orientation map model to be tested on a prenatal phase with spontaneous patterns, followed by a postnatal phase, is Burger and Lang (1999). They found that if a map organized based on uniformly random noise was then trained on natural images (actually, patches from a single natural image), the initial structure was soon overwritten. As described in chapter 2, this is a curious result because animal maps instead appear to keep the same overall structure during postnatal development. Chapter 6 will show orientation maps learned from internally generated activity that has spatial structure, like the retinal waves do. An example is illustrated in figure 3.2. After learning an orientation map from such activity, postnatal experience will smoothly and locally change the map to be a better match to the statistics of the natural environment, without changing the overall structure of the map. This way HLISSOM will show how both the spontaneous activity and natural images can interact to construct realistic orientation maps, an important finding that has not been explained by previous models.

3.2 Computational models of face processing

The models discussed so far have simulated visual processing only up to V1, and therefore did not include any of the higher cortical areas that are thought to underlie face processing abilities. A number of models do simulate face processing, but each of these either bypasses the circuitry up to V1, or treats it as a fixed set of predefined filters (as reviewed in Valentin, Abdi, O'Toole, and Cottrell 1994). There are also many similar systems that are not intended as biological models, but are instead focused on specific engineering applications such as face detection or face recognition (e.g. Rao and Ballard 1995; Rowley, Baluja, and Kanade 1998; Viola and Jones 2001; for review see Yang, Kriegman, and Ahuja 2002, and Phillips, Wechsler, Huang, and Rauss 1998). Given the
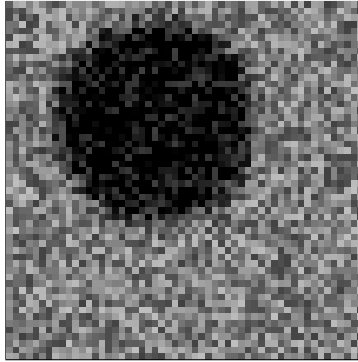
Figure 3.2: **Proposed model for spontaneous activity.** Chapter 6 will show that orientation maps like the neonatal cat map in figure 2.6 can develop from internally generated activity like that shown above, which is modeled after retinal waves (figure 1.1). Unlike the purely random noise used in previous models of spontaneous activity, this pattern contains blobs of activity with oriented edges. The oriented edges are the crucial feature that will drive the development of realistic orientation maps. As long as these blobs are large (relative to a V1 receptive field size), their shape and size are not crucial for the model, nor is the background noise. HLISSOM will learn initial orientation maps based on this type of activity, and then natural images will gradually change the prenatally organized map structure. This process will explain how orientation maps can develop both prenatally and postnatally.

output of the filtering stage, both the biological models and the engineering systems show how face-selective neurons or responses can develop from training with real images (e.g. Acerra, Burnod, and de Schonen 2002; Bartlett and Sejnowski 1997, 1998; Dailey and Cottrell 1999; Gray, Lawrence, Golomb, and Sejnowski 1995; Wallis 1994; Wallis and Rolls 1997). HLISSOM will also show how face-selective neurons can develop, but is the first model to use a self-organized V1 as the input stage, and moreover to use the same algorithm to model processing in V1 and a higher face processing area. HLISSOM thus represents a unification of these higher-level models with the V1 models discussed earlier.

Of the biological models, the Dailey and Cottrell (1999) and Acerra et al. (2002) models are relevant to the work presented in this thesis. Acerra et al. (2002) is the only model that simulates newborn face preferences, and it will be discussed in section 3.3 below. Although Dailey and Cottrell (1999) does not specifically address newborn face preferences, it focuses on genetic aspects of face detection, and it forms the basis for the remainder of this section.

Dailey and Cottrell (1999) show in an abstract model how specific face-selective regions can arise without genetically specifying the weights of each neuron. As reviewed in chapter 2, some of the higher visual areas of the adult human visual system respond more to faces than objects. Others have the opposite property. Moreover, some of the face-selective areas have been shown to occupy the same region of the brain in different individuals (Kanwisher et al. 1997). This consistency suggests that those areas might be genetically specified for face processing.

To show that such explicit prespecification is not necessary, Dailey and Cottrell (1999) set up

a pair of supervised networks that compete with each other to identify faces and to classify objects into categories. They provided one of the networks with real images filtered to preserve low–spatial-frequency information (i.e., slow changes in brightness across a scene), and another with the images filtered to preserve high–spatial-frequency information. These differences correspond to connecting each network to a subset of the neurons in V1, each with different preferred spatial frequencies. They found that the low-frequency network consistently developed face-selective responses, while the high-frequency network developed object-selective responses. Thus they concluded that different areas may specialize for different tasks based on very simple, general differences in their connectivity, and that specific configuration of individual neurons need not be specified genetically to respond to faces.

Like the Dailey and Cottrell (1999) model, HLISSOM is based on the assumption that cortical neurons are not specifically prewired for face perception. However, to make detailed simulations practical, HLISSOM will model only a single higher level region, one that has sufficient spatial frequency information available to develop face selective responses. Other regions not modeled presumably develop similarly, but become selective for objects or other image features instead.

3.3 Models of newborn face processing

The computational models discussed above did not specifically explain why a newborn would show face-selective behavioral responses, as was reviewed in chapter 2. With the exception of Acerra et al. (2002), models of newborn face preferences have all been conceptual, not computational. I will review these models below, and show how they compare to the HLISSOM model.

There are four main previous explanations for how a newborn can show an initial face preference and 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, such as the poor optics of the newborn eye, the sampling density of the retinal ganglion cells, etc. The LSM does not explicitly state how adults develop face processing abilities, but the assumption is that those abilities are learned from postnatal experience with real faces. (2) *Sensory models* are generalizations of the LSM, stating that the spatial frequency response plus other general features of the sensory apparatus together account for newborn face preferences, still without face-specific circuitry present at birth. There are two main models in this class, the Acerra et al. (2002) computational model and the *Top-heavy* conceptual model, which I will discuss separately. (3) the *Haptic* hypothesis posits that newborn face preferences result from the LSM plus prenatal experience with the embryo's own face, either through somatosensory information from the hands resting upon it, or by direct proprioception of facial muscles. (4) *Multiple systems models* propose that there is a hardwired face-specific visual

processing system present at birth, and that this system is later replaced or augmented by a separate, plastic system that operates into adulthood. These four hypotheses will be reviewed below, and a simpler, more effective alternative will be proposed: that only a single, general-purpose visual processing system is necessary if that system is exposed to *internally generated patterns*, such as those in REM sleep.

3.3.1 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 can easily be implemented as a computational model. It is based solely on the newborn's measured contrast sensitivity function (CSF). For a given spatial frequency, the value of the CSF will be high if the early visual pathways respond strongly to that that size of pattern, and lower otherwise. The newborn CSF is limited by the immature state of the eye and the early visual pathways, and thus lower frequencies are more visible than fine detail.

The assumption of the LSM is that newborns simply pay attention to those patterns that when convolved with the CSF give the largest response. Low-contrast patterns and patterns with only very fine detail are only faintly visible, if at all, to newborns (Banks and Salapatek 1981). 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). The LSM also predicts that the newborn will respond equally well to a schematic face regardless of its orientation, because the orientation does not affect the spatial frequency or the contrast. Instead, newborns prefer schematic face-like stimuli oriented right-side-up. Such a preference is found even when the inverted stimulus is a better match to the CSF (Valenza et al. 1996). Thus the CSF alone does not explain face preferences, and a more complicated model is required.

3.3.2 Acerra et al. sensory model

The LSM is only a high-level abstraction of the properties of the early visual system. In particular, it does not take into account that neurons have finite RFs, i.e. that neurons respond to limited portions of the visual field. Incorporating such general limitations, Acerra et al. (2002) recently developed a sensory computational model of processing in V1 that can account for some of the face preferences found in the Valenza et al. (1996) study. Their model consists of a fixed filter-based model of V1 (like that of Dailey and Cottrell 1999), plus a higher level sheet of neurons with modifiable connections. They model two conditions separately: newborn face preferences, and postnatal development

by 4 months. The newborn model includes only V1, because they assume that the higher level sheet is not yet functional at birth.

Acerra et al. showed that the newborn model responds slightly more strongly to the upright schematic face pattern used by Valenza et al. (1996) than to the inverted one, because of differences in the responses of a subset of the V1 neurons. This surprising result replicates the newborn face preferences from Valenza et al. The difference in model responses is because Valenza et al. actually inverted only the internal facial features, not the entire pattern. In the upright case, the spacing is more regular between the internal features and the face outline (compare figure 7.5*d* with 7.5*g*, row Retina, on page 79). As can be inferred from the RF shape in figure 2.2*d*, making the spacing more regular between alternating white and black blobs will increase the response of a neuron whose RF lobes match the blob size. Because of this better match with a subset of V1 neurons, the V1 response was slightly larger overall to the facelike (upright) pattern than to the non-facelike (inverted) pattern.

However, the Acerra et al. model was not tested with patterns from other studies of newborn face preferences, such as Johnson et al. (1991). The facelike stimuli published in Johnson et al. (1991) do not have a regular spacing between the internal features and the outline, and so I do not expect that the Acerra et al. model will replicate preferences for these patterns. Moreover, Johnson et al. used a white paddle against a light-colored ceiling, and their face outlines would have a much lower contrast than the black-background patterns used in Valenza et al. (1996). Thus although border effects may contribute to the face preferences found by Valenza et al., they are unlikely to be explain those measured by Johnson et al.

The Acerra et al. newborn model also does not explain newborn learning of faces, because their V1 model is fixed and the higher level area is assumed not to be functional at birth. Most importantly, the Acerra et al. newborn 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. Because of these differences, I do not expect that the Acerra et al. model will show a significantly higher response overall to photographs of real faces than to other similar images. The HLISSOM model will make the opposite prediction, and will also explain how newborns can learn faces.

To explain learning of real faces in older infants, the Acerra et al. model relies on having face images be strictly aligned on the input, having nothing but faces presented to the model (no objects, bodies, or backgrounds), and on having the eyes in each face artificially boosted by a factor of 10 or 100 relative to the rest of the image. Because of these unrealistic assumptions, it is difficult to evaluate the validity of their postnatal learning model. In contrast, the HLISSOM model learns from faces presented at random locations on the retina, against natural image backgrounds, intermixed with images of other objects, and without special emphasis for faces relative to the other objects.

3.3.3 Top-heavy sensory model

Simion et al. (2001) have also presented a sensory model of newborn preferences; it is a conceptual

model only. They observed that nearly all of the face-like schematic patterns that have been tested with newborns are "top-heavy," i.e. they have a boundary that contains denser patterns in the upper than the lower half. Simion et al. (2001) also ran behavioral experiments showing that newborns prefer several top-heavy (but not facelike) schematic patterns to similar but inverted patterns. Based on these results, they proposed that newborns prefer top-heavy patterns in general, and thus prefer facelike schematic patterns as a special case.

This hypothesis is compatible with most of the experimental data so far collected in newborns. However, face-like patterns have not yet been compared directly with other top-heavy patterns in newborn studies. Thus it is not yet known whether newborns would prefer a facelike pattern to a similarly top-heavy but not facelike pattern. Future experimental tests with newborns can resolve this issue.

To be tested computationally, the top-heavy hypothesis would need to be made more explicit, with a specific mechanism for locating object boundaries and the relative locations of patterns with them. I expect that such a test would find only a small preference (if any) for photographs of real faces, compared to many other common stimuli. For faces with beards, wide smiles, or wide-open mouths, many control patterns would be preferred over the face image, because such face images will no longer be top-heavy. Because the bulk of the current evidence suggests that instead the newborn preferences are more selective for faces, most of the HLISSOM simulations will use training patterns that result in strongly face-selective responses. Section 7.3.3 will show results from patterns with lower levels of face specificity, which would be more comparable to the top-heavy model.

3.3.4 Haptic hypothesis

Bushnell (1998) proposed an explanation very different from that of the sensory models: a newborn may recognize face-like stimuli as a result of prenatal experience with its own face, via manual exploration of its facial features. Some support for this position comes from findings that newborn and infant monkeys respond as strong or more strongly to pictures of infant monkeys as to adults (Rodman 1994; Sackett 1966). However, the process by which a newborn could make such specific connections between somatosensory and visual stimulation, prior to visual experience, is not clear.

Moreover, the haptic explanation does not account for several aspects of newborn face preferences. For instance, premature babies seem to develop face preferences at the same post-conception age regardless of the age at which they were born (Ferrari, Manzotti, Nalin, Benatti, Cavallo, Torricelli, and Cavazzutti 1986). Presumably, the patterns of hand and arm movements would differ between the intrauterine and external environments, and thus the haptic hypothesis would predict that gestation time should have been an important factor. Several authors have also pointed out strong similarities between newborn face preferences and imprinting in newly hatched chicks; chicks, of course, do not have hands with which to explore, yet develop a specific preference for stimuli that resemble a (chicken's) head and neck (Bolhuis 1999; Horn 1985). Some of these

objections are overcome in a variant of the haptic hypothesis by Meltzoff and Moore (1993), who propose that direct proprioception of the infant's own facial muscles is responsible.

However, both variants fail to account for data that suggest the newborns' preferences are specifically visual. For instance, newborns respond as well to patterns with a single blob in the nose and mouth area as to separate patterns for the nose and mouth (Johnson et al. 1991). This finding is easy to explain for visual images: in a blurred top-lit visual image, shadows under the nose blend together with the mouth into a single region. But the nose and mouth have opposite convexity, so it is difficult to see how they could be considered a single region for touch stimulation or proprioception. Similarly, newborn preferences have so far only been found for faces viewed from the front, and it is not clear why manual exploration or proprioception would tend to favor that view in particular. Thus this thesis will assume that newborn face preferences are essentially visual.

3.3.5 Multiple-systems models

The most widely known conceptual model for newborn face preferences and later learning was proposed by Johnson and Morton (1991). Apart from tests in the HLISSOM model, it has not yet been evaluated computationally. Johnson and Morton suggested that infant face preferences are mediated by two hypothetical systems that they dubbed CONSPEC and CONLERN. CONSPEC is a hardwired face-processing system, assumed to be located in the subcortical 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 (for examples see figure 7.3*a*, page 76 and the Retina row of 7.4*c*, page 78). CONLERN is a separate plastic cortical system, presumably the face-processing areas that have been found in adults. The CONLERN system would assume control only after about 6 weeks of age, and would account for the face preferences seen in older infants. Finally, as it learns from real faces, CONLERN would gradually stop responding to static schematic faces, which would explain why face preferences can no longer be measured with schematic patterns (unless their features are moving realistically) by five months (Johnson and Morton 1991).

The CONSPEC/CONLERN 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, McHaffie, and Stein 1997). The model also helps explain why infants after month of age show a reduced interest in faces in the periphery, 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, Mancini, and Leigeois (1998) argue for *three* systems: a subcortical one responsible for facial feature preferences at birth, another one
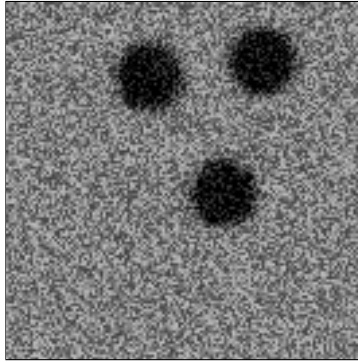
Figure 3.3: **Proposed face training pattern.** Based on the experiments of Johnson and Morton (1991), a training pattern consisting of two blobs of activity above a third might account for the newborn face preferences. This configuration will be used for most face processing experiments in this thesis; figure 7.9 will test other possible patterns.

responsible for newborn learning (of objects and head/hair outlines; Slater 1993), and a cortical system responsible for older infant and adult learning of facial features. And Simion, Valenza, and Umiltà (1998a) proposed that face selectivity instead relies on multiple systems *within* the cortex, maturing first in the dorsal stream but later supplanted by the ventral stream (which is where most of the face-selective regions have been found in adult humans).

In contrast to the increasing complexity of these explanations, I propose that a single general-purpose plastic visual processing system is sufficient, if that system is first exposed to internally generated face-like patterns of neural activity. As reviewed in section 2.2.2, PGO activity waves during REM sleep represent a likely candidate for such activity patterns. This thesis predicts that if the PGO activity patterns have the simple configuration illustrated in figure 3.3 (or similar patterns), they can account for the measured face-detection performance of human newborns. These three-dot training patterns are similar to the hardwired preferences proposed by Johnson and Morton (1991). However, instead of templates implemented in a hard-wired subcortical face-detecting region, in the HLISSOM 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 (as described in chapter 8). In some sense HLISSOM is also a multiple systems model, like CONSPEC/CONLERN, but HLISSOM shows how the genetically specified components can be separated from the visual processing hardware. The specification for the patterns can then evolve separately from the actual visual hardware, which allows the genetic information to be expressed within a large, complex adaptive system. Moreover, the resulting single learning system is plastic throughout, and can thus explain how infants at all ages can learn faces.

3.4 Conclusion

Previous orientation map models have not yet shown how the prenatal and postnatal phases of orientation map development can be integrated smoothly, as appears to happen in animals. They have also not shown how a self-organized V1 network can extract orientations from natural images in a form that can be used by a higher level network, as V1 does in animals. The HLISSOM model will overcome these limitations to show how Hebbian learning can account for orientation maps at birth and their postnatal development.

Previous face processing models have not yet shown how newborns could have a significant preference for real faces at birth, or how newborns could learn from real faces. They have also not accounted for the full range of patterns with which newborns have been tested. Using the self-organized orientation map, the HLISSOM model of face detection will show how face-selective neurons can develop through spontaneous activity, explaining face preferences at birth and later face learning. As will be shown in later chapters, the overall method represents a new approach to designing complex adaptive systems, combining the strengths of evolution and learning.

Chapter 4

The HLISSOM model

This thesis presents a series of simulations with HLISSOM (Hierarchical LISSOM; Bednar and Miikkulainen 2001). HLISSOM is a biologically motivated model of the human visual pathways, focusing on development at the cortical level. It is based on RF-LISSOM (section 3.1.4), extended to handle real images (such as human faces), multiple sources of input activity (from the eyes and brainstem), and multiple levels of cortical processing (V1 and a face selective area). HLISSOM is the first model to bring all of these components and capabilities together, which makes it possible to make detailed predictions and comparisons with experimental data. This chapter describes the HLISSOM architecture and learning algorithm in detail. To make the model concrete and to relate it to previous work, I will also present results from a basic simulation of orientation map development.

4.1 Architecture

I will first give an overview of the model at a high level, then describe the initial state of the network, before learning. Later sections will show how the response to input patterns is computed, how the activity patterns change the components of the model, and how orientation maps can arise through self-organization.

4.1.1 Overview

As reviewed in chapter 2, HLISSOM consists of the key areas of the visual system that are necessary to model how orientation and face processing develop. The architecture is illustrated in figure 4.1. The model is a hierarchy of two-dimensional sheets of neural units representing different brain regions: 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: the primary visual cortex (V1), and a higher level area here called the face-selective area (FSA) and explained below. This overall division into brain regions is assumed to be specified genetically, as is thought to be the case

Face–selective area (FSA)

V1

ON–cells

OFF–cells

LGN

Photoreceptors

PGO pattern generator
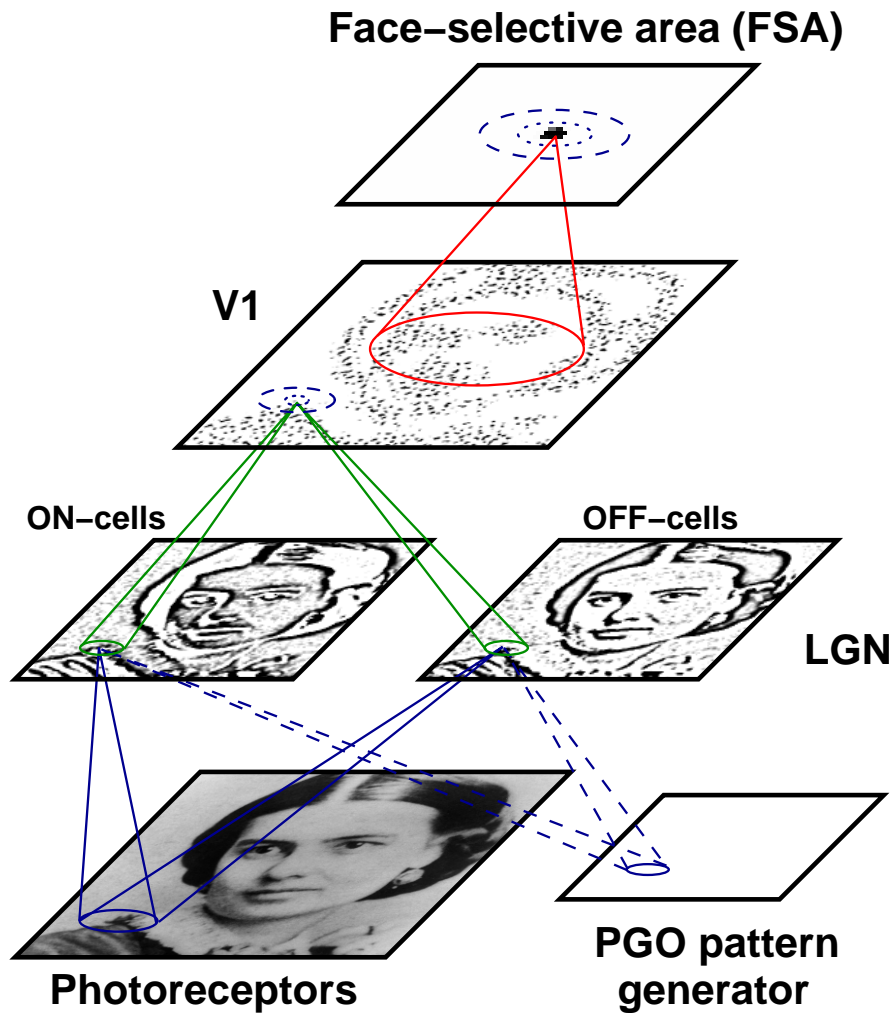
Figure 4.1: **Architecture of the HLISSOM model.** Each sheet of units in the model visual pathway is shown with a sample activation pattern and the connections to one example unit. Grayscale visual inputs are presented on the photoreceptors, and the resulting activity propagates through afferent connections to each of the higher levels. Internally generated PGO input propagates similarly to visual input. Activity is 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. 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 response is patchy because neurons preferring other orientations do not respond. The FSA represents the highest level of abstraction — an FSA neuron responds when there is a high probability of a face being at the corresponding location on the retina.

in animals and humans (Miyashita-Lin et al. 1999; Rakic 1988).

Within the LGN and the cortical sheets, units receive afferent connections from broad over-lapping circular patches of units (i.e., RFs) in the previous sheets in the hierarchy. Units in the cortical sheets also have reciprocal connections to units within the same sheet, modeling lateral interactions between neurons. The strengths of these afferent and lateral connections in the model cortex are set by unsupervised learning.

The learning process is driven by input patterns drawn on the input sheets, either the pho-toreceptors or PGO generator. Both input sheets are represented by an array of numerical values, as in a grayscale bitmap image. Unfortunately, the PGO pathway has not yet been mapped in detail in animals, but the activity that results from the PGO waves appears to be similar to that from visual input (Marks et al. 1995). Thus for simplicity the PGO pathway is modeled with an area like the photoreceptor sheet, connecting to the LGN in the same way.

Activity on the input sheets propagates successively to the higher levels, each of which transforms the activity pattern into a more biologically relevant representation. The ON and OFF layers of the LGN are arrays of units with fixed RFs that filter out large, uniformly bright or dark areas, leaving only edges and lines. The cortical sheets, in V1 and the FSA, consist of initially unselective units that become selective through learning. Each unit in V1 or the FSA corresponds to a vertical column of cells through the six anatomical layers of that area of the cortex (as described in section 2.1.1). As in most models, these units will be called "neurons" even though they strictly correspond to groups of neurons in the cortex. Through self-organization, neurons in V1 become selective for particular orientations of the edges and lines in the LGN activity.

The FSA represents the first region in the ventral processing pathway above V1 that has receptive fields spanning approximately $45°$ of visual arc, i.e. large enough to span a human face at close range. Areas V4v and LO are likely FSA candidates based on adult patterns of connectivity, but the infant connectivity patterns are not known (Rolls 1990; Rodman 1994; Kanwisher et al. 1997; Haxby et al. 1994). The generic term "face-selective area" is used rather than V4v or LO to emphasize that the model results do not depend on the region's precise location or architecture, only on the fact that the region has receptive fields large enough to allow face-selective responses. Through self-organization, neurons in the FSA become selective for patterns similar to faces, and do not respond to most other objects and scenes.

The following subsections describe the specific components of the HLISSOM model in more detail.

4.1.2 Connections to the LGN

Previous models have explained how the connections from the retina to the LGN could develop from internally generated activity in the retina (Eglen 1997; Haith 1998; Keesing, Stork, and Shatz 1992). HLISSOM instead focuses on learning at the cortical level, so all connections to neurons in the ON and OFF channels of the LGN are set to fixed strengths.
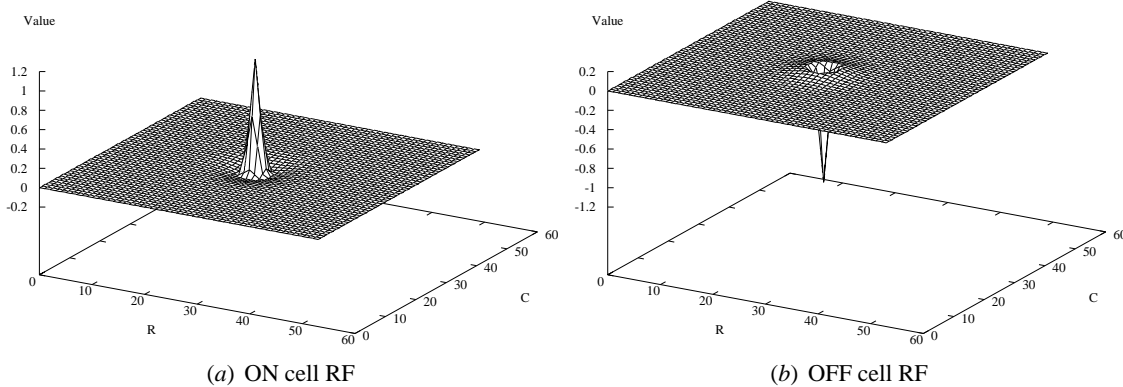
Figure 4.2: **ON and OFF cell RFs.** The ON and OFF cells in HLISSOM are each modeled as a Difference of two Gaussians (DoG), the center and the surround. The Gaussians are normalized to have the same total strength, but the center Gaussian concentrates that strength in a much smaller region. ON cells have an excitatory center and an inhibitory surround (*a*), and OFF cells have an inhibitory center and an excitatory surround (*b*). These RFs perform edge detection at a spatial frequency determined by the center width; they highlight areas of the input image that have edges and lines, and do not respond to large areas of constant illumination.

The strengths were chosen to approximate the receptive fields that have been measured in adult LGN cells, using a standard Difference-of-Gaussians model (Rodieck 1965; see also Cai, DeAngelis, and Freeman 1997; Tavazoie and Reid 2000). First, the center of each LGN receptive field is mapped to the location in the input sheet corresponding to the location of the LGN unit. This mapping ensures that the LGN will have the same two-dimensional topographic organization as the retina. Using the location of each center, the receptor weights are then calculated from the difference of two normalized Gaussians with widths σ_c (center) and σ_s (surround). More precisely, for ON-center cell (a, b) with center (x_o, y_o), the weight $\mu_{ab,xy}$ from each receptor (x, y) in the receptive field is given by the difference of two normalized Gaussians:

$$\mu_{ab,xy} = \frac{\exp(-\frac{(x-x_o)^2-(y-y_o)^2}{\sigma_c^2})}{\Sigma_{xy}[\exp(-\frac{(x-x_o)^2-(y-y_o)^2}{\sigma_c^2})]} - \frac{\exp(-\frac{(x-x_o)^2-(y-y_o)^2}{\sigma_s^2})}{\Sigma_{xy}[\exp(-\frac{(x-x_o)^2-(y-y_o)^2}{\sigma_s^2})]} \tag{4.1}$$

where σ_c determines the width of the central Gaussian and σ_s determines the width of the surround Gaussian. 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. Figure 4.2 shows examples of these ON and OFF receptive fields.

Note that even though the OFF cells have the same weights as ON cells (differing only by the sign), their activity will not be redundant. Each cell will be thresholded to have only positive activations, so ON and OFF cells will never be active at the same cortical location. This thresholding is described in more detail below, and models the fact that firing rates in biological systems cannot

37

be negative. The result is that OFF and OFF channels provide complementary information, both in the model and in the visual system. Separating the ON and OFF channels in this way makes it simpler to compare the model to experimental results.

4.1.3   Initial connections in the cortex

In contrast to the fixed connection weights in the LGN, all connections in cortical regions in HLIS-SOM are modifiable by neural activity. The connections are initially unselective, either Gaussian or random. As will be shown in chapter 5, the specific initial values of these weights have little effect on subsequent self-organization. To maintain consistency with previous publications, the simulations in this thesis generally use random afferent weights and Gaussian lateral weights. Random lateral weights can be used instead, as long as their radial extent is about the same as the extent of the Gaussians.

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, and also include connections to the neuron itself and to its neighbors. This overall center-surround pattern is crucial for self-organization, and approximates the lateral interactions present at high contrasts (section 2.1.1; Sirosh 1995). Long-range excitatory connections can also be included when simulating perceptual grouping and completion phenomena (Choe 2001; Choe and Miikkulainen 2000a, 2002, 1997, 1998, 1996). Figure 4.3 shows examples of the initial weight patterns for one V1 neuron in HLISSOM.

4.2   Activation

Self-organization of the connection weights is driven by a series of t_f input iterations, usually 5,000–20,000. Each iteration consists of presenting an input image, computing the corresponding activation patterns in each sheet, and modifying the weights. Before each input presentation, the activities of all units are initialized to zero, and a grayscale pattern is drawn on the input sheet. The activation of each area is then computed in turn.

Figure 4.4a shows an example input pattern consisting of oriented Gaussians, like those used in the RF-LISSOM and SOM models of orientation map development (Bednar 1997; Bednar and Miikkulainen 2000b; Obermayer, Ritter, and Schulten 1990; Sirosh 1995; Sirosh, Miikkulainen, and Bednar 1996). To generate such input patterns, the activity ξ for every retinal ganglion cell (x, y) is calculated according to the equation

$$
\xi_{xy} = \exp(\quad -\frac{((x - x_o)cos(\theta) - (y - y_o)sin(\theta))^2}{a^2}
$$
$$
-\frac{((x - x_o)sin(\theta) + (y - y_o)cos(\theta))^2}{b^2}) \tag{4.2}
$$

(*a*) V1 Afferent RF   (*b*) V1 Excitatory   (*c*) V1 Inhibitory
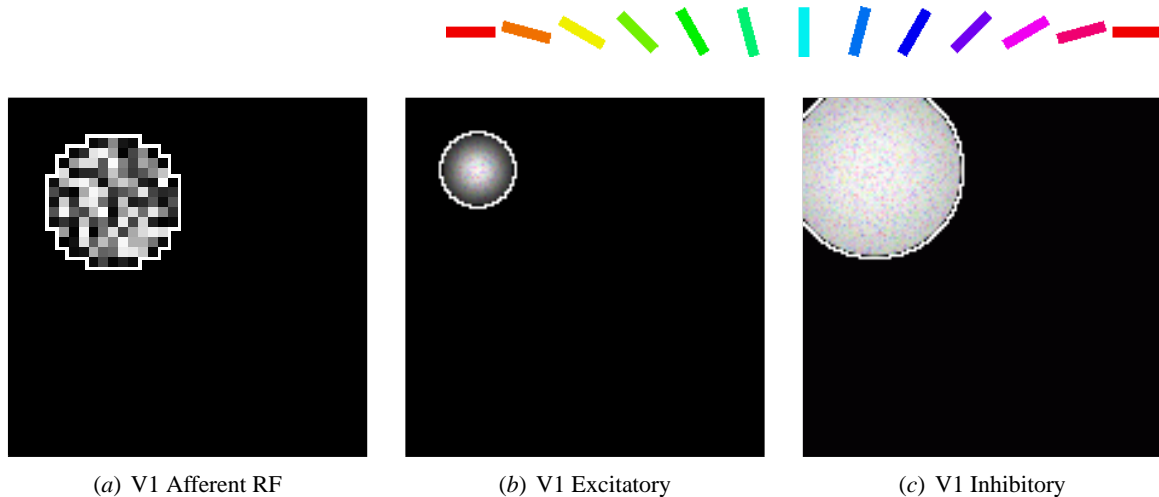
Figure 4.3: **Initial RFs and lateral weights (*color figure*).** These plots show the weights for one V1 neuron before self-organization. Each set of weights is outlined in white, and is plotted on the neural region to which the weights connect. This particular neuron is located at 28 units from the top and 28 units from the left of a 142×142 V1 that will later be used as an orientation map example. The initial afferent RFs (*a*) were random. The weights are plotted on the LGN ON cell sheet, with each weight shown as a brightness level from black to white. The weights to the ON and OFF LGN sheets were initially identical, so only the ON weights are shown. Plots (*b*) and (*c*) show the lateral excitatory weights of this neuron, plotted on V1. The brightness of each neuron in the plot indicates the strength of the connection to that neuron, from black to white. The connections were initially Gaussian here, but can also be random or any other roughly isotropic distribution. The color of each neuron within the lateral connection outlines represents the orientation preference of that neuron, according to the color key along the top. The strength of the color (its saturation) represents the degree of orientation selectivity. Because these random RFs have low selectivity for any particular orientation, the neurons appear nearly white. Later figures will show that selectivity and patchy connectivity will develop through self-organization.

where a^2 and b^2 specify the length along the major and minor axes of the Gaussian, (x_o, y_o) specifies its center, and θ its orientation. At each iteration, the x and y coordinates of the centers are each chosen randomly within the $R \times R$ retinal area, and the orientation θ is chosen randomly from the uniform distribution $0° \leq \theta < 180°$. When multiple patterns are used in the same iteration, the centers are constrained to be at least d_r units apart to avoid overlap. Other artificial patterns can be generated similarly, or natural images can be rendered directly on the input layer.

4.2.1 LGN activation

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 input units in their receptive fields (as in figure 4.4*b*). More precisely, the response η_{ab} of ON or OFF-center cell (a, b) is calculated from the weighted

(*a*) Input pattern	(*b*) LGN response	(*c*) 0: V1 afferent response	(*d*) 0: V1 settled response	(*e*) 10,000: V1 afferent response	(*f*) 10,000: V1 settled response
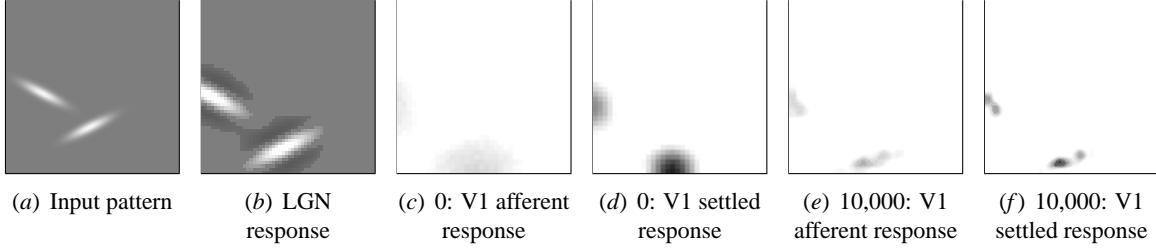
Figure 4.4: **Training pattern activation example.** At each self-organization iteration in HLISSOM, input patterns are drawn on the photoreceptor layer. For this example, multiple oriented Gaussians were drawn with random orientations and at random spatially separated locations on a 54×54 array of receptors (*a*). The 36×36 ON and OFF cell responses are plotted in (*b*) by subtracting the OFF cell responses from the ON. The LGN responds to edges and lines in the input, with high ON cell activations (white in *b*) where the input is brighter than the surround, and high OFF cell activations (black in *b*) where the input is darker than the surround. Before self-organization (i.e., iteration 0), the 142×142 V1 map responds broadly and unspecifically to the input patterns (*c*). The lateral connections focus the response into discrete activity "bubbles" *d*, and connections are then modified. After 10,000 input presentations and learning steps, the V1 response is focused and patchy, extending along the orientation of the stimulus (*e-f*). As will be clear in later figures, the response is patchy because neurons that prefer similar positions but different orientations do not respond.

sum of the retinal activations in its receptive field as

$$\eta_{ab} = \sigma \left(\gamma_A \sum_{\rho xy} \xi_{\rho xy} \mu_{ab,\rho xy} \right), \tag{4.3}$$

where ρ specifies the receptive field (either on the photoreceptors or the PGO), $\xi_{\rho xy}$ is the activation of cell (x, y) in the receptive field, $\mu_{ab,\rho xy}$ is the afferent weight from (x, y) to (a, b), γ_A is a constant afferent scaling factor, and σ is a piecewise linear approximation of the sigmoid activation function (figure 4.5):

$$\sigma(x) = \begin{cases} 0 & x \leq \delta \\ (x - \delta)/(\beta - \delta) & \delta < x < \beta \\ 1 & x \geq \beta \end{cases} \tag{4.4}$$

As in RF-LISSOM and other models, this approximation is used because it can be computed more quickly than a smooth sigmoid function, while retaining its key threshold and saturation properties. Changing γ_A in equation 4.3 by a factor m is mathematically equivalent to dividing β and δ by m. Even so, γ_A is treated as a separate parameter to make it simpler to use the same values of β and δ for different networks. The specific value of γ_A is set manually so that the LGN outputs approach 1.0 in the highest-contrast regions of typical input patterns. This allows each subsequent level to use similar parameter values in general, other than γ_A.

Because of its Difference-of-Gaussians RF, an LGN neuron will respond whenever the input pattern is a better match to the central portion of the RF than to the surrounding portion. The positive
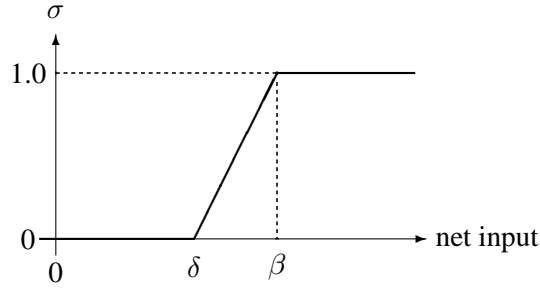
Figure 4.5: **The HLISSOM neuron activation function** σ**.** The neuron requires an input as large as the threshold δ before responding, and saturates at the ceiling β. The output activation values are limited to $[0, 1]$. The activation function is an easy-to-compute approximation of the sigmoid function.

and negative portions of the RF thus have a *push–pull* effect (Hirsch, Alonso, Reid, and Martinez 1998a; Hubel and Wiesel 1962). That is, even if an input pattern activates the ON portion of the LGN RF, the neuron will not fire unless the OFF portion is *not* activated. This balance ensures that the neurons will remain selective for edges over a wide range of brightness levels. Section 4.5 will show that this push–pull effect is crucial when using natural images in the model. Overall, the LGN neurons respond to image contrast, subject to the minimum and maximum activity values enforced by the sigmoid.

4.2.2 Cortical activation

The cortical activation is similar to the LGN, but extended to support self-organization and to include lateral interactions. The total activation of each neuron is computed from both afferent and lateral responses, which will be discussed separately here. First, the afferent response ζ_{ij} of neuron (i, j) in the $N \times N$ V1 sheet is calculated like the retinal sum in equation 4.3, with an additional divisive (shunting) normalization:

$$\zeta_{ij} = \frac{\gamma_A \sum_{\rho ab} \xi_{\rho ab} \mu_{ij,\rho ab}}{1 + \gamma_N \sum_{\rho ab} \xi_{\rho ab}}, \tag{4.5}$$

where $\xi_{\rho ab}$ is the activation of unit (a, b) in receptive field ρ (on either the ON channel or the OFF channel[1]), and $\mu_{ij,\rho ab}$ is the corresponding afferent weight.

The γ_N normalization is an approximation of push–pull afferent weights that allows all afferent weights to be excitatory, so that each can use simple Hebbian learning as in RF-LISSOM. Mathematically, it represents dividing the response from the excitatory weights by the response from a uniform disc of inhibitory weights over the RF. Similar to the LGN activation, the net result is that activity in input locations where there are weak or no afferent weights suppresses the overall

---

[1]Afferent 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. For simplicity and computational efficiency, only a single channel of each type (ON or OFF) was used in the simulations in this thesis.

activation. Activity in locations to which the neuron has a strong connection increases the neuron's activation. As shown below, the push–pull effects in the LGN are the most crucial addition to RF-LISSOM, but this additional normalization in the cortex helps to preserve orientation selectivity with the wide ranges of contrast found across natural images. Simulations with artificial input patterns or covering only a small area of the visual field, such as those in most models and in RF-LISSOM, can omit this normalization and leave γ_N at zero.

HLISSOM simulations with large natural images generally use an initial normalization strength γ_N of zero, but gradually increase it over training as neurons become more selective. To prevent the net responses from decreasing, the scaling factor γ_A is set manually to compensate for each change to γ_N. The goal is to ensure that the afferent response ζ will continue to have values in the full range 0 to 1.0 for typical input patterns, regardless of the γ_N value.

The FSA computes its afferent response just as V1 does, except that ρ in equation 4.5 is an RF on V1, instead of an ON or OFF channel of the LGN. Whether in V1 or the FSA, the afferent response forms the initial response of the neuron, after being passed through the sigmoid activation function:

$$\eta_{ij}(0) = \sigma\left(\zeta_{ij}\right), \tag{4.6}$$

where σ is as shown in equation 4.4. After the initial response, lateral interaction sharpens and strengthens the cortical activity over a very short time scale. At each of these subsequent time steps, the neuron combines the afferent response ζ with lateral excitation and inhibition:

$$\eta_{ij}(t) = \sigma\left(\zeta_{ij} + \gamma_E \sum_{kl} E_{ij,kl}\eta_{kl}(t-1) - \gamma_I \sum_{kl} I_{ij,kl}\eta_{kl}(t-1)\right), \tag{4.7}$$

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, which determine how easily the neuron reaches full activation.

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 4.7, converges into stable focused patches of activity, or activity bubbles (as in figure 4.4d). Mathematically, the settling increases the sparsity of the response, which Field (1994) proposed is a fundamental goal of cortical processing (cf. Sirosh 1995). The practical outcome is that neurons will develop similar properties, as seen in the cortex, because settling ensures that nearby neurons have similar patterns of activity.

4.3 Learning

The learning process is driven by the settled activity patterns. Modifying the weights of each cortical neuron gradually makes them come to represent the correlation patterns seen in the input. Afferent and lateral weights in both V1 and the FSA adapt according to the same biologically motivated mechanism: the Hebb rule (Hebb 1949), normalized so that the sum of the weights is constant:

$$w_{ij,mn}(t + \delta t) = \frac{w_{ij,mn}(t) + \alpha \eta_{ij} X_{mn}}{\sum_{mn} [w_{ij,mn}(t) + \alpha \eta_{ij} X_{mn}]}, \tag{4.8}$$

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. The normalization prevents the weight values from increasing without bound as new patterns are learned, and is an abstraction of the neuronal regulatory processes reviewed by Turrigiano (1999).

At long distances, very few neurons have correlated activity and therefore most long-range connections eventually become weak. The weak connections are eliminated periodically to increase the strength to the remaining connections, resulting in patchy lateral connectivity similar to that observed in the visual cortex.

The radius of the lateral excitatory interactions usually starts out large, but as self-organization progresses, it is decreased until it covers only the nearest neighbors. Such a decrease is a convenient way of varying the balance between excitation and inhibition to ensure that both global topographic order and well-tuned receptive fields develop (Miikkulainen et al. 1997; Sirosh 1995). With sufficient initial order, a more biologically realistic fixed-sized lateral excitatory radius may be used instead.

The following section will show that the architecture, activation procedure, and learning algorithm described here leads to realistic orientation maps, receptive fields, and lateral connections. Because the only explicit representation of orientation is in the input patterns, HLISSOM is also a general model for cortical function, and later chapters will show that it can develop face-selective responses as well as orientation maps. HLISSOM thus represents a unified model of processing in multiple areas of the cortex.

4.4 Orientation map example

This section presents a detailed example of self-organization in HLISSOM. The starting point was a network with the initial weights set as shown in figure 4.3; these parameters were chosen based on earlier RF-LISSOM simulations (Bednar and Miikkulainen 2000b). To this network was presented a series of 10,000 random oriented Gaussian input patterns like the one in figure 4.4.

(*a*) V1 Afferent RF   (*b*) V1 Excitatory   (*c*) V1 Inhibitory
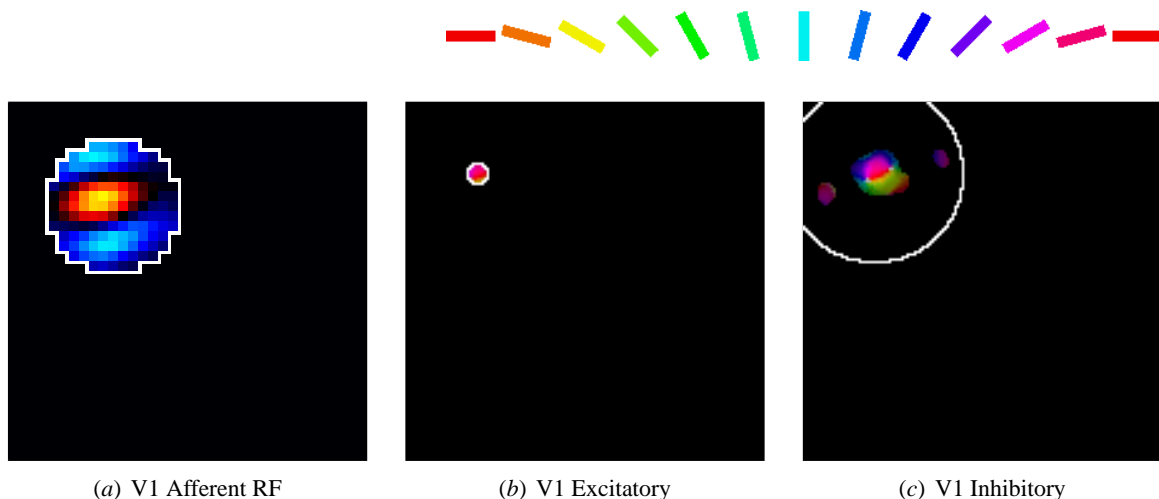
Figure 4.6: **Self-organized receptive fields and lateral weights (*color figure*).** These plots show the weights for the neuron in figure 4.3 after self-organization. In (*a*), the OFF weights were subtracted from the ON weights. Net ON regions are colored red, orange, and yellow, and net OFF regions are colored blue and cyan. This neuron prefers a line oriented at 10°, i.e. nearly horizontal, and will respond most strongly to a white line overlapping the red/yellow portion of its RF, surrounded by black areas overlapping the blue portions of its RF. Other neurons developed similar RFs with different preferred orientations. This type of RF structure is commonly seen in V1 neurons of laboratory animals (Hubel and Wiesel 1962, 1968). Plot (*b*) shows the lateral excitatory weights of this neuron. All connected neurons are strongly colored red, magenta, or purple, i.e. orientations similar to the orientation preference of this neuron. Plot (*c*) shows the lateral inhibitory weights. After self-organization and connection pruning, only connections to neurons with similar orientations remain, and they are extended along the preferred orientation of the neuron. The connection pattern is patchy, because connections to neurons with opposite preferences are weaker or have been pruned away. These patchy, orientation-specific connection patterns are also seen in laboratory animals (figure 2.5; Bosking et al. 1997; Sincich and Blasdel 2001). Thus the HLISSOM model develops realistic receptive fields and intracortical connectivity.

After each presentation, HLISSOM adjusted the weights to each cortical neuron. This incremental learning process required 2.2 hours of computer time on a 1 GHz Athlon-processor machine, and required 457 megabytes of memory for the connections. Parameters for this simulation are listed in section A.3.2.

Figure 4.6 shows that through activity-dependent self-organization, the V1 neurons learned realistic multi-lobed oriented RFs and lateral connections (Bosking et al. 1997; Hubel and Wiesel 1965, 1968; Sincich and Blasdel 2001). The lateral connections extend along the orientation of the neuron, and are patchy because they target orientations similar to this neuron's preferred orientation. The orientation preferences of this and the other neurons in the network form the orientation map shown in figure 4.7. Even with such abstract inputs, the map is a good match to those measured in adult animals (e.g. compare figure 4.7 with figure 2.4 on page 10). This simulation demonstrates that HLISSOM can develop realistic cortical structures.
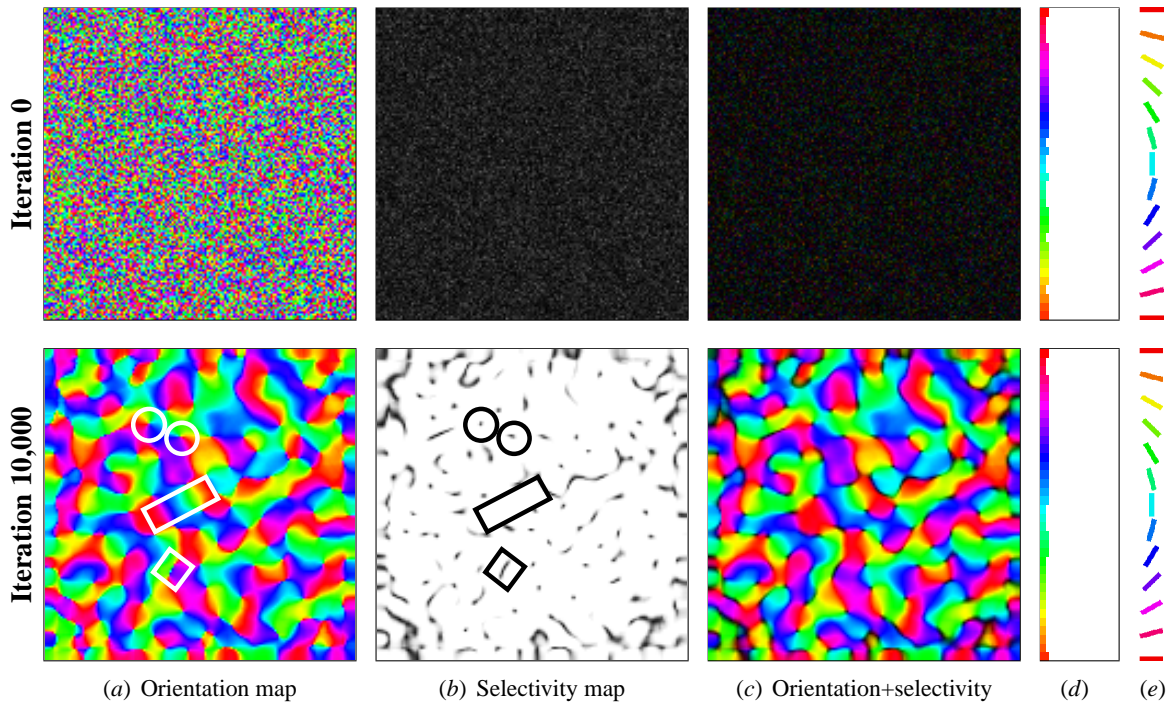
Figure 4.7: **Map trained with oriented Gaussians (*color figure*).** The orientation preference of each neuron in the map from figure 4.4 was computed before (top row, iteration 0) and after self-organization (bottom row, iteration 10,000). Each neuron in the map is colored according to the orientation it prefers, using color key (*e*). (*a*) The preferences are initially random (top). Through self-organization, the network developed a smoothly varying orientation map (bottom). The map contains features found in maps from experimental animals (figure 2.4 on page 10), such as pinwheels (two are circled in white in *a* and black in *b*), linear zones (one is marked with a long white or black rectangle), and fractures (one between green and blue/purple is marked with a white or black square). (*b*) Before self-organization, the selectivity of each neuron for its (random) preferred orientation is very low (black in *b*, top). In contrast, nearly all of the self-organized neurons are highly selective for orientation (white in *b*, bottom). (*c*) Overlaying the orientation and selectivity plots shows that regions of lower selectivity in the self-organized map tend to occur near pinwheel centers and along fractures. Histograms of the number of neurons preferring each orientation are shown in (*d*), and are essentially flat because the initial weight patterns were unbiased and subsequent training inputs represented all orientations equally. These plots show that HLISSOM can develop realistic orientation maps through self-organization based on abstract input patterns.

45

Even though this simulation is primarily an example, it contains two important advances over the earlier RF-LISSOM simulations with oriented Gaussian inputs (Sirosh 1995; Sirosh et al. 1996). First, HLISSOM explains how multi-lobed RFs can develop. RF-LISSOM contained only ON-type connections, and did not model the LGN. Second, HLISSOM neurons have random initial orientation preferences, and the orientation histogram is flat. Because of square connection radii and receptive fields clipped by region borders, most neurons in RF-LISSOM were biased in orientation at the start of training. Such unintentional architectural biases would make it difficult to study how particular types of input patterns can influence development, as in later chapters. The next section will show that the multi-lobed receptive fields introduced in HLISSOM are crucial for modeling natural images, as required by most of the later simulations.

4.5   Role of ON and OFF cells

One of the most important differences between HLISSOM and the previous RF-LISSOM architecture is that HLISSOM includes models of the LGN ON and OFF cells. How does making the model more realistic in this way affect the development of maps, and the functional behavior? Interestingly, when the training patterns are matched exactly, both models develop very similar maps, with many of the orientation blobs in the same relative locations (figure 4.8). This somewhat surprising result explains why models with and without ON and OFF cells have both been able to develop realistic orientation maps. As will be shown in section 5.2, the map shape is determined by the locations of V1 activity blobs evoked by the stream of inputs. With the same stream of oriented patterns, the V1 activity patterns are the same with or without ON and OFF channels. Thus similar final map shapes develop.

Even the orientations of the RFs of individual neurons are usually very similar between the two maps (figure 4.8*f*). Yet the shapes of the RFs are very different. They differ because RF shapes are determined by both the V1 and LGN activities, and the LGN activities for the OFF channel differ greatly from those for the ON channel.

In part because of the different RF shapes, these two models can have very different functional behavior. In particular, figure 4.9 shows that the ON and OFF cells are necessary for neurons to preserve orientation selectivity when given large patterns of input activity or nonzero mean levels of illumination. Natural scenes and retinal waves contain many such features, and thus extending RF-LISSOM with ON and OFF cells was crucial for the experiments in this thesis.

4.6   Conclusion

The HLISSOM model includes the LGN (both ON and OFF channels), V1, and a higher level face-selective region. The model focuses on explaining capabilities that develop at V1 and higher cortical levels, such as orientation and face processing. Connection weights in the LGN were chosen to be
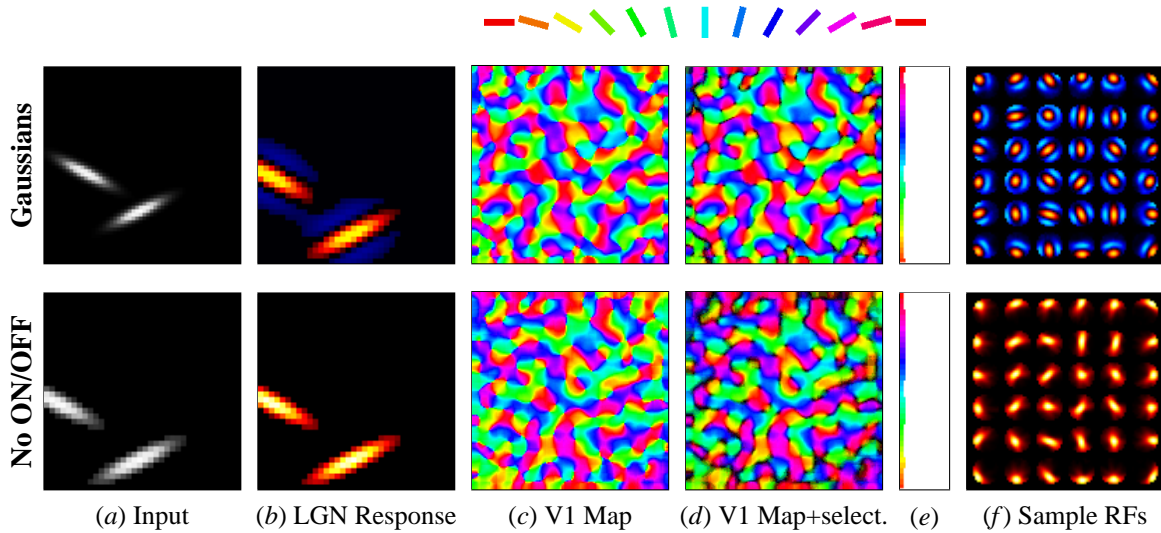
Figure 4.8: **Matching maps develop with or without the ON and OFF channels (*color figure*).** Each row shows the results for a different network trained with the same stream of Gaussian inputs. The networks had similar architectures except that for the network in the bottom row, the ON and OFF LGN units were replaced with a single set of direct connections to the photoreceptor layer. Such an architecture matches previous models without the LGN (e.g. Bednar and Miikkulainen 2000b; Sirosh, Miikkulainen, and Bednar 1996). Column (*b*) shows the LGN response to the sample grayscale input pattern from (*a*); regions with high ON region response are colored red and yellow, and those with high OFF response are colored blue and cyan. Column (*c*) shows the self-organized orientation map for each network. Very similar maps develop with or without the LGN, because the stream of V1 activations caused by the inputs was similar. Column (*d*) shows the map masked by the orientation selectivity. Unselective neurons are colored black, and brighter colors represent more selective neurons. Without the ON and OFF cells, fewer neurons are highly selective, but the improvement in selectivity is relatively small. Column (*e*) shows the histogram of each orientation map, which is nearly flat for both networks. Column (*f*) shows sample RFs for 36 neurons in each map, arranged according to the position of the neuron in the cortex. In the top set of RFs, regions with high ON weights are shown in red and yellow, and those with high OFF weights are in blue and cyan. In the bottom set, regions with high weight values are shown in red and yellow. The orientations of the corresponding RFs in each map are often similar. However, when the ON and OFF cells are included the RFs become more strongly oriented, and have multiple ON and OFF lobes. These results show that similar V1 maps can be developed without modeling the LGN and multi-lobed RFs, but figure 4.9 will show that the functional behavior of these similar-seeming maps is very different.

a good match to the known properties of the LGN in animals, and then connections at the higher levels develop via incremental Hebbian learning of bitmap images. The level of detail in the model was chosen so that the model could be validated against the experimental data available in each domain, and so that it can provide specific predictions for future experiments. HLISSOM develops realistic receptive fields, lateral weights, and map patterns, all of which have been validated against experimental data.
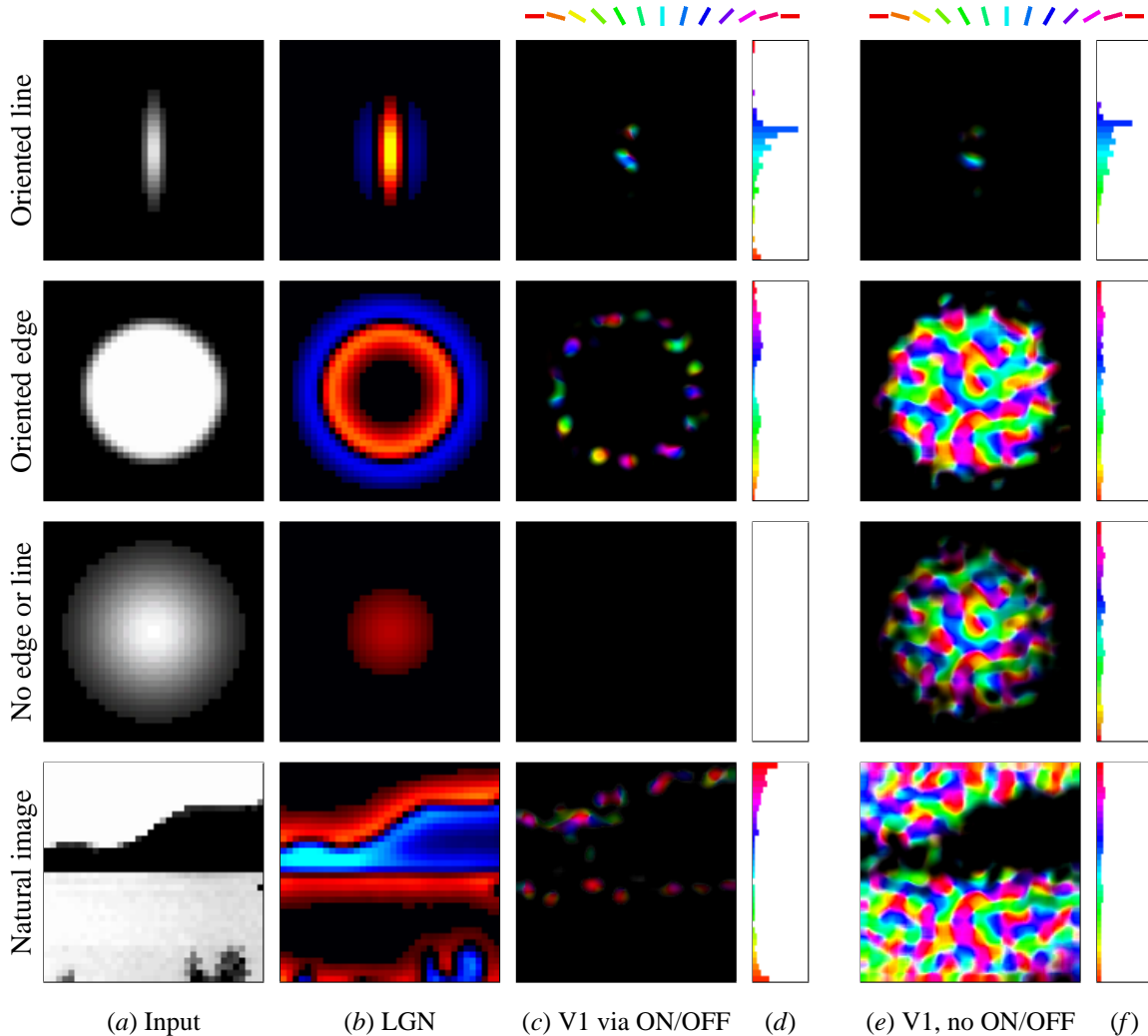
Figure 4.9: **The ON and OFF channels preserve orientation selectivity (*color figure*).** Each row shows the response to the pattern in column (*a*). Columns (*b-d*) show the response of the **Gaussians** network from figure 4.8, and (*e-f*) show the response of the **No ON/OFF** network, where the ON and OFF channels of the LGN are omitted. Both networks respond very similarly to an oriented Gaussian input like those used during training (top row), which is why very similar orientation maps developed in figure 4.8. For both networks, only neurons with orientation preferences matching the input line's orientation respond (histograms *d,f*, top row). However, the networks behave very differently for other types of inputs (remaining rows). The ON and OFF channels filter out smooth, gradual changes in brightness, which ensures that V1 responds only to oriented patterns and sharp edges (*c-d*). Without the LGN, any large enough input pattern causes a correspondingly large and unselective response in V1, activating nearly all neurons regardless of their orientation preferences (*e-f*). Similarly, a constant mean level of illumination will activate *all* of the V1 neurons in the model without ON and OFF cells, regardless of their orientation preferences, while it will activate none of the neurons in the full HLISSOM model (row *No Edge or Line*). Thus the ON and OFF channels are crucial for preserving orientation selectivity for large objects, large, gradual changes in brightness, and non-zero mean levels of illumination, all of which are common in natural images.

48

Chapter 5

Scaling HLISSOM simulations

In the previous chapter I introduced HLISSOM and showed how it can develop realistic maps of orientation preferences. So that the simulations would be practical to run, the maps covered only about a 25mm^2 area out of the full 2500mm^2 area of human V1 (Wandell 1995). Yet despite the small area, each simulation required several hours of computer time and hundreds of megabytes of memory. Experiments with faces will require a much larger area of V1 and the visual field, because most tests with newborns use head-sized face patterns presented at a very close range. Such patterns fill up most of the newborn's visual field. Other important phenomena also require large networks, including long-range contour integration, object recognition, and optic flow.

      In this chapter I will introduce a set of scaling equations that I developed to make such simulations practical. Given a small-scale simulation, the equations make it possible to calculate the parameter settings necessary to perform an equivalent but larger scale simulation. I will show that the original and scaled simulations have similar function and map-level organization; the larger map just has more detail or a greater area, and takes more time and memory to simulate. Although the equations are included here primarily because they will be used in later chapters, they are also an independent contribution that can be used with other incremental Hebbian models with architectures similar to RF-LISSOM or HLISSOM. As will be discussed below, the equations can also be applied to biological systems, calculating how these systems differ in structure and thus how their functions are likely to differ. Finally, the equations allow parameters for a model to be calculated from measurements in animals of a particular species.

5.1   Background

Any particular HLISSOM or RF-LISSOM simulation models a fixed area of the visual field at a fixed retinal density (receptors per visual field area) and a fixed cortical density (neurons per unit of cortical area). Modeling new phenomena often requires using a different area or density. Changing the area corresponds to modeling a larger portion of the visual space, e.g. a larger part of V1 and of

the eye. Changing the density corresponds to modeling a given area at a finer resolution (of either cortical neurons or retinal receptors), as well as modeling a species, individual, or brain area that devotes more neurons or ganglia to the representation of a fixed amount of visual space.

Varying the area or density over a wide range is difficult in a complex nonlinear system like HLISSOM. Parameter settings that work for one size will need to be very different to work properly with other sizes. Furthermore, it is not always clear which parameters need to be adjusted at each size. To eliminate the search for appropriate parameters, I derived equations to compute the values needed for each type of transformation. The equations treat any specific cortical network as a finite approximation to a continuous map composed of an infinite number of units. Similar continuous networks have been studied theoretically by Amari (1980) and Roque Da Silva Filho (1992). Under such an assumption, networks of different sizes represent coarser or denser approximations to the continuous map, and any given approximation can be transformed into another by (conceptually) reconstructing the continuous map and then resampling it. Given an existing retina and cortex of some fixed size, the scaling equations provide the parameter values needed for a smaller or larger retina and cortex to form a functionally equivalent map.

The following subsections will present the scaling equations for area, retinal ganglion density, and cortical density, along with results from simulations of each type. The equations are all linear, so they can also be applied together to change both area and density simultaneously. In each case, what the equations provide is a set of parameters for a new simulation that, when run, will develop similar results to the existing simulation. The equations are first derived on theoretical grounds, and then I verify experimentally that no other parameters need to be scaled, i.e. that these particular equations are sufficient.

5.2 Prerequisite: Insensitivity to initial conditions

One desirable property of a scaling algorithm is that it results in nearly identical final maps even some parts of the network are changed, such as the retina size. Therefore, it is crucial that the map organization not depend on the random initial weights, since those will vary between networks of different sizes. In this section I will show that the HLISSOM algorithm has exactly this property.

In HLISSOM, there are two types of variability between runs with different random numbers: the random order, location, and position of the individual input patterns at each iteration, and the random initial values of the connection weights. Figure 5.1 demonstrates that the orientation map shape does not depend on the random initial values of the weights, as long as the initial weights are drawn from the same random distribution. This result is surprising and significant, because in many commonly used neural networks, the initial weights do play an important role (see e.g. Kolen and Pollack 1990).

Instead, the self-organized orientation map pattern in HLISSOM depends crucially on the stream of inputs, with two streams giving different orientation map patterns even if the streams are

50
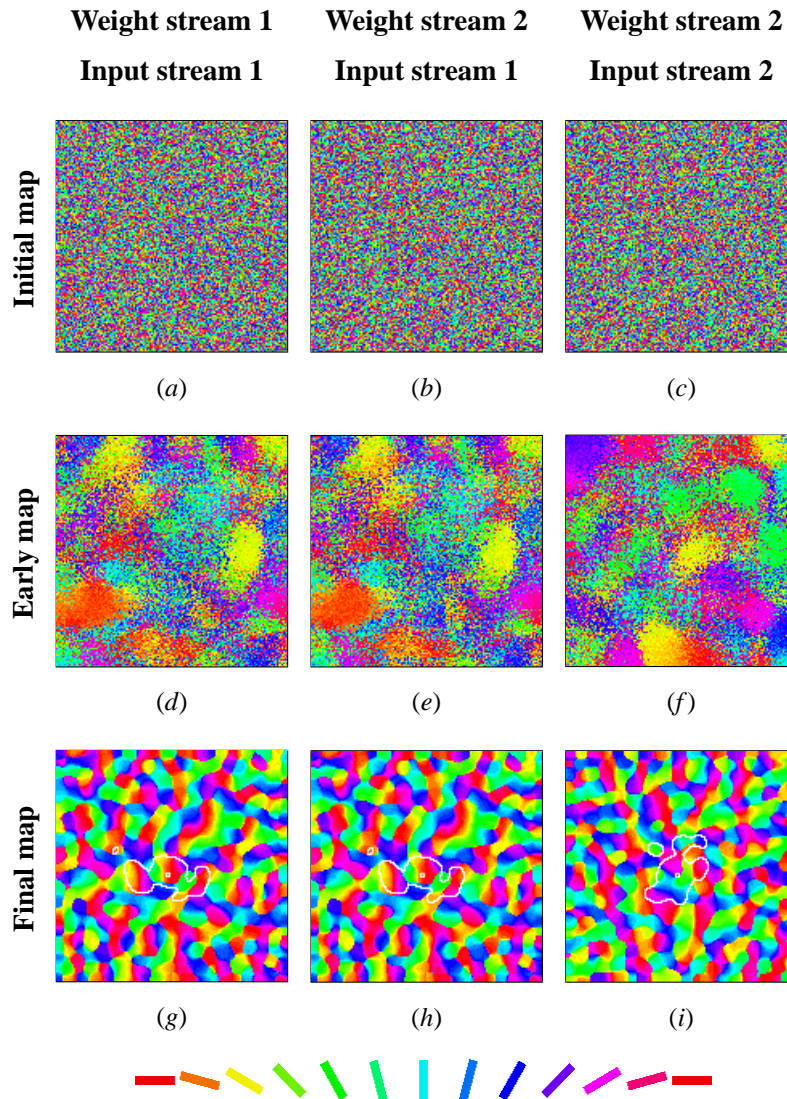
Figure 5.1: **Input stream determines map pattern in HLISSOM (*color figure*).** This figure shows that the self-organized orientation map patterns (e.g. in figure 4.7) do not depend on the random initial values of the weights. They are instead driven by the stream of input patterns presented during training. Using a different stream of random numbers for the weights results in different initial orientation maps (*a* and *b*), but has almost no effect on the final self-organized maps (compare *g* to *h*). In (*g-i*), the lateral inhibitory connections of one sample neuron are outlined in white, and are not affected by changing the weight stream. The final result is the same because lateral excitation smooths out differences in the initial weight values, and leads to similar large-scale patterns of activation at each iteration. (Compare maps *d* and *e* measured at iteration 100; the same large-scale features are emerging in both maps despite locally different patterns of noise caused by the different initial weights.) In contrast, changing the input stream produces very different early and final map patterns (compare *e* to *f* and *h* to *i*), even when the initial weight patterns (and therefore the initial orientation maps) are identical (*b* and *c*). Thus the input patterns are the crucial source of variation, not the initial weights.

drawn from the same distribution. The overall properties of the maps are very similar (number of pinwheel centers, distance between them, etc.), but different input streams lead to different arrangements of blobs and pinwheel centers. In animals, maps are also similar between members of the same species, but differ in the specific arrangements of blobs and pinwheel centers (Blasdel 1992b). Thus the HLISSOM model predicts that the specific orientation map pattern in the adult animal depends primarily on the order and type of activity seen by the cortex in early development, and not on the details of the initial connectivity.

HLISSOM is insensitive to initial weights because of three features common to most other incremental Hebbian models: (1) the scalar product input response function, (2) lateral excitation between neurons, and (3) the initial period with high learning rates. First, a neuron's initial response to an input pattern is determined by the sum of the product of each retinal receptor with its corresponding weight value (equation 4.6). For smoothly varying input patterns and large enough afferent RF size, this sum will have a very similar value regardless of the specific values of the weights to each receptor. Thus, until the weights have self-organized into a smooth, spatially non-uniform distribution, the input response of each neuron will be largely insensitive to the specific weight values. Second, settling due to lateral excitation (equation 4.7) causes nearby neurons to have similar final activity levels, which further reduces the contribution of each random afferent weight value. Third, Hebbian learning depends on the final settled activity levels resulting from an input (equation 4.8), and with a high enough learning rate, the initial weight values are soon overwritten by the responses to the input patterns. As is clear in figure 5.1*d,e*, the large-scale map features develop similarly even before the initial weight values have been overcome. Thus the Hebbian process of self-organization is driven by the input patterns, not the initial weights.

The net result is that as long as the initial weights are generated from the same distribution, their precise values do not significantly affect map organization. Similar invariance to the initial weights should be found in other Hebbian models that compute the scalar product of the input and a weight vector, particularly if they include lateral excitation and use a high learning rate in the beginning of self-organization. If a model does not have such invariance, scaling equations can still generate functionally equivalent maps, but they will not be visually identical like the density-scaled maps presented below.

5.3   Scaling equations

The following subsections will present the scaling equations for area, retinal ganglion density, and cortical density, along with results from simulations of each type.

5.3.1   Scaling the area

In this section I will consider the simplest case, changing the area of the visual space simulated. Scaling up the area allows a model to be developed quickly with a small area, then enlarged to

(*a*) Original cortex   (*b*) Cortex area scaled   (*c*) Original retina   (*d*) Retina area scaled

$N = 54$ $N = 4N_o = 4 \cdot 54 = 216$ $R = 24$ $R = 4R_o = 4 \cdot 24 = 96$

0.4 hours, 8MB   9 hours, 148MB
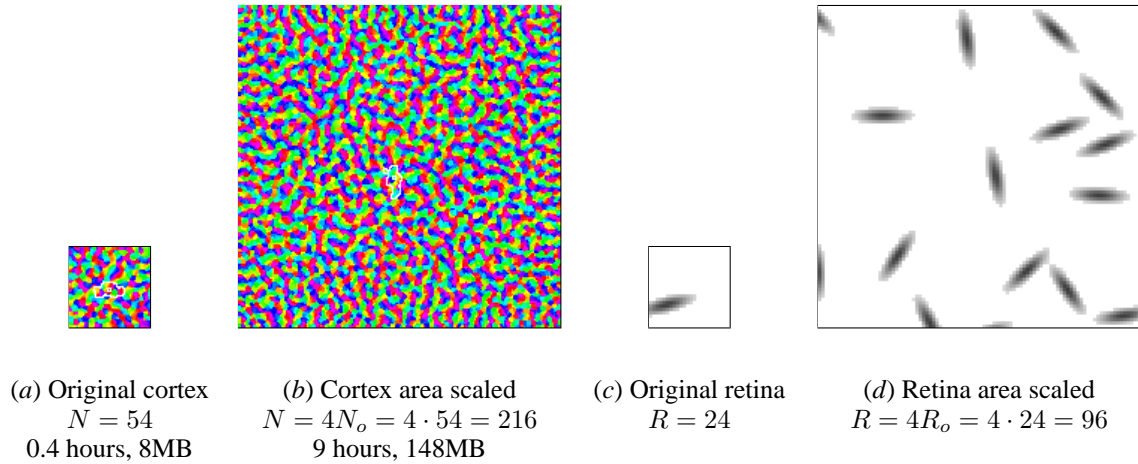
Figure 5.2: **Scaling the total area (*color figure*).** Equations 5.1 can be used to scale the total area simulated, in order to simulate a larger portion of the cortex, retina, and visual field. For large N the number of connections and the simulation time scale approximately linearly with the area, and thus a network four times as wide and four times as tall (above) takes about sixteen times the memory and sixteen times as long to simulate. For discrete input patterns like these oriented Gaussians, larger areas require more input patterns to keep the total learning per neuron and per iteration constant. Because the inputs are generated randomly across the active surface of the retina, each map sees an entirely different stream of inputs, and so the final map patterns always differ when the area differs. The area scaling equations are most useful for testing a model with a small area and then scaling up to eliminate border effects and to simulate the full area of a corresponding biological preparation.

eliminate border effects and to simulate the full area from a biological experiment. To change the area, both the cortex width N and the retina width R must be scaled by the same proportion k relative to their initial values N_o and R_o. (The ON and OFF channels of the LGN change just as the retina does, so their scaling equations are omitted below.)

What is perhaps less obvious is that, to ensure that the resulting network has the same amount of learning per neuron per iteration, the average activity per input receptor needs to remain constant. Otherwise a larger network would need to train for more iterations. Even if such longer learning were not computationally expensive, changing the total number of iterations in HLISSOM has non-linear effects on the outcome, because of the non-linear thresholding performed at each iteration (equation 4.3). Thus when using discrete input patterns (rather than natural images), the average number of patterns $\bar{\imath}$ per iteration must be scaled with the retinal area. Consequently, the equations for scaling the area by a factor k are:

$$N = kN_o, \quad R = kR_o, \quad \bar{\imath} = k^2\bar{\imath}_o \tag{5.1}$$

See figure 5.2 for an example of scaling the retinal and cortical area.

5.3.2 Scaling retinal density

In this section I consider changing the number of retinal units per degree of visual field (the retinal density), e.g. to model species with larger eyes, or parts of the eye that have more ganglia per unit area. This type of change also allows the cortical magnification factor, i.e., the ratio $N{:}R$ between the V1 and retina densities N and R, to be matched to the values measured in a particular species. I will show that scaling equations allow R to be increased to any desired value, but that in most cases (especially when modeling newborns) a low density suffices.

The parameter changes required to change density are a bit more complicated than those for area. The key to making such changes without disrupting how maps develop is to keep the visual field area processed by each neuron the same. For this to be possible, the ratio between the afferent connection radius and R must be constant, i.e. r_A must scale with R.

But when the connection radius increases, the total number of afferent connections per neuron also increases dramatically. Because the learning rate α_A specifies the amount of change per connection and not per neuron (equation 4.8), the learning rate must be scaled down to compensate so that the average total weight change per neuron per iteration remains constant. Otherwise, a given input pattern would cause more total change in the weight vectors of each neuron in the scaled network than in the original. Such a difference would significantly change how the two networks develop. So the afferent learning rate α_A must scale inversely with the number of afferent connections to each neuron, which in the continuous plane corresponds to the area enclosed by the afferent radius. Thus, α_A scales by the ratio $\frac{r_{\mathrm{A}_o}{}^2}{r_\mathrm{A}{}^2}$.

To keep the average activity per iteration constant, the size of the input features must also scale with R, keeping the ratio between the feature width and R constant. For e.g. Gaussian inputs this means keeping the ratio between the Gaussian width σ and R constant; other input types scale similarly. Thus the retinal density scaling equations are:

$$r_\mathrm{A} = \frac{R}{R_o} r_{\mathrm{A}_o} \quad \alpha_\mathrm{A} = \frac{r_{\mathrm{A}_o}{}^2}{r_\mathrm{A}{}^2} \alpha_{\mathrm{A}_o}, \quad \sigma_x = \frac{R}{R_o} \sigma_{x_o},$$
$$\sigma_y = \frac{R}{R_o} \sigma_{y_o} \tag{5.2}$$

Figure 5.3 shows how this set of equations can be used to generate functionally equivalent orientation maps using different retinal receptor densities.

This result may seem surprising, because several authors have argued that the cortical magnification factor $N{:}R$ is a crucial variable for map formation (e.g. Sirosh 1995, p. 138, 164). As is clear in figure 5.3, nearly identical maps can develop regardless of R (and thus $N{:}R$). Instead, the crucial parameters are those scaled by the retinal density equations, specifically r_A and σ. The value of R is unimportant as long as it is large enough to faithfully represent input patterns of width σ, i.e., as long as the minimum sampling limits imposed by the Nyquist theorem are not reached (see e.g. Cover and Thomas 1991). Strictly speaking, even σ is important only if it is larger than the center size of the LGN cells, i.e., only if it is large enough to be passed through to V1. As a result, when using natural images or other stimuli with high-frequency information, the LGN cell

(a) Original retina
$R = 24$, $\sigma_x = 7$, $\sigma_y = 1.5$

(c) Retina scaled by 2.0
$R = 24 \cdot 2 = 48$, $\sigma_x = 14$, $\sigma_y = 3$

(e) (e) Retina scaled by 3.0
$R = 24 \cdot 3 = 72$, $\sigma_x = 21$, $\sigma_y = 4.5$

(b) V1 map for $R = 24$

(d) V1 map for $R = 48$

(f) V1 map for $R = 72$
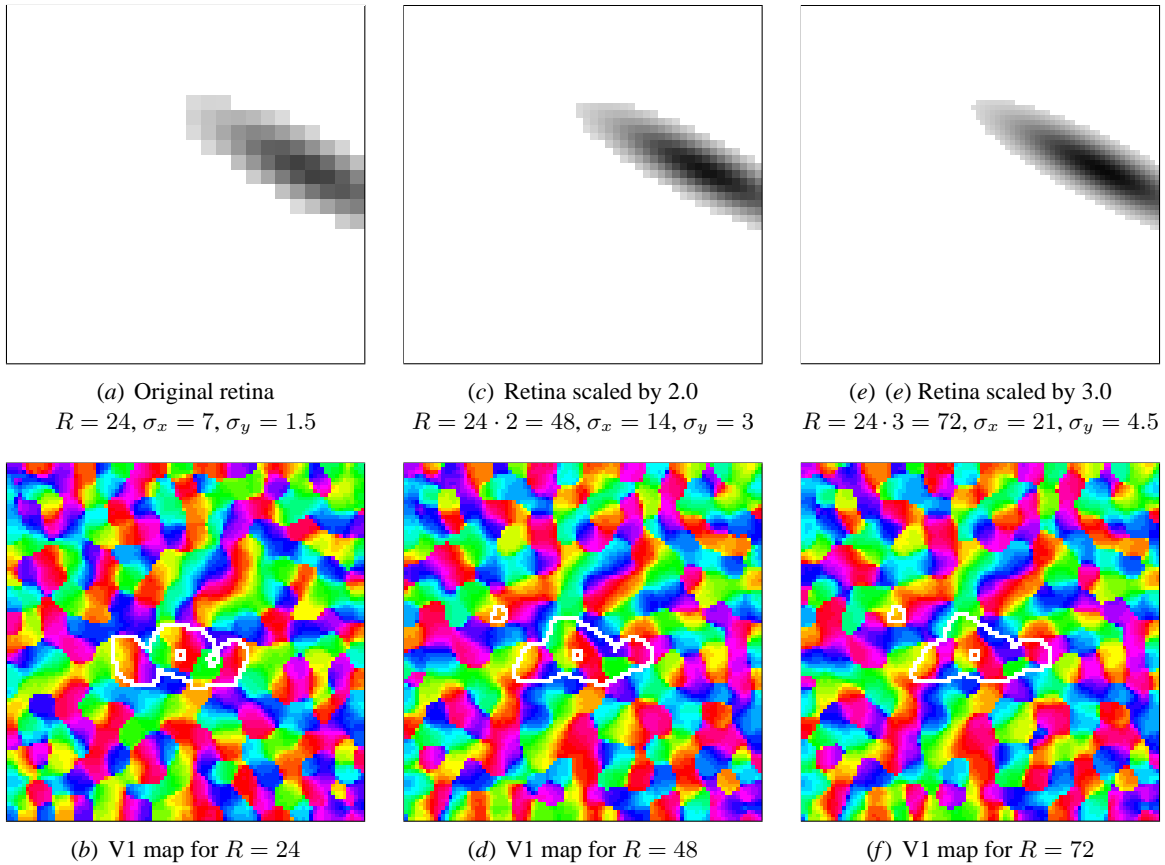
Figure 5.3: **Scaling retinal density (*color figure*).** Equations (5.2) can be used to scale the density of retinal receptors simulated per unit visual area, while keeping the area and the cortex density constant. This is useful for matching the cortical magnification factor in a simulation, i.e., the ratio between the V1 and retina densities, to the values from a particular species. Here each column shows an HLISSOM orientation map from one of three matched 96×96 networks that have retinas of different densities. The parameters for each network were calculated using equations (5.2), and then each network was trained independently on the same random stream of input patterns. The size of the input pattern in retinal units increases as the retinal density is increased, but its size as a proportion of the retina remains constant. All of the resulting maps are similar as long as R is large enough to represent the input faithfully, with almost no change above $R = 48$. Thus a low value can be used for R in practice.

RF size may instead be the limiting factor for map development. In practice, these results show that a modeler can simply use the smallest R that faithfully samples the input patterns, thereby saving computation time without significantly affecting map development.

55

(*a*) 36×36	(*b*) 48×48	(*c*) 72×72	(*d*) 96×96	(*e*) 144×144
0.17 hours, 2.0MB	0.32 hours, 5.2MB	0.77 hours, 22MB	1.73 hours, 65MB	5.13 hours, 317MB
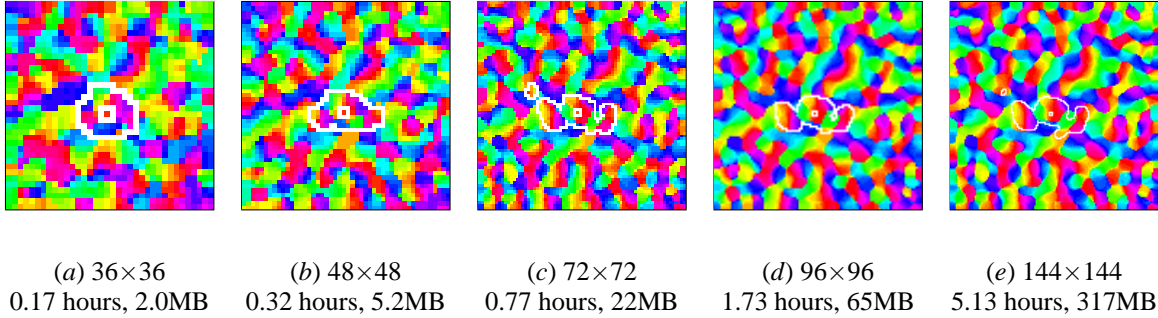
Figure 5.4: **Scaling the cortical density (*color figure*).** Five HLISSOM orientation maps from networks of different sizes are shown; the parameters for each network were calculated using equations (5.3), and then each network was trained independently on the same random stream of input patterns. The size of the network in the number of connections ranged from 2×10^6 to 3×10^8 (2 megabytes to 317 megabytes of memory), and the simulation time ranged from ten minutes to five hours on the same machine, a single-processor 600MHz Pentium III. Simulations of 2048-megabyte 192×192 networks on the massively parallel Cray T3E supercomputer perform similarly. Despite this wide range of simulation scales, the final organized maps are both qualitatively and quantitatively similar, as long as their size is above a certain minimum (here about 64×64). Larger networks take significantly more memory and simulation time, but offer greater detail and will allow other dimensions such as ocular dominance and direction to be simulated in the same map.

5.3.3 Scaling cortical neuron density

Scaling the cortical density allows a small density to be used for most simulations, while making it possible to scale up to the full density of the cortex to see details in the orientation maps. Scaling up is also important when simulating joint maps, such as combined ocular dominance, orientation, and direction maps. Because of the numerous lateral connections within the cortex, the cortical density has the most effect on the computation time and memory, and thus it is important to choose the appropriate value needed for each simulation.

The equations for changing cortical density are analogous to those for retinal receptor density, with the additional requirement that the intracortical connection sizes and associated learning rates must be scaled. The lateral connection radii r_E and r_I should be scaled with N so that the ratio between each radius and N remains constant. For simulations that shrink the lateral excitatory radius, the final radius r_{E_f} must also be scaled. Like α_A in retinal density scaling, α_E and α_I must be scaled so that the average total weight change per neuron per iteration remains constant despite changes in the number of connections. Finally, the absolute weight level D_I below which lateral inhibitory connections are killed at a given iteration must be scaled when the total number of lateral connections changes. The absolute value varies because the weight normalization in equation 4.8 ensures that when there are fewer weights each one is stronger. The value of each weight is proportional to the number of connections, so D_I should scale with the number of lateral inhibitory connections. In the continuous plane, that number is the area enclosed by the lateral inhibitory radius. Thus, the cortical density scaling equations are:

56

$$r_{\mathrm{E}} = \frac{N}{N_o} r_{\mathrm{E}_o}, \quad \alpha_{\mathrm{E}} = \frac{r_{\mathrm{E}_o}{}^2}{r_{\mathrm{E}}{}^2} \alpha_{\mathrm{E}_o}, \quad D_{\mathrm{I}} = \frac{r_{\mathrm{I}_o}{}^2}{r_{\mathrm{I}}{}^2} D_{\mathrm{I}_o}$$
$$r_{\mathrm{I}} = \frac{N}{N_o} r_{\mathrm{I}_o}, \quad \alpha_{\mathrm{I}} = \frac{r_{\mathrm{I}_o}{}^2}{r_{\mathrm{I}}{}^2} \alpha_{\mathrm{I}_o},$$

<div align="right">(5.3)</div>

Figure 5.4 shows how these equations can be used to generate closely matching orientation maps of different cortical densities. In the larger maps, more details are evident, but the overall structure is very similar beyond a minimum size scale.

Although the Nyquist theorem puts limits on the minimum N necessary to faithfully represent a given orientation map pattern, in practice the limiting parameter is the minimum excitatory radius (r_{E_f}). For instance, the map pattern from figure 5.4(e) can be reduced using image manipulation software to 18×18 without changing the largest-scale pattern of oriented blobs. Yet when simulated in HLISSOM even a 36×36 map differs from the larger ones, and (to a lesser extent) so does the 48×48 pattern. These differences result from quantization effects on r_{E_f}. Because units are laid out on a rectangular grid, the smallest neuron-centered radius that includes at least one other neuron center is 1.0. Yet for small enough N, the scaled r_{E_f} will be less than 1.0, and thus must either be fixed at 1.0 or truncated to zero. If the radius goes to zero, the map will no longer have any local topographic ordering, because there will be no local excitation between neurons. Conversely, if the radius is held at 1.0 while the map size continues to be reduced, the lateral spread of excitation will take over a larger and larger portion of the map, causing the orientation blob width of the resulting map to increase. Thus in practice, N should not be reduced so far that r_{E_f} approaches 1.0.[1] Above the minimum bound, adding more cortical density adds more detail to the orientation map, without changing the overall organization. Thus the cortical density equations can be used to select a good tradeoff between simulation detail and computational requirements, and make it easy to scale up to larger densities to simulate more complex phenomena like joint maps of orientation and direction preferences (as in Bednar and Miikkulainen 2003, to appear).

Together, the area and density scaling equations allow essentially any desired cortex and retina size to be simulated without a search for the appropriate parameters. Given fixed resources (such as a computer of a certain speed with a certain amount of memory), they make it simple to trade off density for area, depending on the phenomena being studied. They also make it simple to scale up to larger simulations on supercomputers when both area and density are needed.

5.4 Discussion

Apart from their application to simulations, the scaling equations give insight into how the corresponding quantities differ between individuals, between species, and during development. In essence, the equations predict how the biophysical correlates of the parameters will differ between

---

[1] If working with very small maps is necessary, radius values smaller than 1.0 could be approximated using a technique similar to antialiasing in computer graphics. Before a weight value is used in equation 4.7 at each iteration, it would be scaled by the proportion of its corresponding pixel's area that is included in the radius. This technique should permit smaller networks to be simulated faithfully even with a discrete grid, but has not yet been tested empirically.

any two similar cortical regions that differ in size. The discrepancy between the actual parameter values and those predicted by the scaling equations can help explain why different brain regions, individuals and species will have different functions and performance levels.

For instance, equations (5.3) and the simulation results suggest that learning rates per connection should scale with the total number of connections per neuron. Otherwise neurons from a more densely connected brain area would have significantly greater total plasticity, which (to my knowledge) has not been demonstrated. Consequently, unless the number of synapses per neuron is constant, the learning rate must be regulated at the whole-neuron level rather than being a property of individual synapses. This principle conflicts with assumptions implicit in most incremental Hebbian models that specify learning rates for individual connections directly. Future experimental work will be needed to determine whether such whole-neuron regulation of plasticity does occur, and if not, whether more densely connected regions do have a greater level of overall plasticity.

Similarly, equations (5.3) suggest that the pruning threshold D_I depends on the total number of connections to the neuron, rather than being an arbitrary fixed value. With the divisive weight normalization used in HLISSOM, increasing the number of connections decreases the strength of each one; this procedure is motivated by *in vitro* findings of whole-cell regulation of excitability (Turrigiano, Leslie, Desai, Rutherford, and Nelson 1998). A consequence is that a fixed D_I that prunes e.g. 1% of the connections for a small cortex would prune all of the connections for a larger cortex. This finding provides independent computational and theoretical support for experimental evidence that pruning is a competitive, not absolute, process (Purves 1988).

The scaling equations also provide an effective tool for making cross-species comparisons, particularly between species with different brain sizes. In effect, the equations specify the parameter values that a network *should* use if it is to have similar behavior as a network of a different size. As pointed out by Kaas (2000), different species do *not* usually scale faithfully, probably due to geometrical, metabolic, and other restrictions. As a result, as V1 size increases, the lateral connection radii do not increase as specified in the cortical density scaling equations, and processing becomes more and more topographically local. Kaas (2000) proposes that such limitations on connection length may explain why larger brains such as human and macaque are composed of so many visual areas, instead of just expanding the area of V1 to support greater functionality (see also Catania et al. 1999). The scaling equations combined with HLISSOM provide a concrete platform on which to test these ideas in future multi-region simulations.

5.5   Conclusion

The scaling equations allow an HLISSOM simulation to scale up in area or density to match a biological system for which experimental data is available, and to scale down to save computation time and memory. The scaled-up maps have similar structure and function, but represent a greater portion of the visual field or represent it with finer detail. The larger simulations can then be used

to model large-scale phenomena that cannot be addressed in smaller models, such as joint maps of multiple stimulus features. Later chapters will use these equations to develop large-scale models of V1 and show how they can be used as the first stage of a hierarchy of cortical regions.

Chapter 6

Development of Orientation Perception

This chapter will use the HLISSOM model and scaling equations introduced in the previous chapters to model orientation map development in detail, using realistic inputs. The results show that both internally generated and environmental stimuli are needed to explain the experimentally measured orientation maps. Together these sources of activity can account for the complex process of orientation map development in V1. The simulations also act as a well-grounded test case for the methods used in the face perception experiments in later chapters.

6.1   Goals

In a series of simulations below, I will show that:

1. Orientation maps can develop from either random or structured (blob-like) internally generated activity, but that structured activity is necessary for the maps to match experimental measurements from newborn animals (section 6.2).

2. Orientation maps can also develop from natural images alone, and the distribution of orientation preferences matches the distribution of orientations in those images (section 6.3).

3. Training on internally activity first, then natural images, shows how the orientation map can exist at birth, and then smoothly become a better match to the environment (section 6.4).

Thus either spontaneous activity or natural images alone would be sufficient to drive the development of orientation maps, but both together are required to explain the experimental data.

6.2   Internally generated activity

The simulations in chapter 4 used two-dimensional Gaussians as an idealized representation of oriented features in the input. This representation consists only of single short lines, uncorrelated

with each other or with the edges of larger objects. This section and the next will show how making the input patterns match more closely either to spontaneous activity or to natural images affects the final maps and RFs. So that these separate simulations can be compared to each other and to results from other models, all of them will use the same type of input pattern throughout training, and all will use nearly identical conditions except for the specific input patterns. The goal of these simulations is to determine individually how each type of input affects development, so that it will be clear how each contributes to the full model of prenatal and postnatal development that I will present in section 6.4.

First, I consider internally generated activity alone. As is clear from figure 1.1, a realistic model of patterns like the retinal waves would include large, coherent patterns of activity, as well as background noise. Miller (1994) has argued that such patterns of activity are too large and too weakly oriented to drive the development of orientation preferences in V1. The results in this section show that orientation selectivity and realistic maps can develop in HLISSOM even from large, unoriented patterns of activity, and even with background noise. These results demonstrate that the known properties of retinal waves are sufficient for the development of orientation maps and selectivity. Of course, many sources of internally generated activity other than retinal waves may also share these features, including PGO activity; the retinal waves are simply the best known example of such activity. Besides being sufficient, the results will show that training on the oriented blobs seen in retinal waves actually results in more realistic maps than generated by models that do not include such blobs, such as Miller's.

6.2.1   Discs

To ensure that the internally generated blob patterns have no orientation bias, they were rendered from the following equations for a circular disc with smooth Gaussian falloff in brightness around the edges. (The real patterns often do have orientation biases, but omitting them here shows that the elongated edges are not a necessary feature for orientation map development.) Each disc pattern is specified by the location of the disc center (x_i, y_i), the width w of the full-intensity central portion of the disc, and the half-width σ_f for Gaussian smoothing of the outer border. To calculate the activity for each photoreceptor (x, y), the Euclidean distance d of that photoreceptor from the disc center is calculated:

$$d = \sqrt{(x - x_i)^2 + (y - y_i)^2}.$$ (6.1)

The activity ξ for receptor (x, y) is then

$$\xi_{x,y} = \begin{cases} 1.0 & d < \frac{w}{2} \\ \exp(-\frac{(d - \frac{w}{2})^2}{\sigma_f{}^2}) & d >= \frac{w}{2}. \end{cases}$$ (6.2)

The x and y coordinates of the centers are each chosen randomly, and the orientation is chosen randomly from the uniform distribution in the range $0° \leq \theta < 180°$. The brightness of each pattern was either positive or negative relative to the mean brightness, and was chosen randomly.

61

The **Discs** row of figure 6.1 shows that even with input patterns that do not favor any particular orientation, HLISSOM can develop oriented receptive fields with ON and OFF subregions, as well as a realistic orientation map. As far as I am aware, this finding is the first demonstration that orientation maps can develop from large, unoriented activity patterns. The development is driven by the orientations of small patches around the edge of the circular discs. These oriented edges are visible in the LGN response to the disc pattern in figure 6.1*b*. Because the input patterns had only edges and no lines or bars, most neurons develop two-lobed receptive fields, not three-lobed. Both two-lobed and three-lobed receptive fields are common RF types for simple cells in adult V1 (Hubel and Wiesel 1968), but it is not known what RF types are present in newborns. These results suggest that if orientation map development is driven by large, spatially coherent blobs of activity, newborns will primarily have two-lobed receptive fields in V1.

6.2.2 Noisy Discs

The retinal waves in figure 1.1 contain spatially uncorrelated background noise in addition to the retinal waves. This noise is most likely a combination of measurement noise plus uncorrelated neural activity. It is not clear how much noise is due to either source, but assuming at least some is due to genuine neural activity is reasonable.

Here, I model the noise as a random additive value for each pixel, drawn from the uniform distribution in the range ± 0.5. The **Noisy Discs** row in figure 6.1 shows that realistic maps develop even with substantial noise. However, the final map has fewer highly selective neurons. This result suggests that including the spatially uncorrelated noise may be important for faithfully modeling newborn maps, which also tend to have lower selectivity.

6.2.3 Random noise

As a control, the **Random Noise** row in figure 6.1 shows that orientation maps can even develop from uniformly random noise, where each input pixel is a random number between 0 and 1.0. Such a simulation is in many respects similar to the simpler model of Linsker (1986a,b,c), who also showed that orientation-selective neurons can develop from uncorrelated noise. Even such fully random patterns have local clusters that are brighter or darker than their surround. These clusters lead to patches of activity in the ON channel adjacent to patches in the OFF channel (figure 6.1(*b*)). The result is that orientation-selective neurons develop even from uncorrelated noise. However, the resulting V1 map is less well organized than typically seen in animals, even at birth, and most neurons develop much lower orientation selectivity than for the other patterns above. The lower selectivity is particularly evident in the sample RFs, most of which would be a good match to many different oriented lines.

These results show that even fully random activity can produce orientation maps, but suggest that the inputs need to be spatially coherent for realistic RFs and maps to develop. Thus noise alone
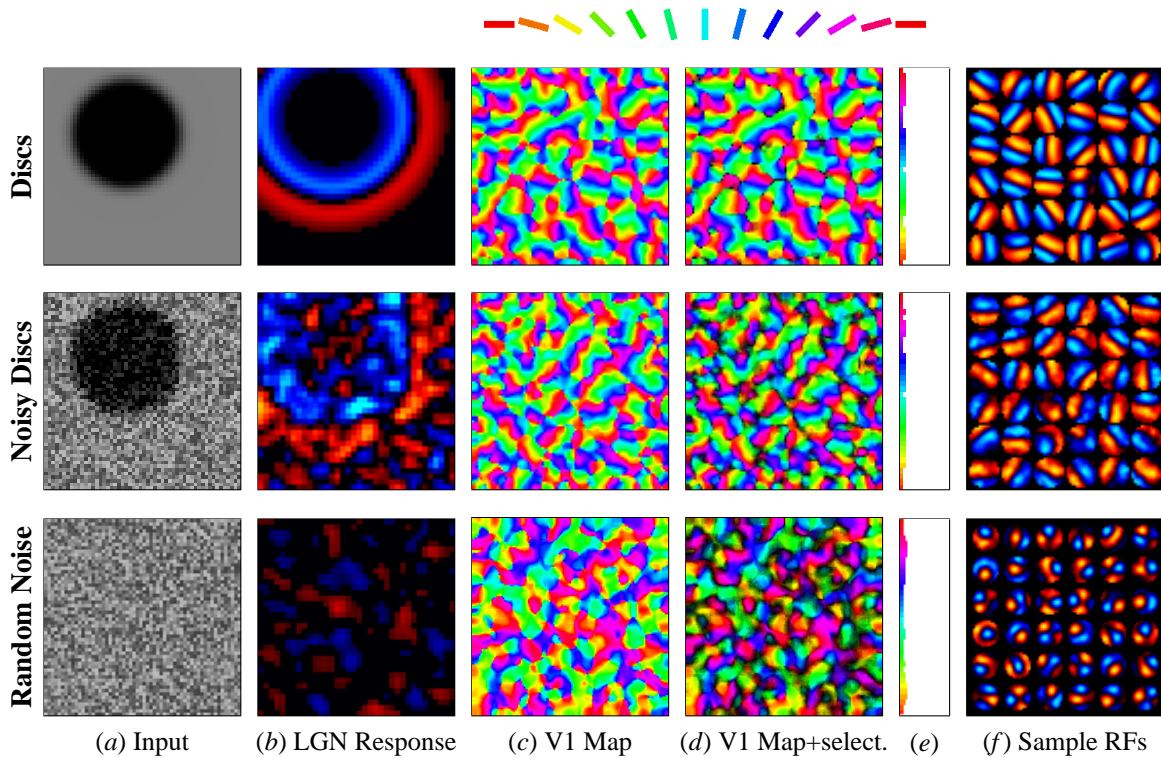
Figure 6.1: **Self-organization based on internally generated activity (*color figure*).** See figure 4.8 on page 47 for the legend for each column. The parameters for the networks plotted on each row were similar except for the training inputs, of which samples are given in (*a*). Spontaneously generated retinal waves are spatially coherent with an oriented edge, like the pattern in the top left. HLISSOM can develop a realistic orientation map from such inputs presented at random locations (*c-f*). Thus even patterns without any dominant orientation can result in the development of orientation selective neurons and orientation maps. Nearly all of the resulting receptive fields have two lobes (edge-selective) rather than three (line-selective); both RF types are found in the visual cortex of adult animals (Hubel and Wiesel 1968). Other sources of spontaneous activity may be better modeled as spatially uncorrelated random noise (bottom row). HLISSOM can also develop an orientation map from such noise. However, the selectivity is significantly lower (*d*), the orientation blob spacing is less regular than in typical animal maps, and most neurons develop less-selective RFs (*f*), unlike those typically found in adult V1. More realistic orientation maps develop when HLISSOM is trained on both random noise and spatially coherent patterns together (middle row). This type of pattern is a close match to the measured retinal waves; compare (*a*) with figure 1.1. Even when the noise is quite strong relative to the coherent patterns (see LGN response in *b*), the resulting map is still significantly more selective than for maps trained on random noise alone (compare *d*, middle with the other two). The selectivity is also visible in the RFs (*f*), which are only mildly affected by the uncorrelated noise. Thus the spatially coherent blobs dominate self organization, even when there is also uncorrelated noise. The combined patterns (**Noisy Discs**) will be used as a model of internally generated activity in later figures, because they are a more realistic model of the known internally generated patterns.

is insufficient. Of these simulations, the **Noisy Discs** patterns are both more realistic and develop maps that more closely match experimental results from newborn animals. I will therefore use these patterns to model the prenatal phase of orientation map training in the combined prenatal and postnatal model in section 6.4.

6.3   Natural images

Even though orientation maps are present at birth, how they develop postnatally depends on experience with real images of the visual environment. This section and figure 6.2 show that HLISSOM can realistic orientation maps and receptive fields from natural images, which has only been shown for a few other models so far. Moreover, the distribution of orientation preferences that develops in HLISSOM depends on the statistical distribution of features in different sets of images. Such dependence on the environment has been found for animal maps (Sengpiel et al. 1999), but to my knowledge has not yet been demonstrated in models.

To be comparable with the previous simulations of spontaneous activity and with other models using natural images, these simulations use natural images throughout the course of self-organization, and do not include a prenatal phase with spontaneous activity. Although the results will show that natural images alone would be sufficient to drive the development of orientation maps, the simulations are not a close match to biology because they do not explain how the maps could be present at birth. In section 6.4 I will combine natural images with internally generated inputs to model the full process of prenatal and postnatal orientation map development, based on these results from simpler conditions.

6.3.1   Image dataset: Nature

The **Nature** row of figure 6.2 shows the result from training an HLISSOM map on randomly selected 36×36 patches of 256×256-pixel natural images. The set of 25 images of naturally occurring objects was taken by Shouval et al. (1997, 1996). The authors specifically excluded images with prominent straight lines, to show that such lines were not required for orientation selectivity to develop. Nearly all of these images are short-range closeups, although there are a few wide-angle landscapes showing the horizon.

When trained on these images, the self-organized HLISSOM orientation map is quite similar to the **Discs** and **Noisy Discs** maps (compare the **Nature** row of figure 6.2 with figure 6.1). Many of the neurons have lower orientation selectivity than for the artificial stimuli, which is expected because the natural images contain many patterns other than pure edges. Even so, most of the RFs are orientation selective, as found in V1 of animals.

Unlike in the networks trained on artificially generated stimuli, the distribution of orientation preferences in the HLISSOM **Nature** map is slightly biased towards horizontal and vertical orientations. This result suggests that the distribution of orientation preferences in both the model

| | (a) Input | (b) LGN Response | (c) V1 Map | (d) V1 Map+select. | (e) | (f) Sample RFs |
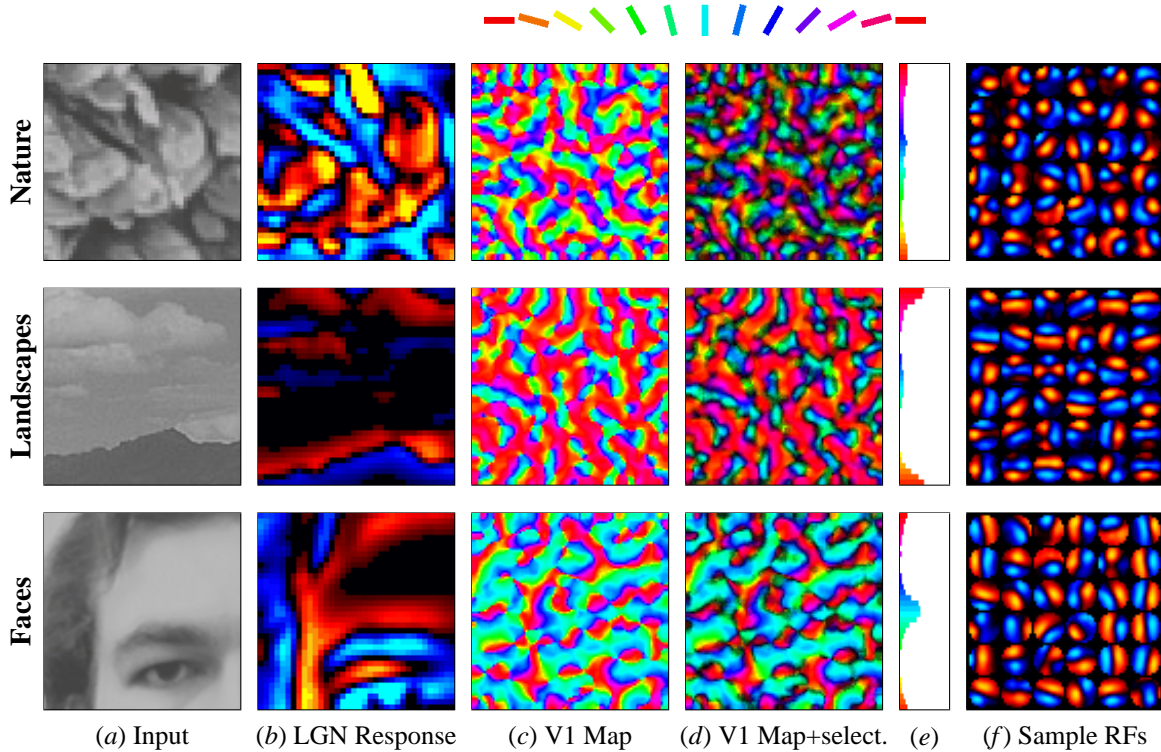
Figure 6.2: **Orientation maps develop with natural images (*color figure*).** See figure 4.8 for the legend for each column. The networks plotted on each row were similar except for the set of training images used. The network in the top row was trained on images of natural objects and (primarily) close-range natural scenes, from Shouval et al. (1997, 1996). It develops a realistic orientation map, albeit with a lower selectivity than for the artificial stimuli. The lower selectivity is unsurprising, because the natural stimuli have many unoriented features. The map is slightly biased towards horizontal and vertical orientations (*e*), as is the natural environment (Switkes et al. 1978). Other image databases with different distributions of oriented patterns can have very different results. The second row shows results for a set of stock photographs from the National Park Service (1995), which are primarily landscapes with abundant horizontal contours. The resulting map is dominated by neurons with horizontal orientation preferences (red), with a lesser peak for vertical orientations (cyan). Results with a database of upright human faces have the opposite pattern, with a strong peak at vertical and a lesser peak at horizontal (bottom row). Thus HLISSOM can self-organize orientation maps and receptive fields from natural images, and the resulting maps reflect the statistics of the image dataset. Later simulations combining natural inputs with internally generated patterns will show that this natural image learning can account for postnatal learning of contours in the environment.

and in animals reflects the over-representation of vertical and horizontal contours in the natural environment (Switkes et al. 1978). The subsection below explores this dependence in more detail, because it illustrates why postnatal learning is important.

6.3.2 Effect of strongly biased image datasets: Landscapes and Faces

Raising animals in artificial environments that are biased towards certain orientations leads leads to a bias in the distribution of orientation-selective cells in V1 (Blakemore and Cooper 1970; Sengpiel et al. 1999). To model being raised in different environments, albeit ones less extreme than in those studies, two networks similar to the **Nature** map were trained on images of landscapes and faces, respectively. The landscapes images were a set of 58 stock photos from the National Park Service (1995). Nearly half of the images in this set are wide-angle photos showing the horizon or other strong horizontal contours; a few also include man-made objects such as fences. The face images are a set of 30 frontal photographs of upright human faces (Achermann 1995). The HLISSOM map resulting from each image set is shown in the **Landscapes** and **Faces** rows of figure 6.2.

The **Landscapes** map is strongly biased toward horizontal contours, with a much weaker bias for vertical. These biases are visible in the orientation map, which is dominated by red (horizontal) and, to a lesser extent, cyan (vertical). The opposite pattern of biases is found for the **Faces** map, which is dominated by cyan (vertical), and to a lesser extent, red (horizontal). These divergent results replicate the single-unit results found in cats raised in skewed environments (Blakemore and Cooper 1970). They also replicate the recent finding that such animals have orientation maps with enlarged orientation domains for the orientations overrepresented in the environment (Sengpiel et al. 1999). Thus the HLISSOM model shows how different environments can influence the development of orientation maps. The result is that the most common contours in the environment are the most well-represented in visual cortex.

6.4 Prenatal and postnatal development

The preceding simulations use only a single type of inputs (either artificial or images) throughout training, to show how each type of input affects development. This section shows that presenting the most realistic internally generated pattern model (the **Noisy Discs**), followed by natural images representing postnatal experience, can account for much of the prenatal and postnatal development of orientation selectivity.

Figure 6.3 demonstrates that (1) an initial map can be self-organized using prenatal patterns, (2) it can smoothly self-organize postnatally with natural images to improve selectivity, and (3) the postnatal learning will closely match the properties of the map to the statistical properties of the natural images. Interestingly, the early postnatal development of the HLISSOM map after iteration 1000 was similar whether internally generated patterns or typical natural images were used. This result replicates Crair et al.'s (1998) finding that normal visual experience typically has only a
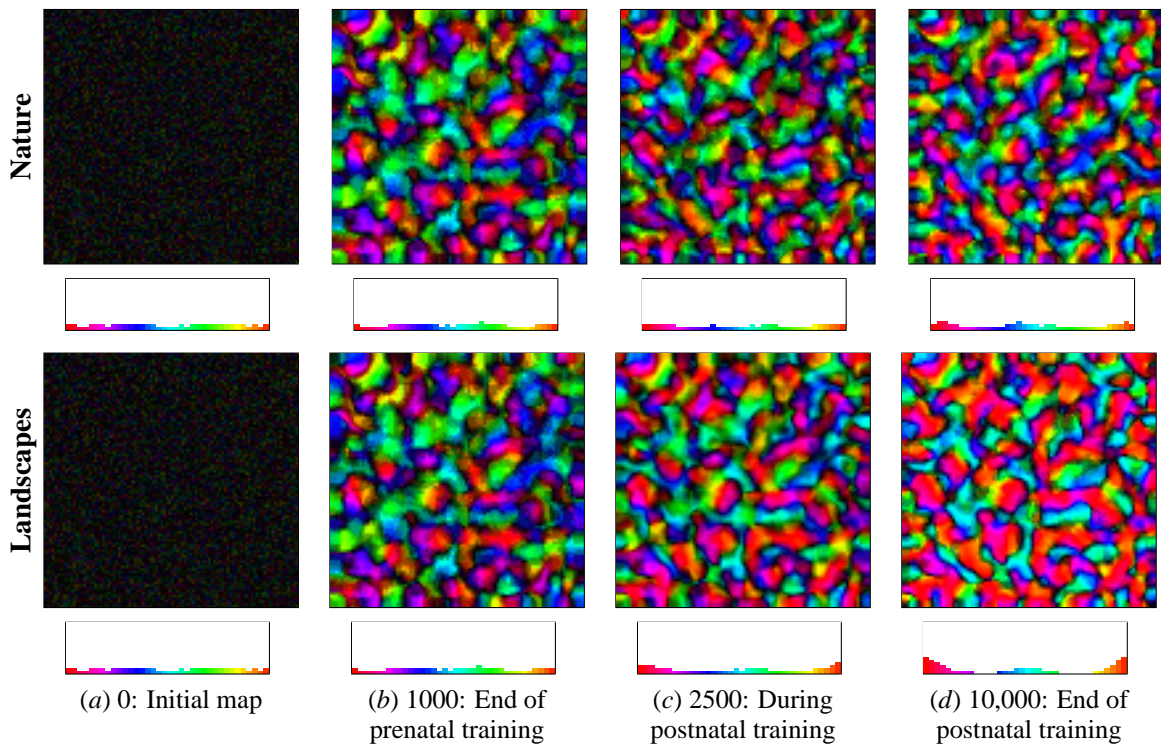
Figure 6.3: **Postnatal training makes orientation map match statistics of the environment (*color figure*).**
Each row shows results from a network trained for 1000 iterations on the **Noisy Discs** patterns from figure 6.1
as a model of internally generated activity, then trained for 9000 further iterations using natural images to
model postnatal visual experience. The orientation map plots (*b-d*) show selectivity as a brightness level,
so that the postnatal improvement in selectivity will be visible. (*a*) and (*b*) are the same in each row. The
top row shows the effect of postnatal training on the **Nature** dataset. With these images, more neurons
become sensitive to horizontal and vertical contours, and the overall selectivity increases. However, the
overall map shape remains similar, as found in laboratory animals (Chapman et al. 1996; compare individual
blobs between maps right to left or left to right). The postnatal changes when trained on the **Landscapes**
dataset are similar but much more pronounced. With these images the network smoothly develops strong
biases for vertical and horizontal contours, within the pre-determined map shape. These results show that
postnatal learning can gradually adapt the prenatally developed map to match the statistics of an animal's
natural environment, while explaining how a genetically specified orientation map can be present at birth.

small effect on map development in kittens, and that similar maps developed whether the eyes were
open or closed. Yet atypical collections of natural images can cause readily visible changes in the
map, which replicates findings from kittens raised with abnormal visual experience (Blakemore and
Cooper 1970; Sengpiel et al. 1999).

Figure 6.4 specifically compares the maps after prenatal and postnatal training to maps
measured at birth and adulthood in animals. The self-organized map is very similar to that found in
newborn ferrets and binocularly deprived cats (Chapman et al. 1996; Crair et al. 1998), showing that
even simple internally generated inputs can account for the initial orientation map. The individual

(a) Prenatal map detail   (b) Neonatal cat (1.9×1.9mm)

(c) Adult map   (d) Adult monkey (5×5mm)
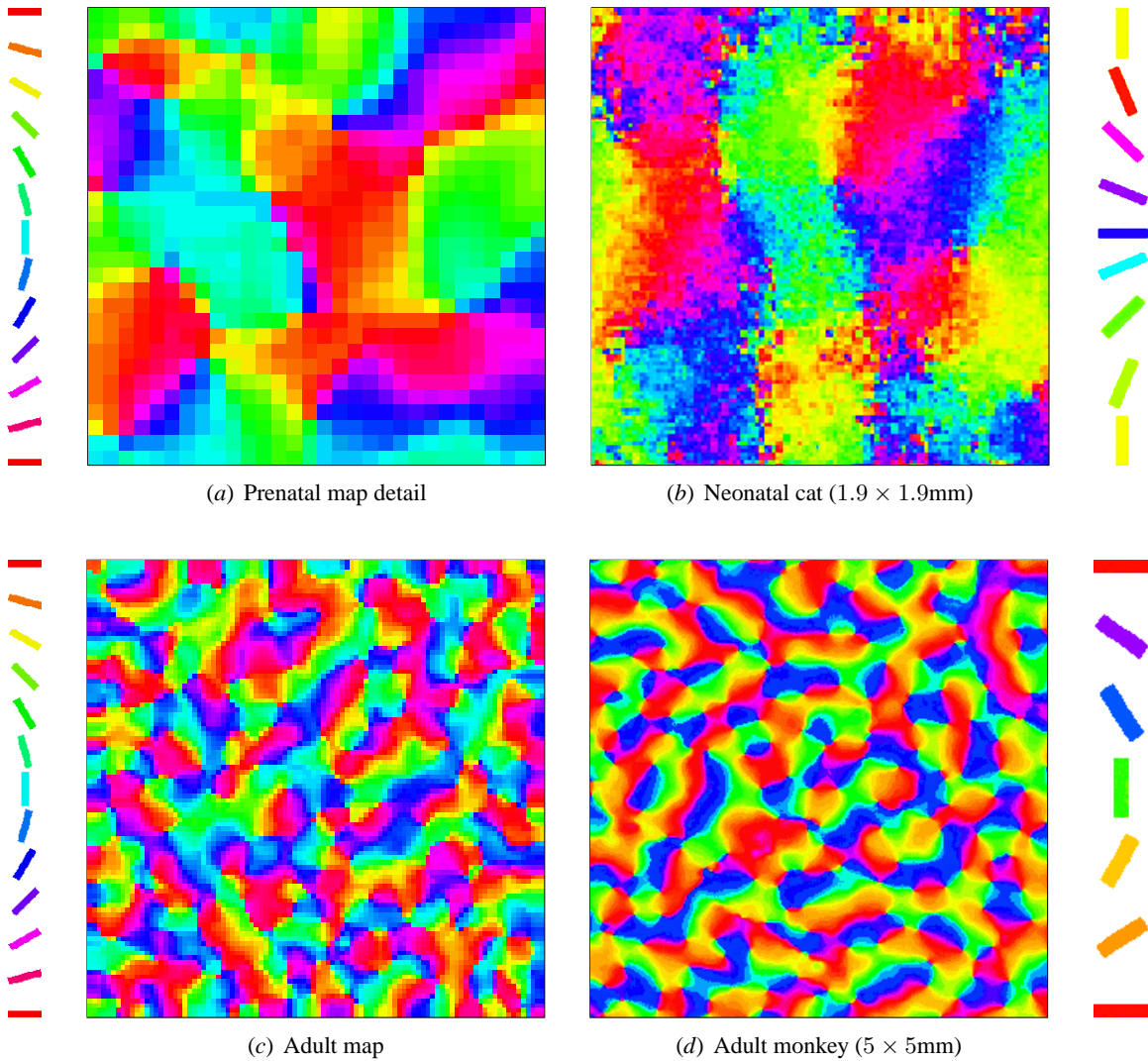
Figure 6.4: **Prenatal and postnatal maps match animal data (*color figure*).** After 1000 iterations of prenatal training, the **Nature** network from figure 6.3 has developed a rudimentary orientation map. Plot (a) shows the central 30×30 region of this map, and plot (b) shows a map measured from a cat without prior visual experience. (Reprinted with permission from Crair et al. (1998), copyright 1998, American Association for the Advancement of Science.) After postnatal training on natural images, the full 96×96 map is quite similar to measurements from experimental animals, including the monkey cortex shown here. (Reprinted with permission from Blasdel 1992b, copyright 1992 by the Society for Neuroscience). Thus the HLISSOM model explains both the prenatal and adult orientation maps found in experimental animals, as self-organization from internally generated and environmental stimuli.
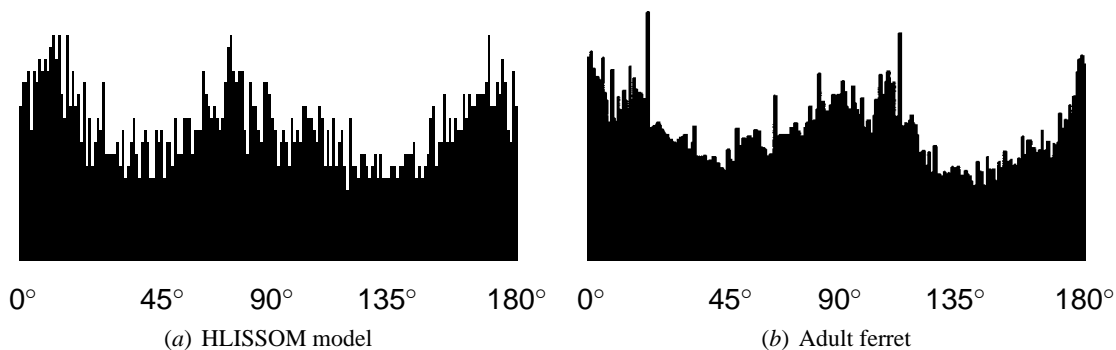
Figure 6.5: **Orientation histogram matches experimental data.** (*a*) Histogram of orientation preferences for the **Nature** network from figure 6.4 and the top row of figure 6.3, which was trained first on **Noisy Discs** and then on the **Nature** dataset. (*b*) Corresponding histogram for a typical adult ferret visual cortex (reprinted from Coppola et al. 1998; copyright National Academy of Sciences, U.S.A.; used with permission.) Both adult ferrets and the HLISSOM model trained on natural images have more neurons representing horizontal or vertical than oblique contours, which reflects the statistics of the natural environment. HLISSOM maps trained on internally generated patterns alone show an approximately flat distribution instead, because they were shown an unbiased distribution of input patterns (round discs and/or random noise).

receptive fields that develop are also good approximations to simple cells in monkey and cat visual cortex (Hubel and Wiesel 1962, 1968). Finally, figure 6.5 shows that postnatal training on natural images causes the HLISSOM map to develop a bias for horizontal and vertical contours, and that the bias matches that found in orientation maps of adult ferrets (Chapman and Bonhoeffer 1998; Coppola et al. 1998). Thus HLISSOM can explain the animal results, and both phases are necessary to account for the experimental data: internally generated patterns before birth, and experience with natural images after birth.

6.5   Discussion

The results in this chapter show that prenatal training on internally generated activity, followed by postnatal training on natural images, can account for much of the development of orientation maps, orientation selectivity, receptive fields, and lateral connections in V1. The same activity-dependent learning rules can explain development based on both types of neural activity, internally and externally generated.

Both types of activity appear to serve important roles in this developmental process, and in particular both are crucial for replicating the experimental data. Postnatal experience with environmental stimuli ensures that the map faithfully represents the environment, while prenatal development ensures that the map is functional even at eye opening. Prenatal development may also be important for the development of higher areas connected to V1 (e.g. V2 and V4), because it ensures that the map organization in V1 is approximately constant after birth (figure 6.3). As a result, the

higher areas can begin learning appropriate connection patterns with V1 even before eye opening.

Comparing orientation maps and RFs trained on random noise versus those trained on images, discs, or Gaussians suggests that oriented features are needed for realistic RFs. Even though maps can develop without such features, the RFs do not match those typically measured in animals. A similar result was recently found independently by Mayer, Herrmann, and Geisel (2001) using single RF simulations. However, Mayer et al. (2001) conclude that natural images are required for realistic RFs, because they did not consider patterns like the **Noisy Discs**. The results here suggest that any pattern with large, coherent blobs of activity will suffice, and thus that natural images are not strictly required for RF development.

In animals, the map present at eye opening is more noisy and and has fewer selective neurons than the prenatally trained maps shown in figures 6.3 and 6.4 (Chapman et al. 1996; Crair et al. 1998). As a result, in animals the postnatal improvement in selectivity is larger than that shown here for HLISSOM. The difference may result from the immature receptive fields in the developing LGN (Tavazoie and Reid 2000). Using a more realistic model of the developing LGN would result in a larger postnatal improvement, but would make the model significantly more complex to analyze. Other factors that could contribute to the lower selectivity at birth include the greater variability of cortical responses in infants, which will cause a lower apparent selectivity. Such behavior could be modeled by adding internal noise to the prenatal neurons, which again would make the model more complex to analyze but could provide a closer match to the biological case.

A recent study has also reported that the distribution of orientation selective cells matches the environment even in very young ferrets, i.e. that horizontal and vertical orientations are over-represented in orientation maps at eye opening (Chapman and Bonhoeffer 1998). One possible explanation for these surprising results is the non-uniform distribution of retinal ganglia along the horizontal and vertical meridians in the retina (Coppola et al. 1998), which could bias the statistics of internally generated patterns. Regardless of such a prenatal bias, HLISSOM shows how visual experience will also drive the map to develop preferences that match the visual environment.

6.6 Conclusion

Overall, the HLISSOM results show that internally generated activity and postnatal learning can together explain much of the development of orientation preferences. Either type of activity alone can lead to orientation maps, but only with realistic prenatal activity and postnatal learning with real images could the model account for the full range of experimental results. These studies also allowed the HLISSOM model to be tested and tuned in a domain that has abundant experimental data for validation, so that it can later be applied to other domains.

Chapter 7

Prenatal Development of Face Detection

In the previous chapter, I showed how internally generated patterns and visual experience can explain prenatal and postnatal development of orientation preferences in V1, a relatively well-studied area. In this chapter and the next I will apply the same ideas to the development of face detection. I focus first on prenatal development, showing how internally generated activity can explain newborn face preferences. Chapter 8 will then model postnatal experience with real faces, and show that such learning can account for experimental results with older infants.

7.1 Goals

The prenatal face detection simulations in this chapter will demonstrate that:

1. A scaled-up HLISSOM model of V1 can extract the local orientations in large natural images, which means that it can be used as the first cortical stage in a hierarchical model of the visual system (section 7.2.1).

2. With this V1 map and internally generated activity, a higher level map can develop three-dot preferences. This result shows that a hierarchy of brain areas can self-organize and function at a detailed, realistic neural level (section 7.2.2).

3. Specific feature preferences at the different levels of the hierarchy can replicate the reported newborn schematic pattern preferences, facelike and non-facelike (section 7.3.1).

4. Preferences for three-dot patterns result in significant preference for faces in natural images (section 7.3.2).

5. The shape of the internally generated patterns is important, but it does not need to be precisely controlled to result in face preferences (section 7.3.3).

71

Together, these results show that a system like the CONSPEC three-dot conceptual model of face detection from Johnson and Morton (1991; described in section 3.3.5) can be implemented and tested computationally. Moreover, they provide a specific, plausible way that such a system for face detection could be constructed in the developing brain.

7.2 Experimental setup

As illustrated in figure 4.1 on page 35, newborn face detection experiments used multiple cortical levels, including both V1 and the FSA (face-selective area). They also included multiple input regions, the photoreceptors and the PGO generating area.

Previous developmental models have been tested only for the small areas typical of the networks in chapters 4–6. However, the newborn face experiments that will be modeled in this chapter presented head-sized stimuli at a distance of about 20cm from the baby's eyes. As a result, the stimuli filled a substantial portion of the visual field (about $45°$). To model behavior at this scale, the scaling equations were used to scale up the **Discs** simulation from figure 6.1 to model a very large V1 area (approximately 1600 mm^2 total) at a relatively low sampling density per mm^2 (approximately 50 neurons/mm^2). The cortical density was first reduced to the minimum value that would show an orientation map that matches animal maps (36×36), and then the visual area was scaled to be just large enough to cover the visual images to be tested (288×288, i.e. width $\times 8$ and height $\times 8$). The FSA size was less crucial, and was set arbitrarily at 36×36. The FSA RF size was scaled to be large enough to span the central portion of a face.

The resulting network 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. There were 80 million connections total in the two cortical sheets, which required 300 megabytes of physical memory.

The RF centers of neurons in the FSA were 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 FSA 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.[1] In the figures in this chapter, only the area mapped directly to V1 will be shown, to ensure that all plots have the same size scale. The input pattern size on the retina was based on measurements of the preferred spatial frequency of newborns, as cited by Valenza et al. (1996). That is, inputs were presented at the spatial scale where the frequency most visible to newborns produced the largest V1 response in the model. The remaining parameters are listed in section A.5.

---

[1]Note that the photoreceptors are modeled as having uniform density for simplicity; there is no increase in sampling density in the fovea.

7.2.1 Development of V1

V1 was self-organized for $10,000$ iterations on 11 randomly located circular discs per iteration, each 50 units wide (figure 7.1*a*). These patterns were used for simplicity rather than **Noisy discs** or a combination of discs and images, because the focus of these simulations is on the development of the higher-level FSA region. 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 $\sigma = 1.5$. After $10,000$ iterations (10 hours on a 600MHz Pentium III workstation), the scaled-up map shown in figure 7.1*e* emerged. Figure 7.2 shows that this scaled-up model will extract the salient local orientations in large images, which has not yet been demonstrated for other models of V1 development.

7.2.2 Development of the FSA

After the V1 training, its 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 two triples of circular dots per iteration, each arranged in a triangular face-like configuration (figure 7.3*a*). Each dot was 20 PGO units wide, and was 0.3 units darker than the surround, which itself was 0.5 on a scale of 0 to 1.0. Each triple was placed at a random location whose center was 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.

Because 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, the FSA in such a simulation would need to be much larger to represent the different sizes, and the resulting patchy FSA responses would require more complex methods of analysis.

Using these three-dot patterns, the FSA was self-organized for $10,000$ iterations (13 hours on a 600MHz Pentium III workstation). The result was the face-selective map shown in figure 7.3*h*, consisting of an array of neurons that respond most strongly to patterns similar to the training patterns. Despite the overall similarities between neurons, the individual weight patterns are unique because each neuron targets specific orientation blobs in V1. Such complicated patterns would be difficult to specify and hardwire, but they arise naturally from internally generated activity.

As FSA training completed, the activation threshold (i.e. β in figure 4.5) was specifically set to a high value, which ensures that only patterns that are a strong match to an FSA neuron's weights will activate it. This way, the presence of FSA activity can be interpreted unambiguously as an indication that there is a face at the corresponding location in the retina, which will be important for measuring the face detection ability of the network.

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

(d) Afferent weights of one V1 neuron   (e) Final V1 orientation map   (f) Detail of orientation map (e)
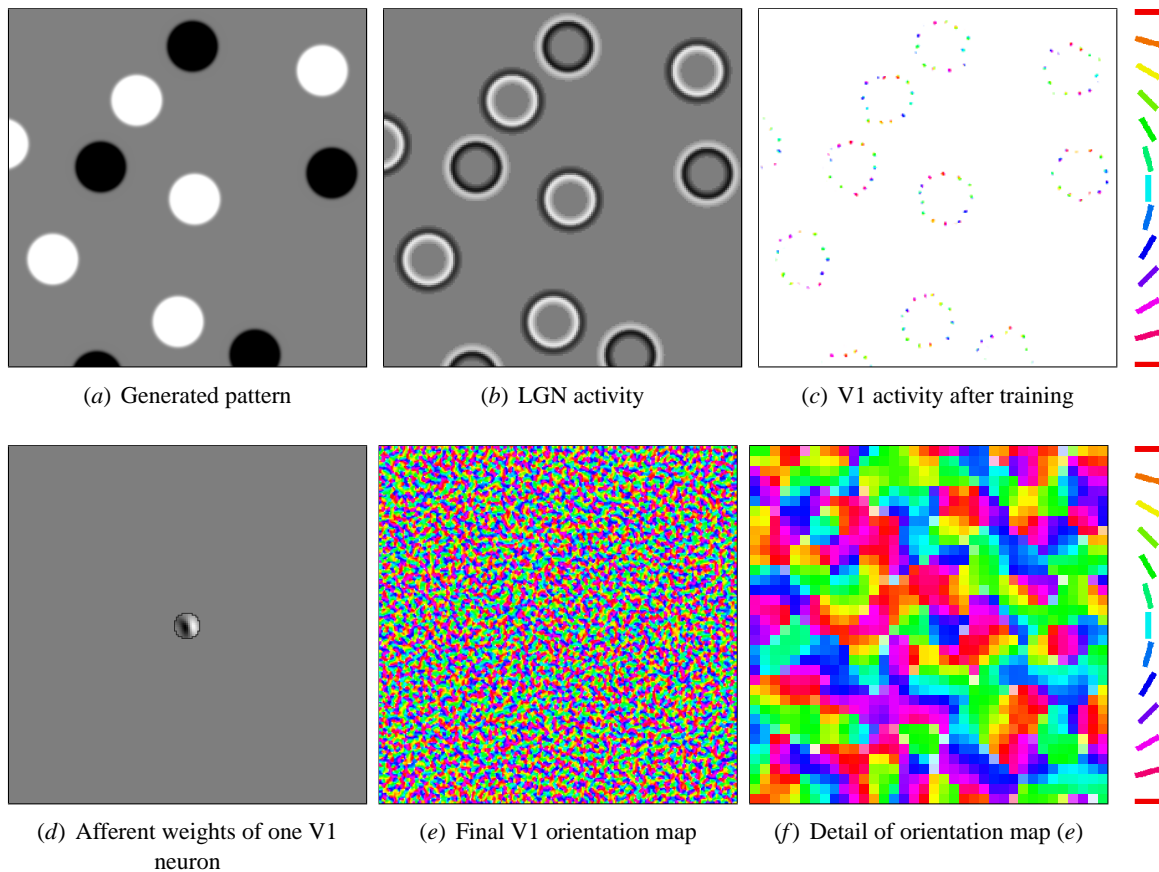
Figure 7.1: **Large-scale orientation map training (*color figure*).** These figures show a scaled-up version of the **Discs** simulation from figure 6.1, 8 times wider and 8 times taller. At this scale, the afferent weights of each neuron span only a small portion of the retina (d), and the orientation map has many more oriented blobs total (e). Zooming in on the central 36×36 portion (f) of the 288×288 map (e) shows that the underlying structure of the orientation map is similar to those in the previous chapter. The plot appears blockier, because the neuron density was reduced to the smallest acceptable value, so that the network would be practical to simulate. Plot (c) shows that the orientation of each neuron that responds to the input is still a good match to the orientation present at that retinal location, and thus that this network is a reasonable approximation to a large area of V1 and the retina. Figure 7.2 will show similar results for this network using natural images.

The cortical maps were then 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 section A.5 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

74

| (*a*) Sample visual image | (*b*) LGN response | (*c*) V1 response |
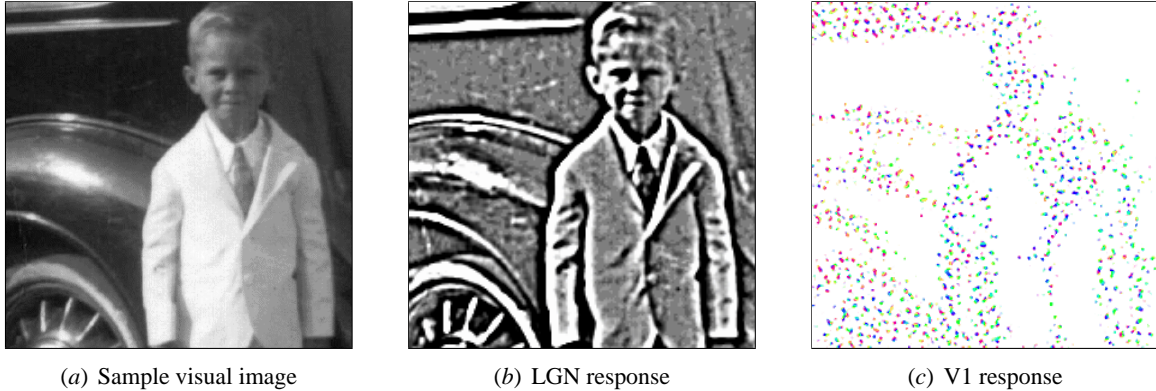
Figure 7.2: **Large-area orientation map activation (*color figure*).** The scaled-up orientation map works well with natural images; each V1 neuron responds if there is a line or edge of its preferred orientation at the corresponding location in the retina. Because the V1 activations preserve the important features of the input, the output from V1 can be used as the input to a higher-level map. As suggested by figure 4.9, earlier models like RF-LISSOM will instead have a broad, unspecific response to natural images, even if they have realistic orientation maps.

schematic images. The different scales model different states of contrast adaptation mechanisms in the retina that have not been implemented explicitly in this model.

7.2.3 Predicting behavioral responses

The HLISSOM model provides detailed neural responses for each neural region. These responses constitute predictions for future electrophysiological and imaging measurements in animals and humans. However, the data currently available for infant face perception is behavioral. Specifically, it consists of newborn attention preferences measured from visual tracking distance and looking time. Thus, validating the model on this data will require predicting a behavioral response based on the simulated neural responses.

As a general principle, I hypothesize that newborns pay attention to the stimulus whose overall neural response most clearly differs from typical stimuli. This idea can be quantified as

$$a(t) = \frac{F(t)}{\overline{F}} + \frac{V(t)}{\overline{V}}, + \frac{L(t)}{\overline{L}}, \tag{7.1}$$

where $a(t)$ is the attention level at time t, $X(t)$ is the total activity in region X at that time, $\overline{X}$ is the average (or median) activity of region X over recent history, and F, V, and L represent the FSA, V1, and LGN regions, respectively. Because most stimuli activate the LGN and V1 but not the FSA, when a pattern evokes activity in the FSA newborns would attend to it more strongly. Yet stimuli evoking only V1 activity could still be preferred over facelike patterns if their V1 activity is much higher than typical.

75

(*a*) Generated pattern	(*b*) LGN activity	(*c*) V1 activity before FSA training	(*d*) Afferent weights of one FSA neuron
(*e*) FSA response before training	(*f*) FSA response after training	(*g*) Initial map	(*h*) Trained map
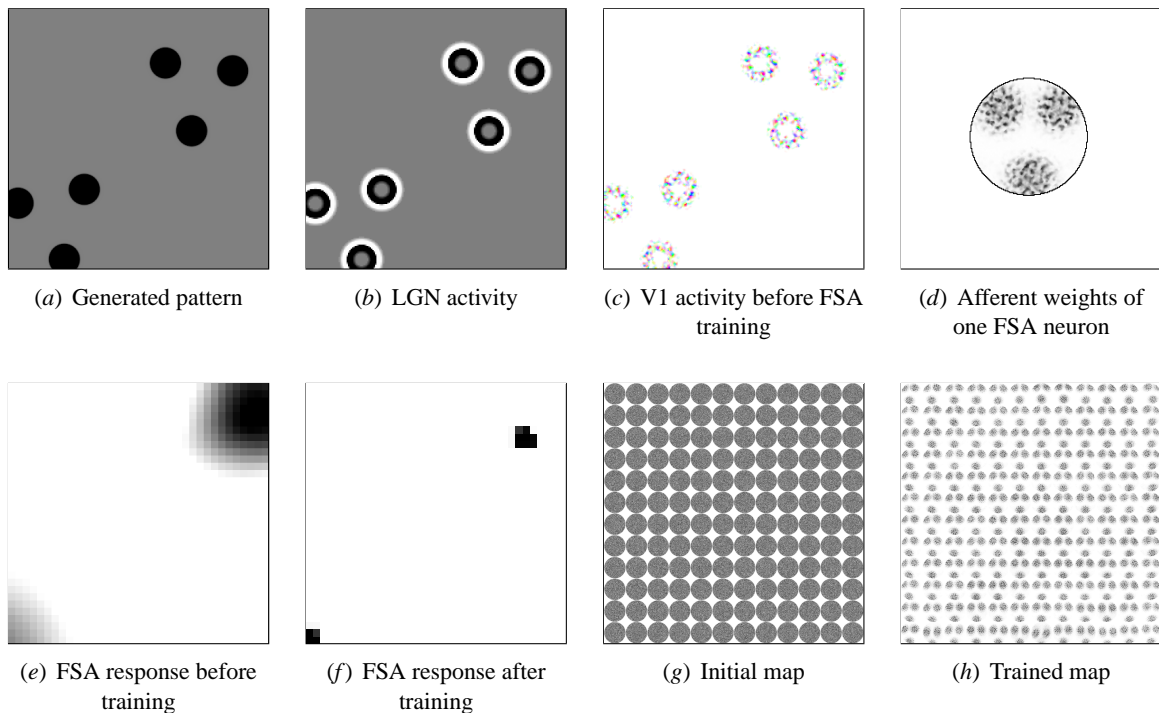
Figure 7.3: **Training the FSA face map.** (*a*) For training, the input consisted of simple three-dot configurations with random nearly vertical orientations presented at random locations in the PGO layer. These patterns were chosen based on the experiments of Johnson and Morton (1991). (*b-c*) The LGN and V1 sheets compute their responses based on this input. Initially, FSA neurons will respond to any activity in their receptive fields (*e*), but after training only neurons with closely matching RFs respond (*f*). Through self-organization, the FSA neurons develop RFs (*d*) that are selective for a range of V1 activity patterns like those resulting from the three-dot stimuli. The RFs are patchy because the weights target specific orientation blobs in V1. This match between the FSA and the local self-organized pattern in V1 would be difficult to ensure without training on internally generated patterns. Plots (*g*) and (*h*) show the afferent weights for every third neuron in the FSA. All neurons develop roughly similar weight profiles, differing primarily by the position of their preferred stimuli on the retina and by the specific orientation blobs targeted in V1. The largest differences between RFs are along the outside border, where the neurons are less selective for three dot patterns. Overall, the FSA develops into a face detection map, signaling the location of facelike stimuli.

Unfortunately, such a formula is difficult to calculate in practice, because the order of presentation of the stimuli in newborn experiments is not generally known. As a result, the average or median value of patterns in recent history is not available. Furthermore, the numerical preference values computed in this way will differ depending on the specific set of patterns chosen, and thus will be different for each study.

Instead, for the simulations below I will use a categorical approach inspired by Cohen (1998) that should yield similar results. Specifically, I will assume that when two stimuli both activate the model FSA, 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, I will assume that the newborn will prefer the pattern that produces FSA activity, unless the V1 activity is vastly greater than for typical patterns (as it is e.g. for a checkerboard pattern). Using these guidelines, the computed model preferences can be validated against the newborn's looking preferences, to determine if the model shows the same behavior as the newborn.

7.3 Face preferences after prenatal learning

This section presents the model's response to schematic patterns and real images after it has completed training on the internally generated patterns. The response to each schematic pattern is compared to behavioral results in infants, and the responses to real images constitute predictions for future experiments.

7.3.1 Schematic patterns

Figures 7.4 and 7.5 show that the model responses match the measured stimulus preference of newborns remarkably well, with the same relative ranking in each case where infants have shown a significant preference between schematic patterns. These figures and the rankings are the main result from this simulation. I will next analyze each of the categories of preference rankings, to show what parts of the model underlie the pattern preferences.

Most non-facelike patterns activate only V1, and thus the preferences between those patterns are based only on the V1 activity values (figure 7.4a,e-i; 7.5f-i). Patterns with numerous high-contrast edges have greater V1 response, which explains why newborns would prefer them. These preferences are in accord with the simple LSM model, because they are based only on the early visual processing (section 3.3.1).

Face-like schematic patterns activate the FSA, whether they are realistic or simply patterns of three dots (figure 7.4b-d; 7.5a-d). The different FSA activation levels reflect the level of V1 response to each pattern, not the precise shape of the face pattern. Again the responses explain why newborns would prefer patterns like the three-square-blob face over the oval-blob face, which has a lower edge length. The preferences between these patterns are also compatible with the LSM, because in each case the V1 responses in the model match newborn preferences.

The comparisons between facelike and non-facelike patterns show how HLISSOM predictions differ from the LSM. HLISSOM predicts that the patterns that activate the FSA would be preferred over those activating only V1, except when the V1 response is highly anomalous. Most V1 responses are very similar, and so the patterns with FSA activity should be preferred over most of the other patterns, as found in infants. However, the checkerboard pattern (7.4a) has nearly three times as much V1 activity as any other similar pattern. Thus HLISSOM explains why the checkerboard would be preferred over other patterns, even ones that activate the FSA.
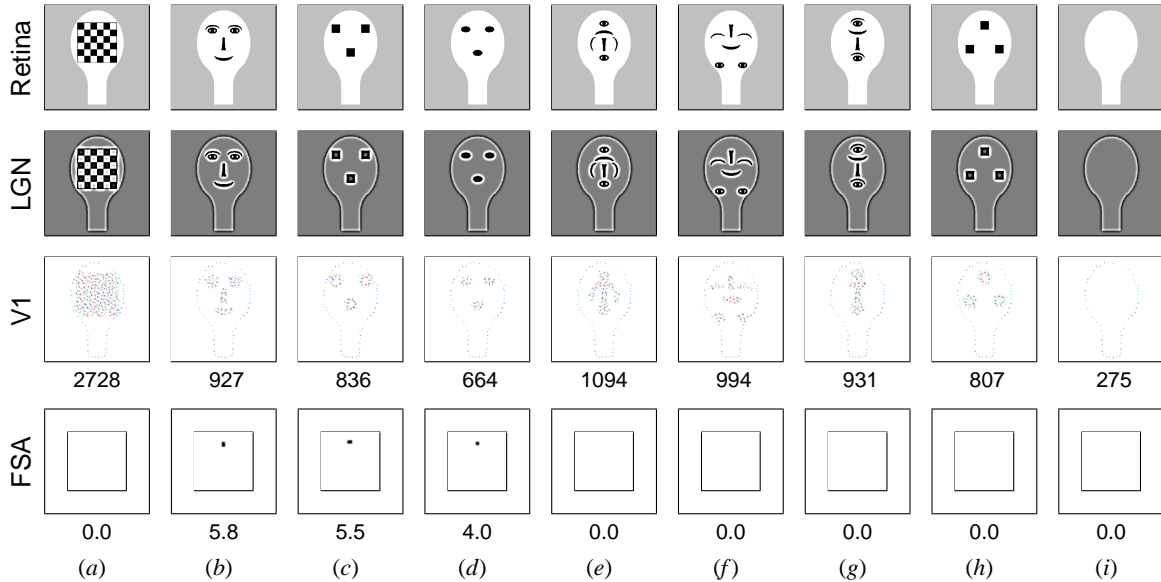
Figure 7.4: **Human newborn and model response to Goren et al.'s (1975) and Johnson et al.'s (1991) schematic images.** The Retina row along the top shows a set of patterns as they are drawn on the photoreceptor sheet. These patterns have been presented to newborn human infants on head-shaped paddles moving at a short distance (about 20cm) from the newborn's eyes, against a light-colored ceiling. In the experimental studies, the newborn's preference was determined by measuring the average distance his or her eyes or head tracked each pattern, compared to other patterns. Below, $x>y$ indicates that image x was preferred over image y under those conditions. Goren et al. (1975) measured infants between 3 and 27 minutes after birth. They found that $b>f>i$ and $b>e>i$. Similarly, Johnson et al. (1991), in one experiment measuring within one hour of birth, found $b>e>i$. In another, measuring at an average of 43 minutes, they found $b>e$, and $b>h$. Finally, Johnson and Morton (1991), measuring newborns an average of 21 hours old, found that $a>(b,c,d)$, $c>d$, and $b>d$. The HLISSOM model has the same preference for each of these patterns, as shown in the images above. The second row shows the model LGN activations (ON minus OFF) resulting from the patterns in the top row. The third row shows the V1 activations, with the numerical sum of the activities shown underneath. If one unit were fully active, the sum would be 1.0; higher values indicate that more units are active. The bottom row shows the settled response of the FSA. Activity at a given location in the FSA corresponds to a facelike stimulus at that location on the retina, and the sum of this activity is shown underneath. The images are sorted left to right according to the preferences of the model. The strongest V1 response by nearly a factor of three is to the checkerboard pattern (*a*), which explains why the newborn would prefer that pattern over the others. The facelike patterns (*b-d*) are preferred over patterns (*e-i*) because of activation in the FSA. The details of the facelike patterns do not significantly affect the results — all of the facelike patterns (*b*)–(*d*) lead to FSA activation, generally in proportion to their V1 activation levels. 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 provides a detailed computational explanation for why these behaviors occur.

78
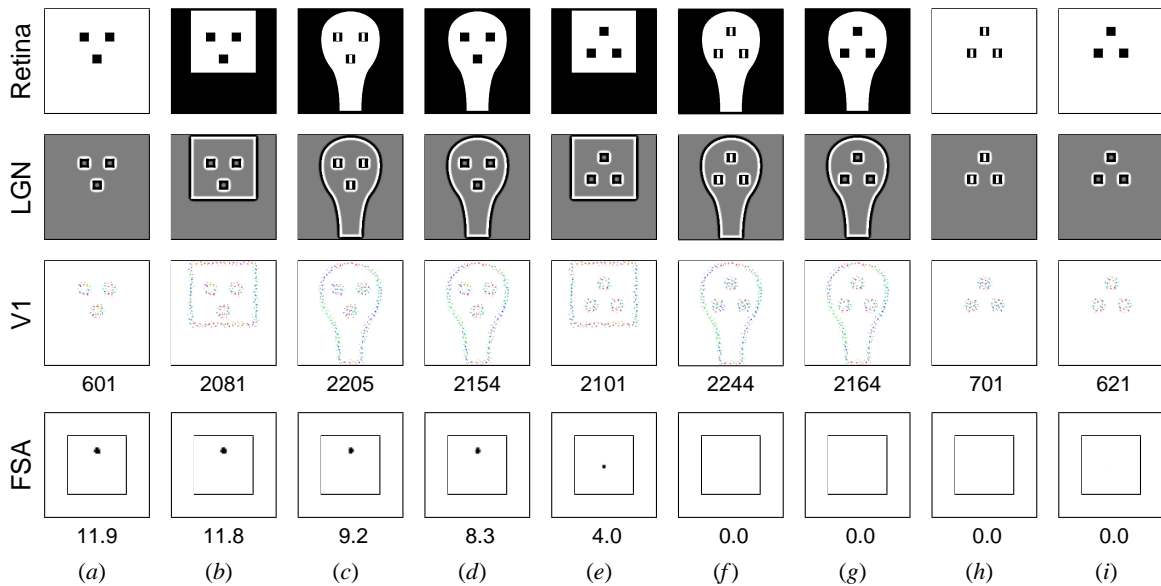
Figure 7.5: **Response to schematic images from Valenza et al. (1996) and Simion et al. (1998a).** Valenza et al. measured preference between static, projected versions of pairs of the schematic images in the top row, using newborns ranging from 24 to 155 hours after birth. They found the following preferences: $d>f$, $d>g$, $f>g$, and $h>i$. Simion et al. similarly found a preference for $d>g$ and $b>e$. The LGN, V1, and FSA responses of the model to these images are displayed here as in figure 7.4, and are again sorted by the model's preference. In all cases where the newborn showed a preference, the model preference matched it. For instance, the model FSA responds to the facelike pattern (*d*) but not to the inverted version (*g*). Patterns that closely match the newborn's preferred spatial frequencies (*f*,*h*) caused a greater V1 response than their corresponding lower-frequency patterns (*g*,*i*). Some non-facelike patterns with high-contrast borders can cause spurious FSA activation (*e*), because part of the border completes a three-dot pattern. Such spurious responses did not affect the predicted preferences, because they are smaller than the genuine responses.

Because the RFs in the model are only a rough match to the schematic patterns, the FSA can have spurious responses to patterns that do not look like faces. For instance, the inverted three-dot pattern in Figure 7.5*e* activated the FSA, because part of the square outline filled in for the missing third dot of an upright pattern. Figure 7.6 shows that such spurious responses should be expected with inverted patterns, even if neurons prefer upright patterns. This result may explain why some studies have not found a significant difference between upright and inverted three-dot patterns (e.g. Johnson and Morton 1991). Because of the possibility of such spurious effects, future studies should include additional controls besides inverted three-dot patterns.

Interestingly, the model also showed a clear preference in one case where no significant preference was found in newborns (Simion et al. 1998a): for the upright 3-blob pattern with no face outline (figure 7.5*a*), over the similar but inverted pattern (7.5*i*). The V1 responses to both patterns are similar, but only the upright pattern has an FSA response, and thus the model predicts that the upright pattern would be preferred.
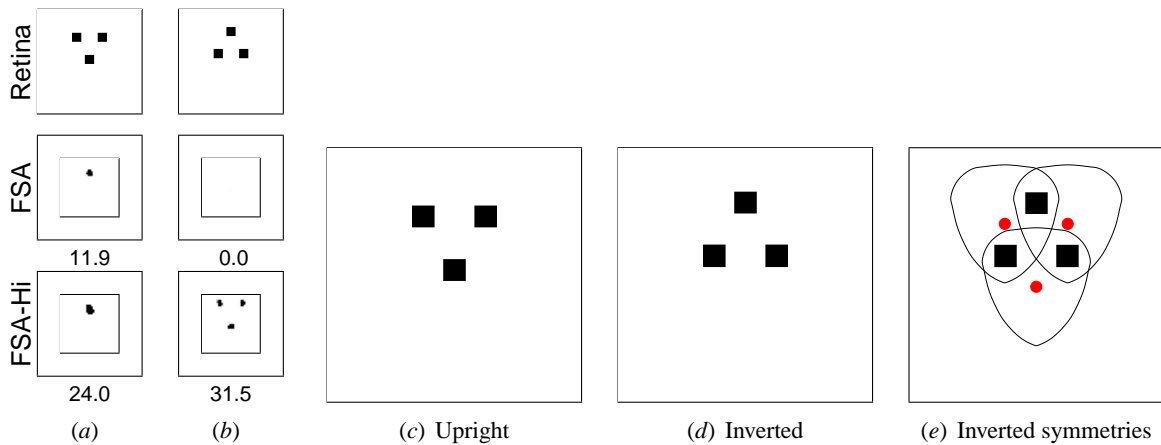
Figure 7.6: **Spurious responses with inverted three-dot patterns.** Several studies have used an inverted three-dot pattern as a non-facelike control for an upright three-dot pattern (e.g. Johnson and Morton 1991; Simion et al. 1998a; Valenza et al. 1996). However, the results here show that this pattern does not make a good control, because of the many axes of symmetry of a three-dot pattern. These symmetries may explain why Johnson and Morton (1991) found no significant difference between the upright and inverted patterns, while Simion et al. (1998a) and Valenza et al. (1996) found that the upright pattern was preferred. The FSA row above is the same as in figure 7.5, and shows that HLISSOM prefers the facelike upright pattern to the control. However, the preference is sensitive to the value of the FSA threshold δ and the FSA input scale γ_A. For instance, if γ_A is increased by 30%, the model FSA will respond more strongly to the inverted pattern (row FSA-Hi). As expected, the inverted pattern is not as good a match for any single neuron's weights, so the blobs of FSA activity are always smaller for the inverted pattern. However, with a high enough γ_A, the FSA responds in three different places (*b*), compared to only one for the upright pattern (*a*). These FSA responses resemble an upright face pattern, but the resemblance is merely coincidental because the retinal stimulus was inverted. Together the three small responses outweigh the single larger response, assuming that the total sum (and not the peak) activity is what determines preferences. Figure (*e*) shows that the three smaller responses are due to matching any two out of the three possible blobs. In (*e*) I have marked the FSA responses on top of the retinal pattern as three small dots. I have also drawn an outline around each of the possible upright three-dot patterns that these responses represent. That is, each response represents one of the three overlapping upright patterns that share two dots with the inverted pattern. As shown in later figures, similar spurious responses can occur in real images when only one or two facial features are matched, if other image features (such as parts of the face or hair outline) form a partial match for the missing features. For the other HLISSOM results with schematics, I set γ_A to a value low enough to prevent such spurious responses, which ensures that FSA neurons respond only to patterns that are a good match to their (upright) RFs. For humans, the γ_A value represents the state of contrast adaptation at a given time, which will vary depending on the recent history of patterns seen (cf. Albrecht et al. 1984; Turrigiano 1999). Thus these results suggest that infants will have no preference (or will prefer the inverted pattern) if they are tested on the high-contrast schematic patterns while being adapted to the lower contrast levels typical of the environment. For instance, an experimental protocol where there is a long time between schematic presentations could have such an effect. Because such adaptation is difficult to control in practice, the inverted pattern is a problematic comparison pattern – negative results like those of Johnson and Morton (1991) may be due to temporary contrast adaptation instead of genuine, long-term pattern preferences.

This potentially conflicting result may be due to postnatal learning rather than capabilities at birth. As will be shown in the next chapter, postnatal learning of face outlines can have a strong effect on FSA responses. The newborns in the Simion et al. study were already 1–6 days old, and e.g. Pascalis et al. (1995) showed that newborns within this age range have already learned some of the features of their mother's face outline. Thus the HLISSOM model predicts that if younger newborns are tested, they will show a preference for upright patterns even without a border. Alternatively, the border may satisfy some minimum requirement on size or complexity for patterns to attract a newborn's interest. If so, the HLISSOM procedure for deriving a behavioral response from the model response (section 7.2.3) would need to be modified to include such a constraint.

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.

7.3.2   Real face images

Most researchers testing newborns with schematic patterns assume that the responses to schematics are representative of responses to real faces. However no experiment has yet tested that assumption by comparing real faces to similar but non-facelike controls. HLISSOM makes testing real faces practical, which is an important way to determine the significance of the behavioral data. The other model for newborn face preferences has not yet been tested with natural images (Acerra et al. 2002), but the results reported here can be compared with future tests of that model. The examples in figure 7.7 show that the prenatally trained HLISSOM model works remarkably well as a face detector for natural images. The response is similar across a range of size scales and viewpoints, as shown in figure 7.8. This range is sufficient to account for typical newborn experiments, which have used a small range of size scales and only frontal views.

The face detection performance of the map was tested quantitatively using two image databases: a set of 150 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; figure 7.7*a* shows an example of a frontal view, and other views are shown in figure 7.8. 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 analyze 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
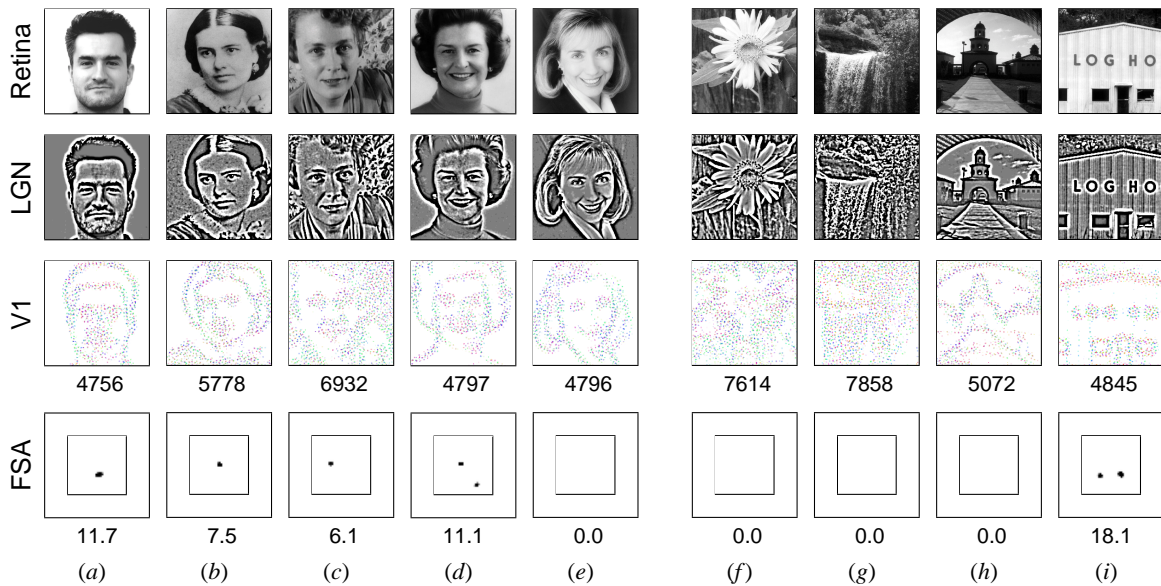
Figure 7.7: **Model response to natural images.** The top row shows a sample set of photographic images. The LGN, V1, and FSA responses of the model to these images are displayed as in figures 7.4 and 7.5. The FSA is indeed activated at the correct location for most top-lit faces of the correct size and orientation (e.g. *a-d*). Image (*a*) is an example of one frontal view from the the Achermann (1995) database, which also included the other viewpoints shown in figure 7.8. For this database, 88% of the faces resulted in FSA activity in the correct location. Just as important, the network is not activated for most natural scenes (*f-g*) and man-made objects (*h*). For example, the FSA responded to only 4.3% of 348 presentations of landscapes and other natural scenes from the National Park Service (1995). 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 and mouth contours (*e*). The model predicts that newborns would show similar responses if tested. Credits: (*a*) copyright 1995 Bernard Achermann, (*b-e*) public domain; (*f-i*) copyright 1999-2001, James A. Bednar.

had spurious responses in 27% (40/150) of the face images, i.e. responses in locations other than the center of the face. Nearly half of the spurious responses were from the less-selective neurons that line the outer border of the model FSA (see figure 7.3*h*); these responses can be ignored because they would be absent in a model of the entire visual field. Most of the remaining spurious responses resulted from a genuine V1 eye or mouth response plus V1 responses to the hair or jaw outlines. If present in humans, such responses would actually serve to direct attention to the general region of the face, and thus contribute to face preferences, 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 were in image regions that coincidentally had a triangular arrangement of three contour-rich areas, surrounded by smooth
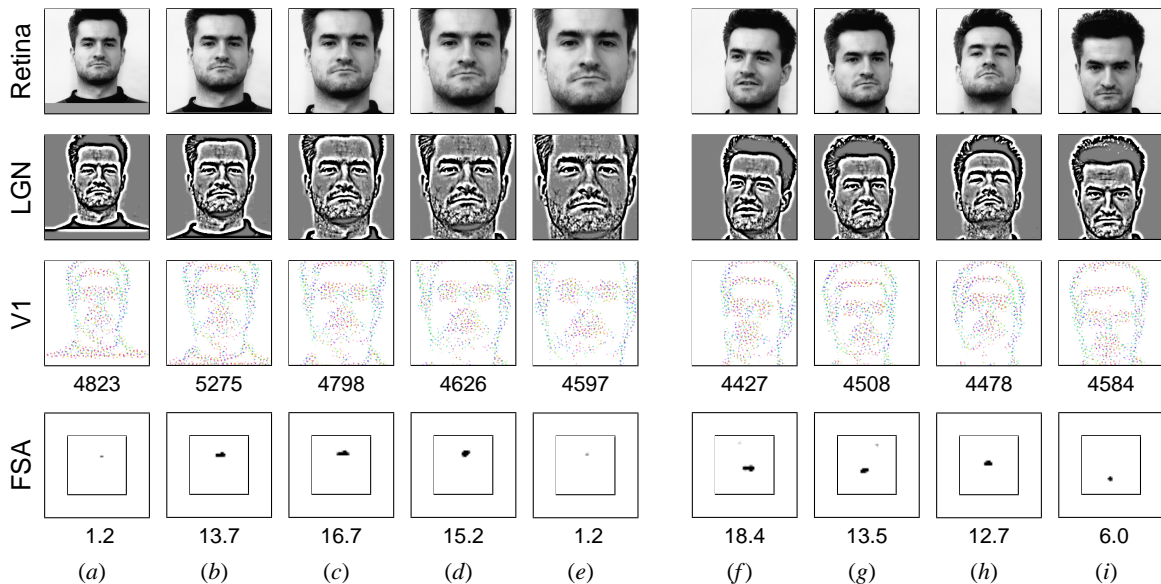
Figure 7.8: **Variation in response with size and viewpoint.** The three-dot training pattern of HLISSOM matches most closely to a single size scale and an upright frontal view. However, it also responds to a range of other sizes (*a-e*). Here the responses of the model are again displayed as in figures 7.4-7.7. Newborns have only been tested with a small range of sizes, and the range of patterns to which HLISSOM responds is sufficient to explain those, even using a single-sized training pattern. The network is also insensitive to moderate changes in viewpoint (*f-i*), responding in the correct FSA location to 88% of this set of 150 images consisting equally of front, left, right, up, and down views. Most of these viewpoints give similar responses, with the biggest difference being that 100% of the faces looking upwards were detected correctly, but only 80% of those looking downward were. Overall, HLISSOM predicts that newborns will respond to real faces even with moderate variation of sizes and viewpoints, as long as they are lit from above like these images.

shading.

In summary, the FSA responds to most top-lit human faces of about 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.

7.3.3 Effect of training pattern shape

The preceding sections show that a model trained on a triangular arrangement of three dots can account for face preferences at birth. This pattern was chosen based on Johnson and Morton's hypothesis that a hard-wired region responding to this pattern might explain newborn preferences. However, the actual shape of most internally generated activity in humans is unknown, and like retinal waves, the shape may not be very precisely controlled in general. Thus it is important to test other possible training pattern shapes, to see what range of patterns can produce similar results.

Accordingly, I ran a series of simulations using a set of nine patterns that I chose to be similar in overall size and shape to the three-dot patterns. The specific patterns are shown in the **Retina** row of figure 7.9. Matching the size is crucial to make the networks comparable, because only a single sized training pattern is used in each network.

To make training and testing so many networks practical, I developed a simpler model without V1 that requires much less time and memory for each simulation (13 megabytes of memory and 40 minutes of training time for self-organization). Without V1, the model does not provide numerical preferences between the non-facelike patterns (i.e., those that activate V1 but not the FSA), but activity in the FSA allows facelike patterns to be distinguished from other types, which is sufficient to measure the face selectivity of each network.

Specifically, the face selectivity S_{F} was defined as the proportion of the average response $\overline{\eta_{\mathrm{F}}}$ to a set of facelike patterns, out of the response to those patterns and the average responses $\overline{\eta_{\mathrm{N}}}$ to a set of non-facelike control patterns:

$$S_{\mathrm{F}} = \frac{\sum \overline{\eta_{\mathrm{F}}}}{\sum \overline{\eta_{\mathrm{F}}} + \sum \overline{\eta_{\mathrm{N}}}}. \tag{7.2}$$

A value of 1.0 indicates a strong preference for facelike patterns. It means that of the patterns tested, only the facelike patterns caused any FSA response. Values less than 0.5 indicate preference for non-facelike schematics. To ensure that the comparisons between networks were fair, the sigmoid activity threshold (δ in equation 4.4 on page 40) was set to maximize the face selectivity of each network. That is, δ was set to the minimum value for each network at which the FSA would respond to every facelike pattern. The upper threshold was then set to $\beta = \delta + 0.48$, for consistency. If there was no response to any non-facelike pattern with these settings, S_{F} would be 1.0, the maximum. δ was set separately for schematics and real images, again to maximize the reported selectivity of each type.

Figure 7.9 shows that even with such a liberal definition of face selectivity, not all patterns result in preferences for facelike schematics and real faces. Furthermore, several different patterns result in high face selectivity. As expected, the results vary most strongly for schematic test images (row **Schematics**). The schematics are all closely matched for size, differing only by the patterns within the face, and thus the results depend strongly on the specific training pattern. Of those resulting in face selectivity, training patterns 7.9*a-e* (three dots, dots and bars, bars, and open triangles) have nearly equivalent selectivity, although the three dot pattern 7.9*a* has the highest.

The results were less variable on real test images, because real faces differ more from objects than the schematic faces differ from other schematic patterns. All training patterns that matched either the overall size of the test faces (7.9*a-f*), or at least the eye spacing (7.9*h-i*), led to face selectivity. In contrast, the single dot pattern (7.9*g*) does not lead to face preferences. Although it is a good match to the eyes and mouth regions in the real faces, it is also a good match to many features in object images. Thus even though most patterns that are matched for size will provide face preferences under these conditions, general patterns like a single eye-sized dot are not sufficient.
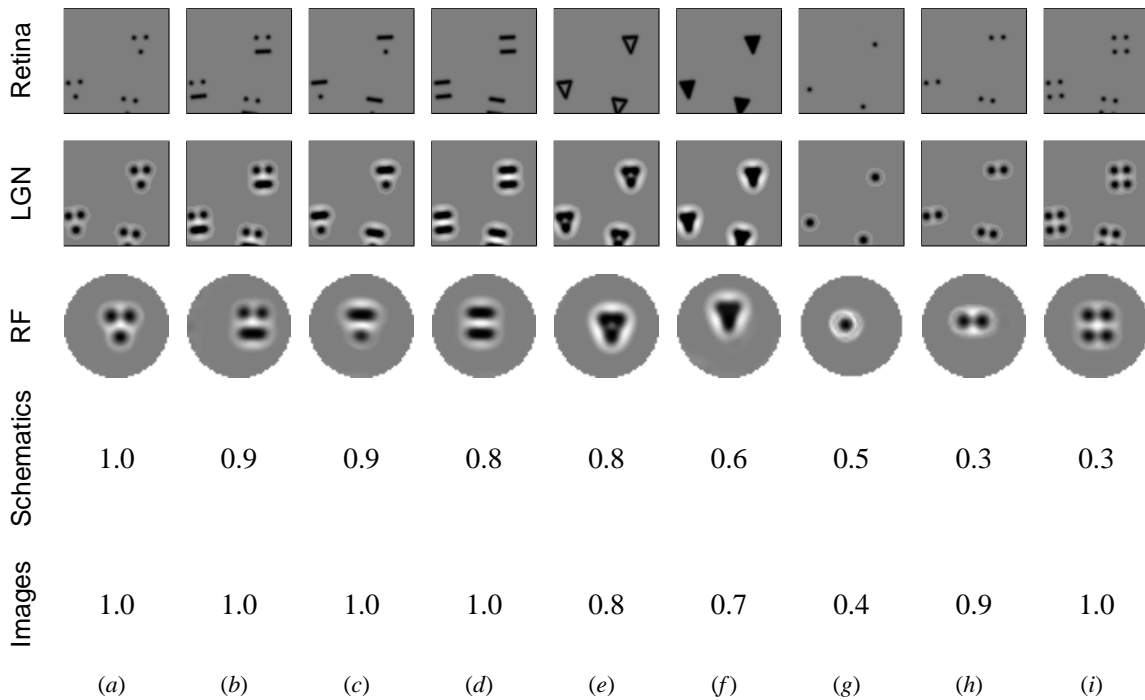
Figure 7.9: **Effect of the training pattern on face preferences.** This figure shows results from nine matched face detection simulations using easier-to-simulate networks without V1 (photoreceptors, LGN, and FSA only). The Retina row shows examples of the training patterns as they were drawn on the photoreceptors in the first iteration of each simulation. At a given iteration, each simulation presented a different pattern at the same locations and orientations. All other training parameters were also the same except for the FSA's γ_A, which was set manually so that each simulation would have the same amount of activity per iteration of training, despite differences in the pattern sizes and shapes. The LGN row shows the LGN response to the retinal input, which forms the input to the FSA. The RF row shows a sample FSA receptive field after self-organization; other neurons learned similar RFs. In each case the HLISSOM network learns FSA RFs similar to the LGN representation of the input patterns. Of course, the FSA RFs are no longer patchy, because they no longer represent the patchy V1 activities. The two numerical rows quantify the face selectivity of each network. The Schematics row shows the selectivity for facelike schematics (figure 7.4*b-d*) relative to non-facelike schematics (figure 7.4*e-i*). The Images row shows the selectivity for the six face images from figure 8.2 on page 90, relative to the six comparable object images in the same figure. The results show that different training patterns gave rise to different selectivities. Pattern (*g*) leads to equal responses for both facelike and non-facelike schematics (selectivity of 0.5), and (*h-i*) have a greater overall response to the *non*-facelike schematics (selectivity lower than 0.5). Thus not all training patterns are sufficient to explain preferences for schematic faces, even if they match some parts of the face. Similarly, the single dot pattern (*g*) has a selectivity below 0.5 for real faces, indicating a stronger response for the objects than for real faces. The other training patterns are all matched either for size with the real faces, or match at least two parts of the face, and thus have face selectivities larger than 0.5 for real images. Overall, the shape of the training pattern is clearly important for face selectivity, both for schematics and real faces, but it need not be controlled very tightly to result in face-selective responses.

Overall, these results show that matching the large-scale pattern size provides selectivity for faces. Further matching the pattern shape provides further selectivity, which can be tested with the schematic patterns. Of all training patterns tested, the three-dot pattern provides the most selectivity, while simply matching a single low-level feature like eye size is not enough to provide face preferences. Because of its high face selectivity, the three-dot pattern is a good default choice for pattern generation simulations, so that the predictions will most clearly differ from those of other models.

7.4 Discussion

The HLISSOM simulations show that internally generated patterns and self-organization can together account for newborn face detection. Importantly, predictions from the HLISSOM model differ from those of other recent models of newborn face preferences. These predictions can be tested by future experiments.

One easily tested prediction of HLISSOM is that newborn face preferences should not depend on the precise shape of the face outline. The Acerra et al. (2002) model (section 3.3.2) makes the opposite prediction, because in that model preferences arise from precise spacing differences between the external border and the internal facial features. HLISSOM also predicts that newborns will have a strong preference for real faces (e.g. in photographs), while I argue that the Acerra et al. model predicts only a weak preference for real faces, if any.

Experimental evidence to date cannot yet decide between the predictions of these two models. For instance, Simion et al. (1998a) did not find a significant schematic face preference in newborns 1–6 days old without a contour surrounding the internal features, which is consistent with the Acerra et al. model. However, the same study concluded that the shape of the contour "did not seem to affect the preference" for the patterns, which would not be consistent with the Acerra et al. model. As discussed earlier, younger newborns may not require any external contour, as HLISSOM predicts, until they have had postnatal experience with faces. Future experiments with younger newborns should compare model and newborn preferences between schematic patterns with a variety of border shapes and spacings. These experiments will either show that the border shape is crucial, as predicted by Acerra et al., or that it is unimportant, as predicted by HLISSOM.

The HLISSOM model also makes predictions that differ from those of the Simion et al. (2001) "top-heavy" model (reviewed in section 3.3.3). The top-heavy model predicts that any face-sized border that encloses objects denser at the top than the bottom will be preferred over similar schematic patterns. HLISSOM predicts instead 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. These two models represent very different explanations of the existing data, and thus testing such patterns should offer clear support for one model over the other.

On the other hand, many of the predictions of the fully trained HLISSOM model are similar to those of the hardwired CONSPEC model proposed by Johnson and Morton (1991). In fact, the reduced HLISSOM face preference network in section 7.3.3 (which does not include V1) can be seen as the first 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 2000a). The primary functional difference between the trained HLISSOM network and CONSPEC is that only cortical regions contain face-selective neurons 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). If future brain imaging studies do discover face-selective neurons in subcortical areas in newborns, HLISSOM will need to be modified to include such areas. Yet the key principles would remain the same, because the subcortical regions may also be organized from internally generated activity. Thus experimental tests of HLISSOM vs. CONSPEC should focus on how the initial system is constructed, and not where it is located.

7.5 Conclusion

Internally generated patterns explain how genetic influences can interact with general adaptation mechanisms to specify and develop newborn face processing circuitry. 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. Chapter 8 will show that the prenatally trained map can learn from real images after birth, and that such learning can explain how face processing develops postnatally.

Chapter 8

Postnatal Development of Face Detection

As shown in the previous chapter, prenatal learning of internally generated patterns can lead to face preferences at birth. In this chapter I show that the prenatally trained system can learn from faces in real images. This learning process suggests novel explanations for experimental phenomena, such as the disappearance of schematic face preferences, and provides concrete predictions for future behavioral experiments. Together, the face processing simulations show that learning of activity patterns can explain how face preferences develop prenatally and postnatally in young infants.

8.1   Goals

The postnatal face detection simulations in this chapter will demonstrate that:

1. Prenatal training leads to faster and more robust learning after birth, compared to a system exposed only to environmental stimuli (section 8.3.1).

2. The decline in infants' response to schematic patterns in the periphery after one month may simply result from learning real faces, not from a shift to a separate system (section 8.3.2).

3. Learning of both facial features and outlines, holistically, can explain the development of a mother preference, including why it disappears when the outline is masked (section 8.3.3).

These simulations focus on peripheral vision, which is what most tests of newborn and young infant face preferences measure (Johnson and Morton 1991). However, as discussed in section 8.4, similar mechanisms can also explain the later development of face preferences in central vision. The results suggest that the delay in central vision is simply due to postnatal development of the retina, again without a shift to a separate cortical subsystem.

8.2   Experimental setup

Postnatal experiments in newborns have tested only basic face selectivity, not detailed preferences between patterns of different spatial frequencies like in the newborn tests. As a result, the reduced HLISSOM model (without V1) is sufficient to model postnatal learning. Using the reduced model makes it practical to simulate much larger FSA RFs, which will be crucial for simulating postnatal face outline learning. For these experiments, the model was first trained "prenatally" as shown in figure 7.9*a* on page 85.

8.2.1   Control condition for prenatal learning

To determine whether the prenatal patterns bias subsequent learning (goal 1 above), a control network was also simulated, without prenatally organized receptive fields. This control will be called the *naïve* network, because it models neurons that have not had experience with coherent activity patterns until after birth. The intent is to see if neurons that are initially face selective, due to prenatal training on internally generated patterns, will learn faces more robustly than neurons that are initially unselective and learn only from the environment postnatally.

So that the naïve and prenatally organized networks would match on as many parameters as possible, I constructed the naïve network from the prenatally trained network *post hoc* by explicitly resetting afferent receptive fields to their uniform-Gaussian starting point. This procedure removed the prenatally developed face selectivity, but kept the lateral weights and all of the associated parameters the same. The activation threshold δ for the naïve FSA network was then adjusted so that both networks would have similar activation levels in response to the training patterns; otherwise the parameters were the same for each network. This procedure ensures that the comparison between the two networks will be as fair as possible, because the networks differ only by whether the neurons have face-selective weights at birth. Figure 8.1 shows the state of each network just before postnatal learning.

8.2.2   Postnatal learning

The experiments in this chapter simulate gradual learning from repeated encounters of specific individuals and objects against different backgrounds, over the first few months of life. Figure 8.2 shows the people and objects that were used and figure 8.3 describes how the training images were generated. The prenatally trained and naïve networks were each exposed to the same random sequence of 30,000 images, so that the influence of prenatal training on postnatal learning will be clear.

The RF changes in these simulations over time were larger than in the postnatal learning experiments in chapter 6, because the weights were initially concentrated only in the center (figure 8.1). As the weights spread out over the receptive field, the strength of the neural responses to typical training inputs varied significantly. To compensate for these changes, I periodically adjusted

(*a*) Prenatally trained network   (*b*) Naïve network
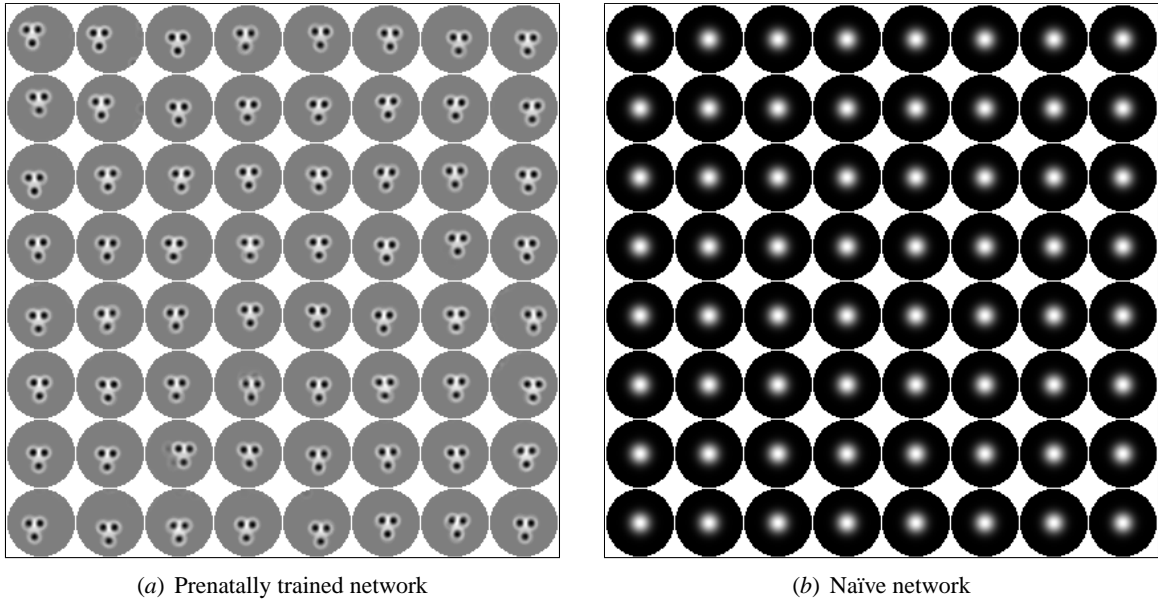
Figure 8.1: **Starting points for postnatal learning.** These plots show the RFs for every third neuron from the 24×24 array of neurons in the FSA. For the prenatally trained network (*a*), the RFs were visualized by subtracting the OFF weights from the ON. The result is a plot of the retinal stimulus that would most excite that neuron. As in chapter 7, the prenatally trained network consists of an array of roughly facelike RFs. In contrast, the neurons in the naïve network are initially uniformly Gaussian. The ON and OFF weights were identical, so only the ON weights are shown in (*b*). Later figures will compare the postnatal learning of each network.

Figure 8.2: **Postnatal learning source images.** Simulations of postnatal learning used the above person and object images (person images adapted from Rowley et al. 1998; objects are from public domain clip art and image collections.) As shown in figure 8.3, each of these items was presented at random locations in front of randomly chosen natural scenes.

90
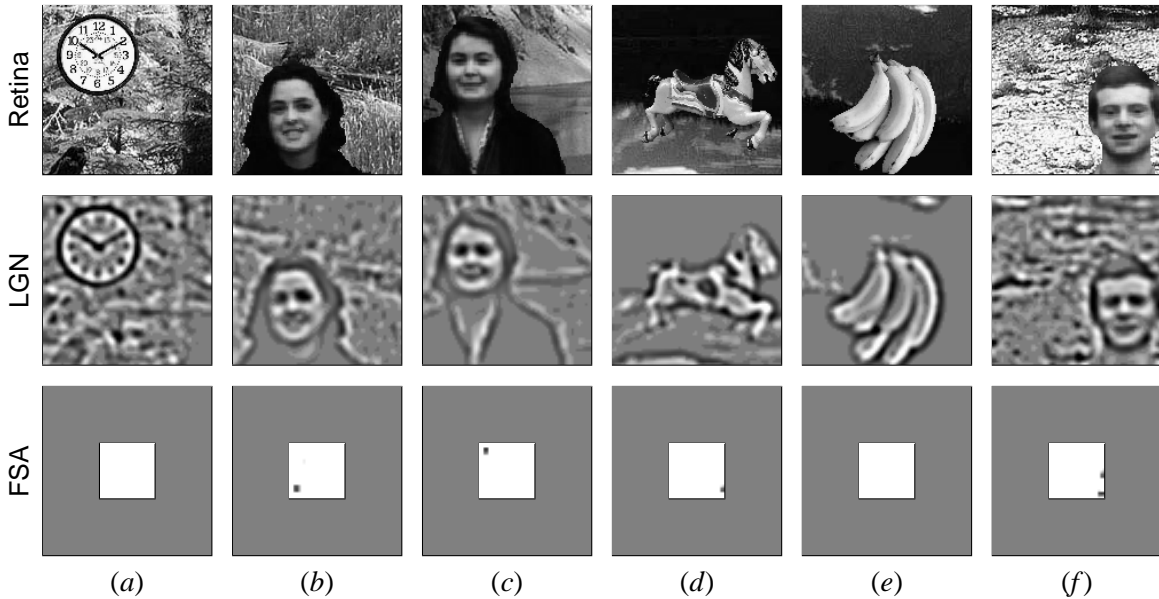
Figure 8.3: **Sample postnatal learning iterations.** The top row shows six randomly generated images drawn on the retinal photoreceptors at different iterations. Each image contains a foreground item chosen randomly from the images in figure 8.2. The foreground item was overlaid onto a random portion of an image from a database of 58 natural scenes (National Park Service 1995), at a random location and at a nearly vertical orientation (drawn from a normal distribution around vertical, with $\sigma = \pi/36$). The second row shows the LGN response to each of these sample patterns. The bottom row shows the FSA response to each pattern at the start of postnatal training. For the FSA, only neurons with complete receptive fields (those in the unshaded inner box) were simulated, because those in the gray area would have RFs cut off by the edge of the retina. The gray area shows the FSA area that corresponds to the same portion of the visual field as in the LGN and retina plots; with this representation, points in the FSA are mapped directly to the corresponding location in the LGN and retina. At the start of postnatal training, the FSA responds to groups of dark spots on the retina, such as the eyes and mouths in (*b-c,f*) and the horse's dark markings in (*d*); the location of the FSA activity corresponds to the position of the group of retinal patterns that caused the response. Subsequent learning in the FSA will be driven by these patterns of activity. The prenatal training biases the activity patterns towards faces, and so postnatal self-organization will also be biased towards faces.

the sigmoid threshold δ from equation 4.4 for each network. Without such compensation, the networks eventually fail to respond to any training input, because the fixed total amount of weights has become spread out over a much larger area than initially. The specific parameter values are listed in section A.4. Chapter 9 will discuss how this process can be simplified by extending HLISSOM with automatic mechanisms for setting the threshold based on the recent history of inputs and responses, as found in many biological systems (Turrigiano 1999).

8.2.3 Testing preferences

In the previous chapter, pattern preferences were measured in a fully organized network that had similar RFs across the visual field. With such a uniform architecture, image presentations generally obtained similar responses at different retinal locations. In this chapter, the network pattern preferences will be measured periodically during early postnatal learning, before the network has become fully uniform.

To allow clear comparisons in this more-variable case, preferences will be measured here by presenting input stimuli at 25 different retinal locations, averaging the results, and computing the statistical significance of the difference between the two distributions. All comparisons will use the Student's t-test, with the null hypothesis being that the network responds equally strongly to both stimuli. To match the psychological experiments, I will consider a calculated $p \leq 0.05$ to be evidence of a significant difference between responses to the two stimuli.

8.3 Results

8.3.1 Bias from prenatal learning

Figure 8.4 shows that with postnatal exposure to real images, both the naïve and prenatally trained networks develop RFs that are averages (i.e. prototypes) of faces and hair outlines. RFs in the prenatally trained network gradually become more face-selective, and eventually nearly all neurons become highly selective (figure 8.4*b*).

Postnatal self-organization in the naïve network is less regular, and the final result is often less face selective. For example, the postnatal network often develops neurons selective for the clock. The clock has a high-contrast border that is a reasonably close match to real face outlines, and thus the same naïve network neurons tend to respond to both the clock and to real faces during training. But the clock is a very weak match to the three-dot training pattern in the prenatally trained network, and so that network rarely develops clock-selective neurons. These results suggest that the prenatal training biases postnatal learning towards biologically relevant stimuli, i.e. faces.

8.3.2 Decline in response to schematics

Figure 8.5 shows that the HLISSOM model replicates the disappearance of peripheral schematic face preferences after one month (Johnson et al. 1991). In HLISSOM, the decrease results from the afferent weight normalization (equation 4.8). As the FSA neurons in HLISSOM learn the hair and face outlines typically associated with real faces, the connections to the internal features necessarily become weaker. Unlike real faces, the facelike schematic patterns match only on these internal features, not the outlines. As a result, the response to schematic facelike patterns decreases as real faces are learned. Eventually, the response to the schematic patterns approaches and drops below the fixed activation threshold (δ in equation 4.4). At that point, the model response is no longer

(a) Prenatally trained network

(c) Naïve network

| 0 | 6,000 | 10,000 | 12,000 | 16,000 | 20,000 | 30,000 |

(b) Snapshots of typical prenatally trained RFs

| 0 | 6,000 | 10,000 | 12,000 | 16,000 | 20,000 | 30,000 |

(d) Snapshots of naïve network RFs
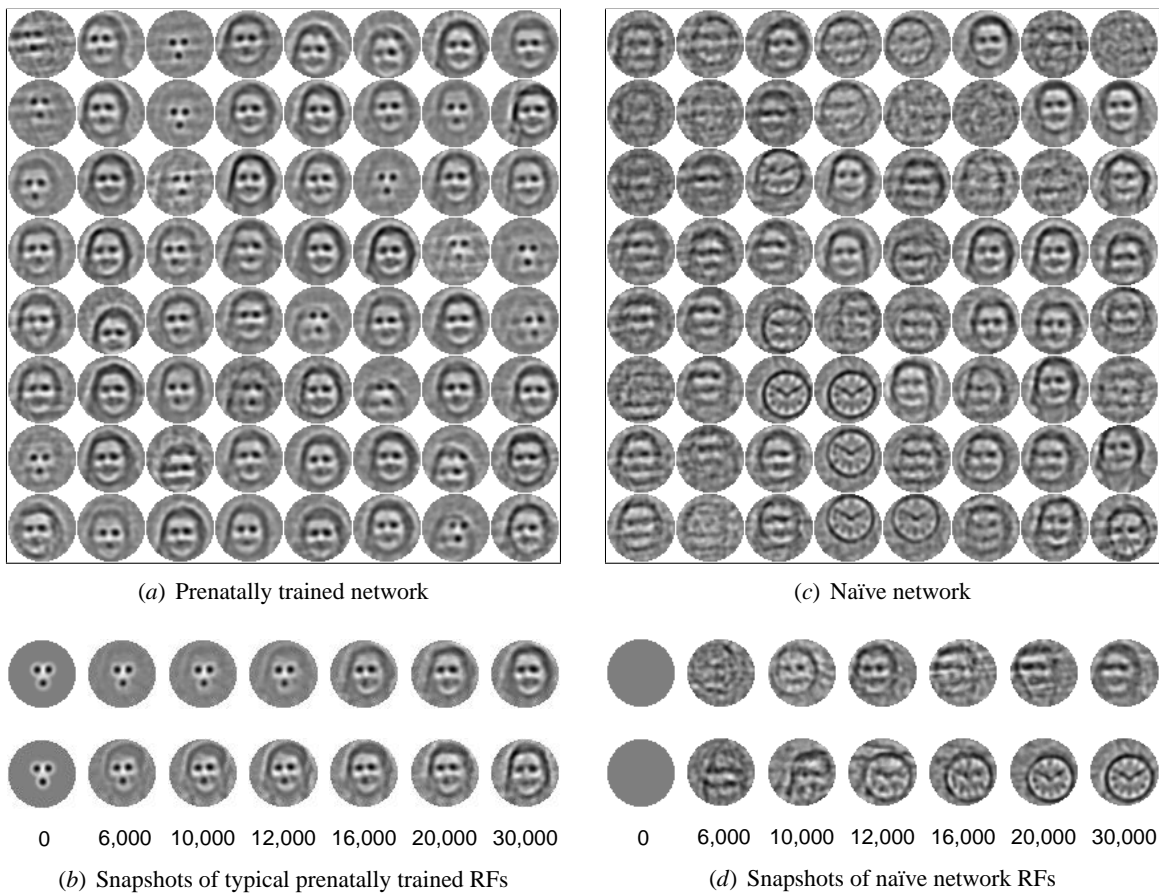
Figure 8.4: **Prenatal patterns bias postnatal learning in the FSA.** Plots (*a*) and (*c*) show the final RFs for every third neuron from the 24×24 array of neurons in the FSA, visualized as in figure 8.1(*a*). As the prenatally trained network learns from real images, the RFs morph smoothly into prototypes, i.e. representations of average facial features and hair outlines (*b*). By postnatal iteration 30,000, nearly all neurons have learned face-like RFs, with very little effect from the background patterns or non-face objects (*a*). Postnatal learning is less uniform for the naïve network, as can be seen in the RF snapshots in (*d*). In the end, many of the naïve neurons do learn face-like RFs, but others become selective for general texture patterns, and some become selective for objects like the clock (*c*). Overall, the prenatally trained network is biased towards learning faces, while the initially uniform network more faithfully represents the environment. Thus prenatal learning can allow the genome to guide development in a biologically relevant direction.

93
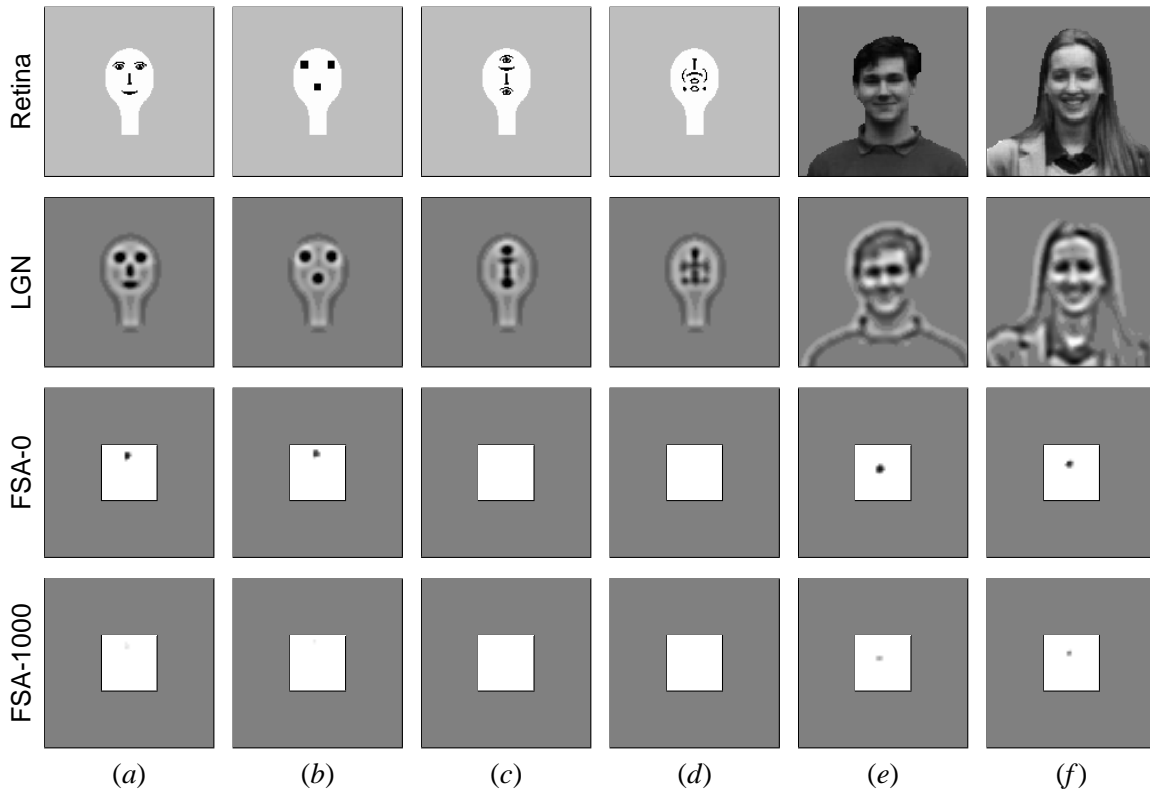
Figure 8.5: **Decline in response to schematic faces.** Before postnatal training, the prenatally trained FSA (third row from top) responds significantly more to the facelike stimulus (*a*) than to the three-dot stimulus (*b*; $p = 0.05$) or the scrambled faces (*c-d*; $p = 10^{-8}$). Assuming that infants attend most strongly to the stimuli that cause the greatest neural response, these responses replicate the schematic face preferences found by Johnson and Morton (1991) in infants up to one month of age. Some of the Johnson and Morton 1991 experiments found no significant difference between (*a*) and (*b*), which is unsurprising given that they are only barely significantly different here. As the FSA neurons learn from real faces postnatally, they respond less and less to schematic faces. The bottom row shows the FSA response after 1000 postnatal iterations. The FSA now rarely responds to (*a*) and (*b*), and the average difference between them is no longer significant ($p = 0.25$). Thus no preference would be expected for the facelike schematic after postnatal learning, which is what Johnson and Morton (1991) found for older infants, i.e. 6 weeks to 5 months old. The response to real faces also decreases slightly through learning, but to a much lesser extent (*e-f*). The response to real faces declines because the newly learned average face and hair outline RFs are a weaker match to any particular face than were the original three dot RFs. That is, the external features vary more between individuals than do the internal features, as can be seen in figure 8.2, and thus their average is not a close match to any particular face. Even so, there is only a comparatively small decrease in response to real faces, because real faces are still more similar to each other than to the schematic faces. Thus HLISSOM predicts that older infants will still show a face preference if tested with more-realistic stimuli, such as photographs.

94

higher for schematic faces (because there is no FSA response, and V1 responses are similar). In a sense, the FSA has learned that real faces typically have both inner *and* outer features, and does not respond when either type of feature is absent or a poor match to real faces.

Yet the FSA neurons continue to respond to real faces (as opposed to schematics) throughout postnatal learning (figure 8.5*e-f*). Thus the model provides a clear prediction that the decline in peripheral face preferences is limited to schematics, and that if infants are tested with sufficiently realistic face stimuli, no decline in preferences will be found.

This prediction is an important difference from the CONSPEC/CONLERN model, where the decline was explained as an effect of CONLERN maturing and beginning to inhibit CONSPEC. CONSPEC/CONLERN predicts that the decline in preferences for schematic faces also represents a decline in preference for real faces. HLISSOM predicts instead that an initially CONSPEC-like system is also like CONLERN, in that it will gradually learn from real faces. These divergent predictions can be tested by presenting real faces in the periphery to infants older than one month; HLISSOM predicts that the infant will prefer the real face to other non-facelike stimuli.

8.3.3   Mother preferences

Figure 8.6*a-b* shows that when one face (i.e. the mother) appears most often, the FSA response to that face becomes significantly stronger than to a similar stranger. This result replicates the mother preference found in infants a few days old (Bushnell 2001; Pascalis et al. 1995; Walton and Bower 1993; Walton, Armstrong, and Bower 1997; Bushnell et al. 1989).

Interestingly, figure 8.6*c-d* shows that the mother preference disappears when the hair outline is masked, which is consistent with Pascalis et al.'s claim that newborns learn outlines only. However, Pascalis et al. (1995) did not test the crucial converse condition, i.e. whether newborns respond when the facial features are masked, leaving only the outlines. Figure 8.6*e-f* shows that there is no response to the head and hair outline alone either. Thus this face learning is clearly *not* outline-only.

In the model, the decreased response with either type of masking results from holistic learning of *all* of the features typically present in real faces. As real faces are learned, the afferent weight normalization ensures that neurons respond only to patterns that are a good overall match to all of the weights, not simply matching on a few features. Many authors have argued that adults also learn faces holistically (e.g. Farah et al. 1998). These results suggest that newborns may learn faces in the same way, and predict that newborns will no prefer their mother when her hair outline is visible but her facial features are masked.

8.4   Discussion

The results in this chapter show that a prenatally trained map can explain how schematic face preferences will gradually decay, due to the learning of real faces. However, they do not address one
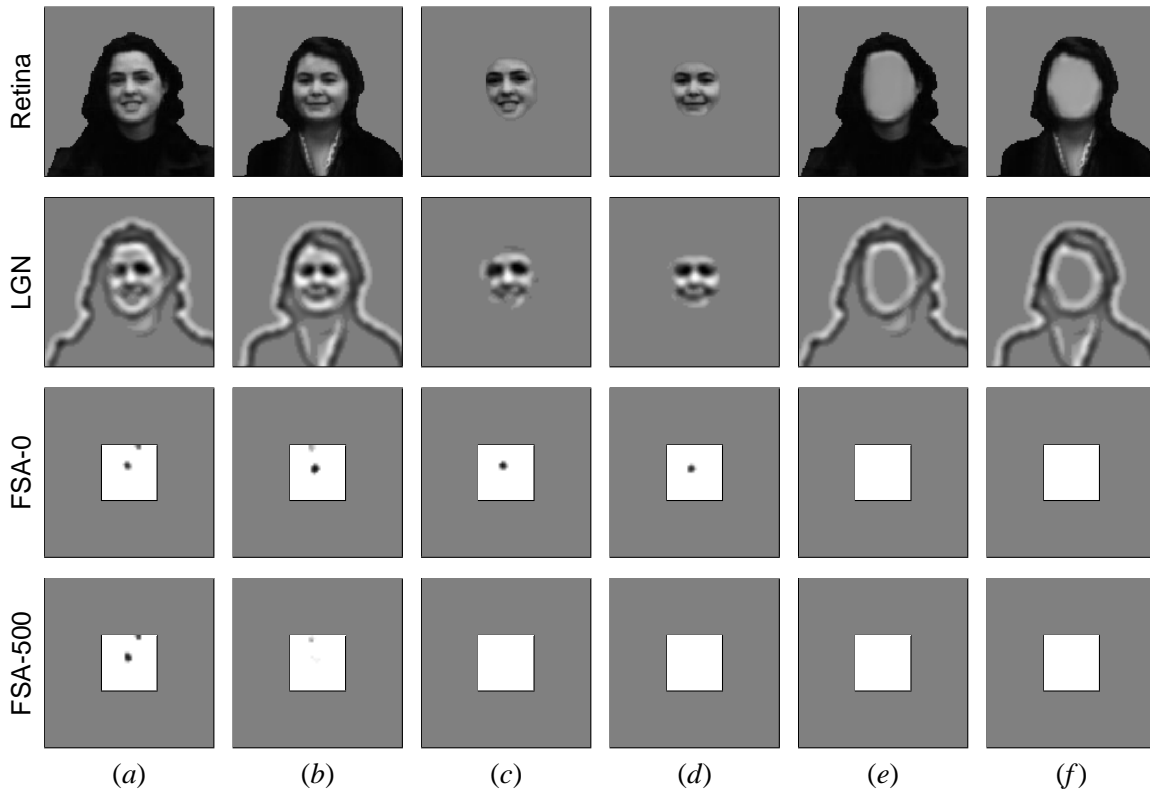
Figure 8.6: **Mother preferences depend on both internal and external features.** Initially, the prenatally trained FSA responds well to both women above (*a-b*; FSA-0), with no significant difference ($p = 0.28$). The response is primarily due to the internal facial features (*c-d*; FSA-0), although there are some spurious three-dot responses due to alignment of the hair with the eyes (*a-b*; top of FSA-0). Designating image (*a*) as the mother, I presented it in 25% of the postnatal learning iterations. (This ratio is taken from Bushnell 2001, who found that newborns look at their mother's face for an average of about one-fourth their time awake over the first few days.) Image (*b*), the stranger, was not presented at all during training. After 500 postnatal iterations, the response to the mother is significantly greater than to face (*b*) ($p = 0.001$). This result replicates the mother preference found by Pascalis et al. (1995) in infants 3–9 days old. The same results are found in the counterbalancing condition — when trained on face (*b*) as the mother, (*b*) becomes preferred ($p = 0.002$; not shown). After training with real faces, there is no longer any FSA response to the facial features alone (*c-d*), which replicates Pascalis et al.'s (1995) finding that newborns no longer preferred their mother when her face outline was covered. Yet contra Pascalis et al. (1995), we cannot conclude that what has been learned "has to do with the outer rather than the inner features of the face", because no preference is found for the face outline alone either (*e-f*). Thus face learning in HLISSOM is holistic. Face learning in adults is also thought to be holistic (Farah et al. 1998), and these results show that we do not need to assume that newborns are using a different type of face learning than adults.

interesting phenomenon: in central vision, preference for schematic faces is not measurable until 2 months of age (Maurer and Barrera 1981), and is gone by 5 months (Johnson et al. 1991) This time course is delayed relative to peripheral vision, where preferences are present at birth but disappear by 2 months.

As reviewed in section 3.3.5, Johnson and Morton (1991) propose two separate explanations for the peripheral and central declines in face preferences. That is, in the periphery the preferences disappear because CONLERN matures and inhibits CONSPEC, while in central vision they disappear because CONLERN learns properties of real faces and no longer responds to static schematic patterns. HLISSOM provides a unified explanation for both phenomena: a single learning system stops responding to schematic faces because it has learned from real faces.

Why, then, would the time course differ between peripheral and central vision? As Johnson and Morton acknowledged, the retina changes significantly over the first few months. In particular, at birth the fovea is much less mature than the periphery, and may not even be functional yet (Abramov, Gordon, Hendrickson, Hainline, Dobson, and LaBossiere 1982; Kiorpes and Kiper 1996). Thus schematic face preferences in central vision may be delayed relative to those in peripheral vision simply because the fovea matures later. A single cortical learning system like HLISSOM is thus sufficient to account for the time course of both central and peripheral schematic face preferences.

Central and peripheral differences may also have a role in mother preferences. In a recent study, Bartrip, Morton, and de Schonen (2001) found that infants 19–25 days old do not exhibit a significant preference for their mothers when either the internal features or external features are covered. This result partially confirms the predictions from section 8.3.3, although full confirmation will require tests with newborns only a few days old, as in Pascalis et al. (1995). Interestingly, Bartrip et al. also found that older infants, 35–40 days old, do show a mother preference, even when the external outline is covered. The gradual maturation of the fovea may again explain these later-developing capabilities. Unlike the periphery, the fovea contains many ganglia with small RFs, and which connect to cortical cells with small RFs (see e.g. Merigan and Maunsell 1993). These neurons can learn smaller regions of the mother's face, and their responses will allow the infant to recognize the mother even when other regions of the face are covered. Thus simple, documented changes in the retina can explain why mother preferences would differ over time.

8.5 Conclusion

The HLISSOM face detection simulations in this chapter and in chapter 7 show that internally generated patterns and a self-organizing system can together account for newborn face preferences, neonatal face learning, and longer term development of face detection. The results suggest simple but novel explanations for why newborn learning appears to depend on the face outline, and why the response to schematic faces decreases over time. Unlike other models, the same principles apply to

both central and peripheral vision, and the results differ only because the development of the retina is not uniform.

These explanations and simulation results lead to concrete predictions for future infant experiments. Over the first two months, the response to real faces in the periphery should continue even as response to schematics diminishes, and the mother preference of newborns should disappear when the facial features are masked. The results also suggest that internally generated patterns allow the genome to steer development towards biologically relevant domains, making learning quicker and more robust.

Chapter 9

Discussion and Future Research

The results in the previous chapters focused specifically on biological and psychological studies, but they open up future investigations in a number of basic and applied research fields. Most immediately, they suggest specific psychological experiments that can be run on human infants, to validate or refute the model. I summarize these experiments below. The model also suggests experiments with animals, which I will also describe. Next, I outline some extensions to the HLISSOM model that will allow new phenomena to be explained and make the model more useful in new domains. Finally, I will describe applications of the pattern generation approach to other fields, focusing on how it can be a general-purpose problem-solving method.

9.1   Proposed psychological experiments

Chapters 7 and 8 listed a number of concrete predictions derived from HLISSOM that can be tested in psychological experiments with human infants:

- Newborns will prefer real faces and face images over other similar stimuli that do not have three-dot patterns.

- Newborns will still prefer the facelike patterns from Valenza et al. (1996), even if they have an enclosing border that does not have the spacing assumed by the Acerra et al. (2002) model.

- Newborns will prefer a facelike pattern to one where the eyes have been shifted over to one side, contrary to the "top-heavy" explanation for face preferences.

- Newborns will no longer prefer their mother when her facial features are covered, just as when her hair outline is covered.

- Even though infants stop preferring schematic faces in the periphery by 2 months of age, they will continue to prefer sufficiently realistic stimuli, such as photographs or video of faces.

Nearly all of these experiments can be run using the same techniques already used in previous experiments with newborns. They will help determine how specific the newborn face preferences are, and clarify what types of learning are involved in newborn mother preferences.

9.2   Proposed experiments in animals

In addition to behavioral tests in infants, the HLISSOM results suggest a number of specific experiments to be run with laboratory animals.

9.2.1   Measuring internally generated patterns

The key to the pattern generation explanation of how sensory systems develop is the patterns themselves. HLISSOM shows that pattern shapes that are known, such as retinal waves, are sufficient to develop the types of circuitry seen in V1 before animals open their eyes. Other sources of internally generated activity would also suffice, if they consist of coherent patterns of activity, but none of these patterns have been measured in detail so far. Similarly, in the HLISSOM simulations a number of different patterns resulted in a preference for schematic faces and images of real faces. To understand how these patterns affect development, it will be crucial to measure in detail what kinds of spontaneous visual system activity are present in developing animals and during REM sleep. Recent advances in imaging equipment may make measurements of this type possible (Rector et al. 1997), which would allow the assumptions and predictions of the HLISSOM model to be tested directly.

Experimental studies with animals should also clarify how internally generated patterns differ across species, and whether these differences correspond to differences in their cognitive and sensory-processing abilities. One of the most important tests will be to modify the internal patterns in animals, testing to see if their brains develop differently in a systematic way. In the long term, detailed analysis of the pattern-generating mechanisms in the brainstem could provide clues to how the pattern generator is specified in the genome. These studies would also suggest ways for representing the generator effectively in a computational genetic algorithm, which is currently an open question. Higher level studies of genetics and phylogeny could also determine whether pattern generation was instrumental in the emergence of complex organisms with postnatal flexibility.

9.2.2   Measuring receptive fields in young animals

The results in chapter 6 show that different types of internally generated activity patterns will result in different types of V1 receptive fields. For instance, uniformly random noise tends to lead to four-lobed RFs, while discs alone lead to two-lobed RFs. Although orientation maps have been measured in young animals, very little data is available about what receptive fields these neurons might have.

Future experiments should try measure to measure the RFs in the same newborn animals where orientation maps are measured, so that it will be clear what types of RFs are present at each stage of orientation map development. This data will help to narrow down the range of plausible orientation map models, e.g. by rejecting either uniformly random noise or retinal waves as the source of the initial orientation preferences.

9.3   Proposed extensions to HLISSOM

The HLISSOM model introduces a number of significant extensions to earlier self-organizing models, which will be refined in future work. This section describes the future extensions expected to be the most productive.

9.3.1   Push-pull afferent connections

HLISSOM includes a divisive normalization term not present in simpler models (see equation 4.5). This change allows neurons to respond selectively to a wide range of contrasts in natural images. However, as a side-effect it also attenuates the response to closely spaced stimuli. That is, V1 can respond less to a high-frequency square-wave grating than to the same pattern with every other bar removed, because the normalization term penalizes any activity in the anatomically circular receptive fields that does not match the neuron's weights. In future work, it may be possible remove this unrealistic limitation by using a push-pull arrangement of weights rather than full-RF normalization (see Ferster 1994; Hirsch, Gallagher, Alonso, and Martinez 1998b; Troyer, Krukowski, Priebe, and Miller 1998). With a push-pull RF, cortical neurons receive both excitatory and inhibitory afferent input from different parts of the retina, rather than the purely excitatory input in HLISSOM and most other incremental Hebbian models. One difficulty with using push-pull weights is that the inhibitory weights need to connect to anti-correlated regions of the retina, rather than to the correlated regions that Hebbian learning connects. Thus either a new learning rule or a more complicated local circuit in the cortex will need to be developed so that push-pull weights can self-organize.

9.3.2   Threshold adaptation

One of the most important components of any detection or recognition system, and for neural network models in general, is the threshold or criterion used. For HLISSOM and RF-LISSOM, the parameter δ of the sigmoid function acts as a threshold for each neuron's response: if the inputs in a neuron's receptive field sum to a value below δ, the neuron does not activate. Above δ, the neuron activates linearly, until an upper threshold.

If set properly, the lower threshold allows the system to respond to appropriate inputs while ignoring low-level noise or other nonspecific activation. For instance, the threshold can allow an orientation-selective neuron to respond to stimuli near its preferred orientation, and not to other

orientations. If the threshold is set too high, the system will never respond to any input presented, and if the threshold is too low, the system will respond indiscriminately to all inputs. In terms of signal detection theory, the threshold is a balance between the false alarm rate and the false positive rate; in general this tradeoff is unavoidable (see e.g. Egan 1975 for more details). Thus it is important to make this tradeoff appropriately.

In HLISSOM and RF-LISSOM, the thresholds are set by hand through a process of trial and error. This procedure corresponds to adaptation over evolutionary timescales, which could presumably result in animals that are born with appropriate threshold settings. However, an automatic method of setting thresholds may be more realistic, and for at least four reasons such a mechanism would be very desirable : (1) it is time consuming for the modeler to find a good threshold value, (2) the threshold setting process is subjective, which often prevents rigorous comparison between different experimental conditions (e.g. to determine what features of the model are required for a certain behavior), (3) different threshold settings are needed even for different types of input patterns, depending on how strongly each activates the network (e.g. high-contrast schematic patterns vs. low-contrast real images), and (4) the optimal threshold value changes over the course of self-organization, because weights become concentrated into configurations that match typical inputs, increasing the likelihood of a response. The underlying difficulty is that the self-organizing weight pattern of a neuron interacts nonlinearly with the likelihood of seeing similar patterns in the input. Neither the weight pattern nor the environment is entirely predictable, and thus setting the threshold appropriately is not always a trivial task.

Adding automatic mechanisms for setting the thresholds would make the model significantly more complex mathematically. Even so, it may make the model much simpler to use in practice, particularly with real images. As reviewed by Turrigiano (1999), such "homeostatic regulation" mechanisms have recently been discovered in a variety of biological systems. Some of these regulatory mechanisms are similar to the weight normalization already used in HLISSOM (equation 4.8). Others more directly adjust how excitable a neuron is, so that its responses will cover a useful range (i.e., neither always on nor always off). Extending HLISSOM with additional homeostatic mechanisms will make it easier to use with different input pattern types, and will make it simpler to compare different experimental conditions rigorously.

9.4 Maintaining genetically specified function

The simulations in this thesis used two separate learning phases, prenatal and postnatal. This way, the influence of the internally generated and environmentally driven stimuli would be perfectly clear. Such a separation is a good model of spontaneous activity in the developing sensory areas, such as retinal waves, because the waves disappear at eye opening. But other activity, such as that during REM sleep, continues throughout development and adulthood. Thus internally generated patterns may also have a significant role beyond prenatal development.

Specifically, I propose that postnatal spontaneous activity is interleaved with waking experience to ensure that environment-driven postnatal development does not entirely overwrite the prenatal organization. These postnatal patterns may explain why there are limits to the learning of altered environments (as found by Sengpiel et al. 1999), and why the duration of REM sleep correlates so strongly with the degree of neural plasticity during development (Roffwarg et al. 1966).

A system that does not adapt would not require such postnatal maintenance; the weights of each neuron could simply be fixed to their prenatally developed values. But if the adult system remains plastic and learns by Hebbian rules, it can eventually drift far from its starting state. Being fully adaptable like this may be desirable for some brain areas and tasks, particularly for those that change significantly over the course of a lifetime. For others, i.e. tasks that remain unchanged for millennia, adaptability may be irrelevant or even detrimental. Postnatal patterns may provide a simple way for specific types of processing to trade off between adaptability and explicit specification of function. Future models can study how this interleaving interacts with experience, perhaps using the Sengpiel et al. (1999) results as a starting point.

9.5 Embodied/situated perception

The HLISSOM simulations focused only on explaining face detection at a particular size scale, and did not address how full face processing and object processing abilities develop in children older than a few months. To fully model this development, simulations would require a training regimen that includes sequences of images of people from different angles, in different lighting, with hairstyles and clothing that change over time, against different background scenes, and over the course of months or years. Unless the visual experiences vary along such dimensions, we would not expect to see neurons develop that respond to the same individual regardless of hairstyle or clothing.

Einhauser et al. (2002) have recently taken first steps in collecting such realistic training data by attaching a camera to an adult housecat's head, and collecting video while it freely roams the environment. Einhauser et al. argue that this approach is a good approximation to the visual experience of the adult cat, because cats tend to move their heads (and not just their eyes) when scanning a visual scene. However, the visual experience of a kitten may be quite different from an adult's. For instance, very young kittens probably move around their environment much less than adult cats do, particularly during the earliest times after eye opening while their visual systems are first developing.

Moreover, it seems impractical to extend this approach to human infants, or even to closely related primates. It is true that video cameras have become small enough to be mounted unobtrusively on a headband. But apart from any ethical considerations, the video thus recorded would not necessarily be a good approximation to the infant's visual experience. In humans, eye movements can show significantly different trends from head movements. Humans have larger heads that often

remain still while we scan a scene or fixate on different parts of an object. Eye-tracking equipment exists, but not in a form that could be attached to an infant or animal unobtrusively for several months.

Finally, even if a realistic set of training data were obtained, several researchers have argued that video alone will not capture the crucial features of vision, particularly at higher, cross-modal areas of the nervous system. These authors argue that vision is an active process of interaction with the environment, driven by attention (as reviewed in Findlay 1998). The video data can record the pattern of light falling on the eye, but it cannot capture the causal link between visual system activity, visual attention, and the eye movements that determine the specific images seen, and thus future activity patterns. Given the difficulty in replicating this feedback loop, it will be impractical to approach human levels of visual system performance by simply making training patterns more realistic.

For these reasons, an "embodied" or "situated" approach may be a good next step towards understanding postnatal learning (as reviewed in Markman and Dietrich 2000; Pfeifer and Scheier 1998; Pylyshyn 2000; Thompson and Varela 2001). For instance, it may be useful to study a relatively simple robot as an example of an individual interacting with its environment. Such a system may provide insights that could not be obtained from a detailed computational study of human development, because it can preserve the causal link between visual system activity, eye and head movements, and the changes in activity patterns that result from such movements.

9.6   Engineering complex systems

This dissertation focused on modeling orientation and face processing, because those two biological capabilities show the clearest evidence of how genetic and environmental influences interact. Beyond these domains, the simulations can be seen as an example of a general-purpose problem-solving approach that could be applied to a variety of fields, perhaps making it practical to develop much larger computing systems than we use today. Such work can use either (a) specifically designed pattern generators, as in the simulations in this thesis, or (b) patterns that are generated by an evolved mechanism, via a genetic algorithm (GA; Holland 1975).

(a) Specifically designing a pattern generator allows an engineer to express a desired goal, within an architecture capable of learning. That is, the engineer can bias the patterns that will be learned by the system, without hard-coding a particular, inflexible architecture. Similarly, the engineer can also bootstrap a learning system with hand-coded patterns, allowing it to solve problems that are difficult for learning algorithms alone. For instance, the designer can feed in simpler patterns before real data, which could make subsequent learning easier by avoiding local minima in the search space of solutions (cf. Elman, Bates, Johnson, Karmiloff-Smith, Parisi, and Plunkett 1996; Gomez and Miikkulainen 1997; Nolfi and Parisi 1994). Such bootstrapping may also allow a designer to avoid expensive and laborious manual collection and/or tagging of training datasets, as

in tasks like handwriting recognition and face detection. For instance, a three-dot training pattern could be used to detect most faces, and only the patterns that were not detected would need to be tagged manually (cf. Viola and Jones 2001).

(b) Evolving a pattern generator with a genetic algorithm may also be an important technique. Evolving the generator eliminates the need for the domain-specific knowledge that is necessary to design one explicitly. For instance, studying real faces may lead one to suggest that a three-dot pattern would be a good training pattern to use to bootstrap a face detector, but it would be good to have an algorithm to choose such patterns automatically. Indeed, a learning system, pattern generator, and genetic algorithm together can be considered a single, general-purpose adaptive algorithm. What benefits would such a system have over other adaptive systems, such as genetic algorithms and learning networks alone? Essentially, the combination of learning and genetic algorithms represents a balance between adaptation at widely different time scales: in the short term, those particular circumstances that an individual network encounters, and in the long term, the cumulative fitness evaluations of all such networks evaluated by the genetic algorithm.

Adaptation at both time scales can be vitally important. Short-term adaptation allows an individual network to become particularly well-suited to the particular tasks it is tested on. Long-term adaptation (selection in the genetic algorithm) can ensure that short-term learning does not reduce generality. For instance, the GA can select training patterns that ensure that a system remains able to handle events that occur very rarely, yet are vitally important over the long term. As described for specifically designed generators above, the GA can also select pattern generators that get the system "in the ballpark", to increase the chance that learning will succeed. Thus by combining GAs and learning using pattern generators, it should be possible to evolve systems that perform better than using either approach alone.

9.7 Conclusion

This chapter outlined a variety of research areas opened up by the work in this thesis. Future experiments should measure more internally generated patterns to get clues to how brain systems are being constructed, and to inspire new approaches to building complex systems in general. The predictions of the HLISSOM model can also be tested in experiments with newborns and animals, to validate the model and to increase our understanding of the visual system.

Chapter 10

Conclusions

The results in this thesis showed that internally generated activity and postnatal learning can together explain much of how orientation maps and face preferences develop in animals and humans. Spontaneous activity ensures that the system develops robustly from the start, and learning from the environment ensures that the system is well-tuned to its actual surroundings. Together, a pattern generator, a learning algorithm, and the environment can construct a complex adaptive system. Below I summarize the contributions from each results chapter.

In chapter 4, I introduced HLISSOM, a new model of how the visual system develops. HLISSOM is the first model that shows in detail how genetic and environmental influences interact in multiple cortical areas. This level of detail is crucial for validating the model on experimental data, and for making specific predictions for future experiments in biology and psychology. As an example of how HLISSOM works, I used it to develop an orientation map simulation using abstract inputs. Results with this map showed that the ON and OFF channels introduced in HLISSOM are crucial for the model to be used with natural images in the later experiments. They also provide multi-lobed receptive fields that are a good match to those from biological experiments.

In chapter 5, I derived a set of size scaling equations and showed how they can be used to develop quantitatively equivalent orientation maps over a wide range of simulation sizes. I first showed that orientation maps depend on the stream of input patterns seen during development, not on the initial connection weight values. As a result, it is possible to develop nearly identical maps with different retina and cortex sizes. This capability is very useful for running simulations, and makes the results with large-scale orientation maps in later chapters possible.

In chapter 6, I showed how V1 neurons could develop biologically realistic, multi-lobed receptive fields and patterned intracortical connections, through unsupervised learning of spontaneous and visually evoked activity. I also showed how these neurons organize into biologically realistic topographic maps, matching those found at birth and in older animals. Postnatal experience gradually modifies the orientation map into a precise match to the distribution of orientations present in the environment. This smooth transition has been measured in animals, but not yet demonstrated

106

computationally.

In chapter 7, I showed how newborn human preferences for facelike patterns can result from learning of spontaneous activity, and that the self-organized face-selective map detects faces in real images. The hypothesis that newborn face preferences result from spontaneous activity follows naturally from experimental studies in early vision, but had not previously been proposed or tested.

In chapter 8, I showed that postnatal learning with real faces can explain how newborns learn to prefer their mothers, and why preferences for schematic faces disappear over time. Independently of the pattern generation hypothesis, the results show that researchers will need to reinterpret psychological studies that claim newborns learn face outlines, and that responses to faces in the periphery decrease over the first month of age. Instead, newborns may learn all parts of the face, and any decline might be specific to schematic stimuli alone.

Together, these results demonstrate a comprehensive approach to understanding the development and function of the visual system. They suggest that a simple but powerful set of self-organizing principles can account for a wide range of experimental results from animals and infants. These principles lead to concrete experimental predictions. The same principles can be applied to the development of future complex adaptive systems, allowing them to integrate prior information with environment-driven development and ongoing adaptation.

Appendix A

Parameter values

All of the simulations in this thesis used the same set of default parameters, with small modifications to these defaults as necessary to study different phenomena. The default parameters consist of values optimized in earlier work to produce realistic orientation maps (Bednar and Miikkulainen 2000b). The first section below will list and describe the defaults, and later sections will show how each simulation differed from the defaults.

A.1   Default simulation parameters

Because this thesis presents results from a large number of closely related simulations, the default values will be listed in a format that makes it simple to calculate new values when some of the defaults are changed. That is, instead of listing numeric values, most parameters are shown as equations derived from the scaling equations in chapter 5. The scaling equations require one particular network size to be used as a reference, from which parameters for other sizes can be calculated. Table A.1 lists these reference values, which are used as constants in later equations.

Using these reference values, the default parameters are calculated as shown in tables A.2 and table A.3. The parameters in table A.2 are constant for any particular simulation, but different simulations can override some of the defaults listed there. The parameters in table A.3 vary systematically over a single simulation, as shown in the table.

The sections below will explain how to use these three tables to compute the parameters for each of the different simulations. As an example, let us first consider how to use them to compute the parameters for the Bednar and Miikkulainen (2000b) orientation map simulation. To do this, one can go through table A.2 line by line, calculating the numerical value of each parameter by filling in the constants from table A.1 and the previous lines. For instance, parameter N_d can be calculated as 192, R_d as 24, and β_i as 0.65. The remaining values can be calculated similarly.

Note that most of the parameters in table A.2 are temporaries used only in later entries in the table and in table A.3. These are introduced here for notational convenience, and are not used by the HLISSOM model itself. For actual HLISSOM parameters, the tables list the equation or section where the parameter is used. Once these values are known, the temporary values can be discarded.

Constant	Value	Description
N_o	192	reference value of N, the length and width of the cortex, in number of units
R_o	36	reference value of R, the length and width of the retina, in number of units
r_{Ao}	6.5	reference value of r_A, the maximum radius of the afferent connections
r_{Eo}	19.5	reference value of r_E, the maximum radius of the lateral excitatory connections
r_{Io}	47.5	reference value of r_I, the maximum radius of the lateral inhibitory connections
a_o	7.5	reference value of a, the radius (σ) of the major axis of ellipsoidal Gaussian input patterns
b_o	1.5	reference value of b, the radius (σ) of the minor axis of ellipsoidal Gaussian input patterns
t_{f_o}	20,000	reference value of t_f, the number of training iterations
D_{Io}	0.0003	reference value of D_I, the lateral inhibitory connection death threshold

Table A.1: **Constants from original reference simulation.** These constants will be used in the equations in table A.2, and have the same value for every simulation. The subscript "o" in each name stands for original, as used in the scaling equations in chapter 5. The value of each constant is the corresponding parameter value from the simulation in Bednar and Miikkulainen (2000b). For example, constant r_{I_o} is the value of parameter r_I from that simulation.

Parameter	Value	Used in	Description
N_d	N_o	Table A.2	cortical density, i.e. width (and height) of a unit area of the cortex
R_d	$R_o - 2(r_{Ao} - 0.5)$	Table A.2	retinal density, i.e. width (and height) of a unit area of the retina (the area that projects to N_d)
n_A	1	Table A.2	number of regions from which the cortex receives afferent connections
s_a	1.0	Table A.2	area scaling factor
s_d	1.0	Table A.2	input density scale (ratio between average cortical activity from one oriented Gaussian to the average for the actual patterns used)
t_{si}	9	Table A.3	initial number of settling iterations
δ_i	0.1	Tables A.2, A.3	initial lower threshold of the sigmoid activation function
β_i	$\delta_i + 0.55$	Table A.3	initial upper threshold of the sigmoid activation function†
γ_A	1.0	Equation 4.5	scaling factor for the afferent weights
γ_E	0.9	Equation 4.7	scaling factor for the lateral excitatory weights
γ_I	0.9	Equation 4.7	scaling factor for the lateral inhibitory weights
γ_N	0.0	Equation 4.5	strength of divisive gain control

Table A.2: **Defaults for constant parameters.** *(Table continued on next page.)* These equations show how to construct the default values for parameters, using the reference values from table A.1. Parameter values that change during the simulation will be shown later in table A.3. In the equations marked with a dagger (†), the numerical constants were calculated from earlier simulations, primarily Bednar and Miikkulainen (2000b). Parameters that are listed as being used only in this and later tables are temporaries used for notational convenience, while those listed as being used in equations and sections of the text are parameters of the HLISSOM model itself.

Parameter	Value	Used in	Description
r_A	$\frac{R_d}{4} + 0.5$	Section 4.1.3	maximum radius of the afferent connections†
r_I	$\frac{N_d}{4} - 1$	Section 4.1.3	maximum radius of the lateral inhibitory connections†
$r_{\mathrm{E}i}$	$\frac{N_d}{10}$	Table A.3	initial maximum radius of the lateral excitatory connections, before shrinking†
$r_{\mathrm{E}f}$	$\max(1.5, \frac{N_d}{44})$	Table A.3	minimum final value of the r_E after shrinking†
N	$s_a N_d$	Section 4.2.2	length and width of the cortex, in number of units
R	$s_a R_d + 2(r_\mathrm{A} - 0.5)$	Section 4.2	length and width of the retina, in number of units
s_r	$\left(\frac{R}{R_d + 2(r_\mathrm{A} - 0.5)}\right)^2$	Table A.2	retinal area scale, i.e. area of the retina relative to the retinal area in the reference simulation†
s_t	$\frac{1}{s_d}$	Tables A.2, A.3	iteration scaling factor, which can be adjusted to use fewer iterations if input patterns are more dense at each iteration, or vice versa
σ_A	$\frac{r_\mathrm{A}}{1.3}$	Section 4.1.3	radius (σ) of the initial Gaussian-shaped afferent connections (when not using uniform random weights)†
σ_E	$0.78 r_{\mathrm{E}i}$	Section 4.1.3	radius (σ) of the initial Gaussian lateral excitatory connections†
σ_I	$2.08 r_\mathrm{I}$	Section 4.1.3	radius (σ) of the initial Gaussian lateral inhibitory connections†
i	$\max(1, s_d s_r)$	Section 4.2	number of discrete input patterns per iteration (e.g. Gaussians)
s_w	$\frac{r_{\mathrm{A}o} + 0.5}{r_\mathrm{A}}$	Table A.2	Scale of r_A relative to the default
a	$\frac{a_o}{s_w}$	Equation 4.2	radius (σ) of the major axis of ellipsoidal Gaussian input patterns
b	$\frac{b_o}{s_w}$	Equation 4.2	radius (σ) of the minor axis of ellipsoidal Gaussian input patterns
w	$\frac{50}{s_w}$	Section 6.2.1	width of the full-brightness portion of disc-shaped patterns†
σ_f	$\frac{b}{s_w}$	Section 6.2.1	radius (σ) of the Gaussian falloff in brightness at the edge of disc-shaped patterns†
t_f	$t_{f_o} s_t$	Section 4.2	number of training iterations
D_I	$D_{\mathrm{I}o} \frac{r_{\mathrm{I}o}^2}{r_\mathrm{I}^2}$	Section 4.3	lateral inhibitory connection death threshold
t_D	t_f	Section 4.3	iteration at which inhibitory connections are first pruned
d_r	$2.2 r_\mathrm{A}$	Section 4.2	minimum separation between the centers of multiple input patterns†
$\alpha_{\mathrm{A}i}$	$\frac{0.0070}{n_\mathrm{A} s_t s_d}$	Table A.3	initial afferent learning rate†
$\alpha_{\mathrm{E}i}$	$\frac{0.002 r_{\mathrm{E}o}^2}{s_t s_d r_\mathrm{E}^2}$	Table A.3	initial lateral excitatory learning rate†
α_I	$\frac{0.00025 r_{\mathrm{I}o}^2}{s_t s_d r_\mathrm{I}^2}$	Equation 4.8	lateral inhibitory learning rate†
σ_c	$\frac{0.5}{s_w}$	Equation 4.1	radius (σ) of center Gaussian of DoG, for simulations with the LGN†
σ_s	$4\sigma_c$	Equation 4.1	radius (σ) of surround Gaussian of DoG, for simulations with the LGN†
γ_L	2.33	Equation 4.3	γ_A for the LGN†
R_p	R	Section 4.2	for discrete patterns, width (and height) of the retinal area in which centers are randomly selected

(Table A.2 continued from previous page)

Parameter	Initial value	Used in	Description
r_E	r_{Ei}	Section 4.1.3	maximum radius of the lateral excitatory connections
δ	δ_i	Equation 4.4	lower threshold of the sigmoid activation function
β	β_i	Equation 4.4	upper threshold of the sigmoid activation function
t_s	t_{si}	Section 4.2.2	number of settling iterations
α_A	α_{Ai}	Equation 4.8	afferent learning rate
α_E	α_{Ei}	Equation 4.8	lateral excitatory learning rate

Iteration	Settings				
$0s_t$	$r_E = \max(r_{Ef}, r_{Ei})$	$\delta = \delta_i$	$\beta = \beta_i$	$t_s = t_{si}$	$\alpha_A = \alpha_{Ai}$
$200s_t$	$r_E = \max(r_{Ef}, 0.600 r_{Ei})$	$\delta = \delta_i + 0.01$	$\beta = \beta_i + 0.01$	$t_s = t_{si}$	$\alpha_A = \alpha_{Ai}$
$500s_t$	$r_E = \max(r_{Ef}, 0.420 r_{Ei})$	$\delta = \delta_i + 0.02$	$\beta = \beta_i + 0.02$	$t_s = t_{si}$	$\alpha_A = \frac{50}{70}\alpha_{Ai}$
$500s_t$	$\alpha_E = 0.5\alpha_{Ei}$				
$1000s_t$	$r_E = \max(r_{Ef}, 0.336 r_{Ei})$	$\delta = \delta_i + 0.05$	$\beta = \beta_i + 0.03$	$t_s = t_{si}$	$\alpha_A = \frac{50}{70}\alpha_{Ai}$
$2000s_t$	$r_E = \max(r_{Ef}, 0.269 r_{Ei})$	$\delta = \delta_i + 0.08$	$\beta = \beta_i + 0.05$	$t_s = t_{si} + 1$	$\alpha_A = \frac{40}{70}\alpha_{Ai}$
$3000s_t$	$r_E = \max(r_{Ef}, 0.215 r_{Ei})$	$\delta = \delta_i + 0.10$	$\beta = \beta_i + 0.08$	$t_s = t_{si} + 1$	$\alpha_A = \frac{40}{70}\alpha_{Ai}$
$4000s_t$	$r_E = \max(r_{Ef}, 0.129 r_{Ei})$	$\delta = \delta_i + 0.10$	$\beta = \beta_i + 0.11$	$t_s = t_{si} + 1$	$\alpha_A = \frac{30}{70}\alpha_{Ai}$
$5000s_t$	$r_E = \max(r_{Ef}, 0.077 r_{Ei})$	$\delta = \delta_i + 0.11$	$\beta = \beta_i + 0.14$	$t_s = t_{si} + 2$	$\alpha_A = \frac{30}{70}\alpha_{Ai}$
$6500s_t$	$r_E = \max(r_{Ef}, 0.046 r_{Ei})$	$\delta = \delta_i + 0.12$	$\beta = \beta_i + 0.17$	$t_s = t_{si} + 3$	$\alpha_A = \frac{30}{70}\alpha_{Ai}$
$8000s_t$	$r_E = \max(r_{Ef}, 0.028 r_{Ei})$	$\delta = \delta_i + 0.13$	$\beta = \beta_i + 0.20$	$t_s = t_{si} + 4$	$\alpha_A = \frac{30}{70}\alpha_{Ai}$
$20000s_t$	$r_E = \max(r_{Ef}, 0.017 r_{Ei})$	$\delta = \delta_i + 0.14$	$\beta = \beta_i + 0.23$	$t_s = t_{si} + 4$	$\alpha_A = \frac{15}{70}\alpha_{Ai}$

Table A.3: **Default parameter change schedule.** The first section above lists the initial values for the six parameters that varied over the course of each simulation. The second section lists how the values of these parameters at each subsequent iteration were computed from the initial values and the parameters in table A.2. At each iteration listed, the new values were calculated using the equations shown. This parameter change schedule was adapted from Bednar and Miikkulainen (2000b).

A.2 Choosing parameters for new simulations

Despite the seemingly large number of parameters, few of them need to be adjusted when running a new simulation. The most commonly changed parameters are the cortical density (N_d) and area scale (s_a), because those parameters directly determine the time and memory requirements of the simulations. The default N_d of 192 for a 5mm×5mm area is a good match to the density of columns in V1 orientation maps (Bednar, Kelkar, and Miikkulainen 2002), but in practice much smaller values often work well. The default s_a of 1.0 covers an area large enough to show several orientation columns in each direction, but more area is useful when processing larger images.

Apart from the simulation size parameters, most simulations differ primarily by the choice of input patterns. Starting from a working simulation, usually only one or a few parameters need to be changed to obtain a similar simulation based on a new set of patterns. If the new pattern is similar in overall shape, often all that is needed is to set the afferent input scale (γ_A) or sigmoid threshold (δ_i) to a value that, on average, produces similar cortical activity levels. The result will usually be a quantitatively similar map, as shown in the simulations with and without ON/OFF channels in section 4.5.

For a large change in pattern shape or size, such as using natural images instead of Gaussian spots, two parameters usually need to be adjusted. First, the modeler will need to adjust the input scale or threshold to get results as similar to the original working simulation as possible. Second, the modeler will adjust the input density scale (s_d) to compensate for the remaining differences in the amount of input per iteration. Of course, because the system is nonlinear, it is not always possible to compensate completely.

As an example, if Gaussian input patterns are replaced with large, sharp-edged squares, each input will produce multiple activity bubbles in V1 instead of one bubble. The researcher would then set the input scale γ_A to a value that results in similarly-sized bubbles, and set s_d to the average number of bubbles per iteration in the new simulation. For input types with large, spread-out areas of activity, the lateral interaction strength (γ_E and γ_I) can also be increased to ensure that distinct activity bubbles form. Finally, for simulations where it is crucial that all active areas of an image result in cortical response, the afferent divisive gain control (γ_N) can gradually be increased as neurons become more selective during training. This increase will reduce the sensitivity to contrast and thus increase the consistency of the responses across an image and to different images. Other parameters do not usually need to be changed when changing the input patterns.

A.3 V1 simulations

The sections below show the specific differences from the default parameters, for each of the simulations in this thesis. To determine the actual parameters used, one can (conceptually) begin with a copy of tables A.2 and A.3, make the changes mentioned below, then calculate each of the values

starting at the top of each table. For instance, for a simulation that changes δ_i to 0.05, parameter β_i would become 0.6 instead of 0.65 (table A.2). Other changes can be computed similarly.

A.3.1 Gaussian, no ON/OFF

In this thesis, the Gaussian-input orientation map simulations without ON and OFF channels (figures 4.8 and 4.9*e-f*) had the default parameter values listed in tables A.2 and A.3, except that they used a slightly smaller cortical density ($N_d = 142$) to reduce memory requirements, and used two input patterns per iteration (by setting $s_d = 2.0$) to reduce the total training time.

A.3.2 Gaussian, ON/OFF

As described in section 4.5, simulations that include the ON and OFF layers of the LGN are otherwise nearly identical to those that bypass the LGN. For the orientation map, the most important parameter that had to be adjusted after adding the LGN was to set the sigmoid thresholds so that the total cortical response would be the same (on average) as without ON and OFF channels.

Specifically, the Gaussian-input simulations with ON and OFF channels (figures 4.3, 4.4, 4.6, 4.7, 4.8, and 4.9*b-d*) were identical to the Gaussian orientation map simulations without the LGN, as listed in section A.3.1, except that they had two LGN input regions (ON and OFF; thus $n_A = 2$), and used a lower input threshold ($\delta_i = 0.083$) to compensate for the lower average value of LGN activity compared to photoreceptor activity.

A.3.3 Uniform random

The simulations with uniform random noise (figure 6.1 **Random Noise**) were identical to the Gaussian ON/OFF simulations in section A.3.2, except that they used only one input pattern per iteration (by setting $s_d = 1.0$), and had LGN parameters optimized to produce activity for such low-contrast inputs ($\sigma_c = 0.75$, $\sigma_s = 3.0\sigma_c$, and $\gamma_L = 2.5$).

A.3.4 Discs

The simulations with circular discs (figure 6.1 **Discs**) were identical to the Gaussian ON/OFF simulation in section A.3.2, except that they used a smaller cortical density ($N_d = 96$) to reduce computation time, had stronger lateral interactions to allow long contours to be separated into distinct activity bubbles ($\gamma_I = 2.0$, $\gamma_E = 1.2$), had a higher input density scale ($s_d = 2$) because each input resulted in more than one activity bubble, and used a slightly stronger LGN afferent scale ($\gamma_L = 3$) because each activity bubble was slightly weaker than in the Gaussian simulation from section A.3.2.

The disc input patterns were at full strength for a circular region (radius $w = \frac{50}{s_w}$) around their centers, and the strength then fell off according to a Gaussian ($\sigma_f = 3$). Each input center was

separated far enough so that input patterns would never overlap ($d_r = 0.75w$). Because the disc stimuli are large compared to R_d, the area in which disc centers are chosen was increased so that even the neurons at the borders will be equally likely to see all parts of the discs. Thus instead of the $R_p = R$ used in simulations with oriented Gaussians, $R_p = R + w$ for the disc simulations. The number of inputs per eye was then corrected to reflect this larger area ($i = \max(1, \left(\frac{w+R_d s_a}{w+R_d}\right)^2)$.

The simulations with noisy circular discs (figure 6.1 **Noisy Discs**) were identical to the simulations with noiseless circular discs, except that uniform random noise in the range $[-0.5, 0.5)$ was added after the discs were drawn. The simulations with discs followed by natural images (figures 6.3, 6.4, and 6.5) used the parameters from the noisy circular discs simulations until iteration 1000, then used the parameters from section A.3.5 (*Natural images*).

A.3.5 Natural images

The orientation map simulations trained on natural images (figure 6.2) were identical to the Gaussian ON/OFF simulation in section A.3.2, except that they used a smaller cortical density ($N_d = 96$), had a higher input density scale ($s_d = 8$) because each input resulted in about eight activity bubbles on average, used a fixed iteration scale ($s_t = 1.0$) instead of the calculated $\frac{1}{s_d}$ to save computation time, had LGN parameters optimized to produce activity for low-contrast inputs ($\gamma_L = 4.7$), had stronger lateral interactions to allow long contours to be separated into distinct activity bubbles ($\gamma_I = 2.0$, $\gamma_E = 1.2$), and used a smaller sigmoid threshold ($\delta_i = 0.076$) to allow responses to lower-contrast stimuli.

A.4 FSA simulations

Simulations with only an FSA (no V1) used parameters adapted from Bednar and Miikkulainen (2000a). Table A.4 lists defaults for these simulations, each of which overrides a default listed in table A.2 as the default for FSA simulations. Individual FSA simulations can then override these defaults or the other parameters listed in table A.2.

The prenatal portion of the FSA-only simulations in chapter 8 (i.e., figure 8.1) was identical to the default FSA-only parameters in table A.4, except that the simulations needed only a small cortical density ($N_d = 24$) because there was little variation between FSA units, initialized the afferent weights with a fixed-width Gaussian ($\sigma_A = \frac{9.5}{1.3}$) instead of random noise to allow the enclosing radius r_A to be varied without affecting the size of the initial activity bubbles, and used a larger afferent radius to allow face outlines to be learned ($r_A = 25.5$).

The simulations with different face training pattern types (figure 7.9) were identical to the other FSA-only simulations except that each used a different γ_A to ensure that the average FSA activity was the same for each pattern. The γ_A values for the pattern types shown in the subfigures of figure 7.9 are listed in table A.5. These values were determined by presenting a set of random

Parameter	Value
n_A	2
r_{Ei}	$\frac{N_d}{6}$
r_I	$\frac{N_d}{2.4}$
r_A	9.5
s_w	$\frac{r_{Ao}}{9.5}$
s_d	1.5
s_t	0.5
d_r	$20(\frac{6.5}{9.5s_w})$
i	$\max(1, 0.5s_d s_r)$
γ_A	$\frac{2.14}{n_A}$
γ_L	10.2
σ_c	$\frac{0.75}{s_w}$
σ_s	$1.6\sigma_c$

Table A.4: **Defaults for FSA simulations.** FSA simulations used the defaults from table A.2 modified as shown above. These values were chosen to match FSA parameters optimized in earlier work (Bednar and Miikkulainen 2000a). Some of the FSA simulations override some of these defaults, as described in the text.

inputs while adjusting γ_A until the sum of the cortical response was the same as for the three-blob case (7.9*a*).

The postnatal portions of the FSA-only simulations in chapter 8 (figure 8.4) continued with the same parameters as they had at the end of prenatal training, except that α_A continued to be reduced as shown in table A.6, the sigmoid range was reduced ($\beta = \delta + 0.48$), and the lower threshold on the sigmoid (δ) was set separately for each network as described in section 8.2.2 and shown in table A.6.

A.5 Combined V1 and FSA simulations

The simulations in chapter 7 combine both V1 and the FSA. The retina, LGN, and V1 parameters were identical to the Discs simulation above (section 6.2.1), except that the combined simulations used a very small cortical density ($N_d = 24$) to reduce memory and computational requirements, a very large area ($s_a = 8$) to allow large retinal stimuli to be tested, a slightly smaller LGN radius ($\sigma_c = 0.4$) to match earlier simulations, and continued past 20,000s_t (as described below). The FSA parameters were identical to the default FSA-only parameters above (section A.4), except that they used a small cortical density ($N_d = \frac{36}{0.94}$), a very large retinal density ($R_d = 170$) to match the

Figure	Pattern	γ_A
7.9a		1.070
7.9b		0.697
7.9c		0.697
7.9d		0.550
7.9e		0.577
7.9f		0.490
7.9g		1.983
7.9h		1.416
7.9i		0.948

Table A.5: **Parameters for different types of face training patterns.** Each of the simulations used in the subfigures of figure 7.9 used the same parameters, except that the γ_A was adjusted by hand until the average activity resulting from each pattern was similar. The resulting γ_A is shown for each pattern above.

full size of V1, an area scale corresponding to the full area of V1 ($s_a = 0.94$), and the parameter change schedule listed in table A.7. The FSA also had only one input (V1) instead of two (the ON and OFF cell layers), and so $n_A = 1$ for the FSA.

Because FSA training followed V1 training, the FSA parameters treated training iteration $20{,}000s_t$ as if it were iteration 0, and thus e.g. iteration $28000s_t$ used parameters from iteration $8000s_t$ of the default FSA simulation. After iteration $28000s_t$, the V1 and FSA parameters followed the schedule shown in table A.7. The V1 parameters were chosen to ensure that the network responds well to large natural images. First, the value of γ_N was gradually increased to make the responses less dependent on image contrast. This way, similar shapes that have different contrasts in different areas of an image would give comparable V1 responses. The other parameters were then adjusted to compensate for the effect of γ_N, ensuring that V1 responses to the highest-contrast patterns were near 1.0, and lower contrasts gave lower V1 responses. The parameters of the FSA were chosen similarly, except that the FSA sigmoid threshold δ was also gradually increased so that the response would be nearly binary. In this way the FSA response was used as an unambiguous criterion for the presence of a face-like pattern on the input, as described in chapter 7.

Iteration		Prenatally trained	Naïve
$20000s_t$	$\alpha_A = \frac{10}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.120$	$\delta = \delta_i + 0.070$
$20400s_t$	$\alpha_A = \frac{10}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.120$	$\delta = \delta_i + 0.070$
$22000s_t$	$\alpha_A = \frac{9}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.120$	$\delta = \delta_i + 0.070$
$24000s_t$	$\alpha_A = \frac{8}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.090$	$\delta = \delta_i + 0.070$
$28000s_t$	$\alpha_A = \frac{7}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.070$	$\delta = \delta_i + 0.070$
$32000s_t$	$\alpha_A = \frac{7}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.070$	$\delta = \delta_i + 0.070$
$36000s_t$	$\alpha_A = \frac{6}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.050$	$\delta = \delta_i + 0.080$
$40000s_t$	$\alpha_A = \frac{6}{70}\alpha_{Ai}$	$\delta = \delta_i + 0.045$	$\delta = \delta_i + 0.090$

Table A.6: **Postnatal FSA parameter change schedule.** The FSA simulations in chapter 8 continued past 20,000 iterations to model postnatal learning, and the equations above show how the parameters were computed at these additional iterations. Together, table A.3 and these values make up the parameter change schedule for the postnatal FSA simulations.

Iteration	V1 settings				
$28000s_t$	$\delta = \delta_i + 0.140$	$\gamma_N = 0$	$\gamma_A = 1.90$	$\gamma_E = 0.9$	$\gamma_I = 0.9$
$30000s_t$	$\delta = \delta_i + 0.267$	$\gamma_N = 1$	$\gamma_A = 2.70$	$\gamma_E = 1.0$	$\gamma_I = 1.1$
$35000s_t$	$\delta = \delta_i + 0.317$	$\gamma_N = 2$	$\gamma_A = 2.90$	$\gamma_E = 1.1$	$\gamma_I = 1.3$
$40000s_t$	$\delta = \delta_i + 0.417$	$\gamma_N = 4$	$\gamma_A = 3.25$	$\gamma_E = 1.2$	$\gamma_I = 1.4$

Iteration	FSA settings				
$28000s_t$	$\delta = \delta_i + 0.130$	$\gamma_N = 0$	$\gamma_A = 3.0$	$\gamma_E = 0.9$	$\gamma_I = 0.9$
$30000s_t$	$\delta = \delta_i + 0.250$	$\gamma_N = 2$	$\gamma_A = 5.0$	$\gamma_E = 0.5$	$\gamma_I = 0.7$
$35000s_t$	$\delta = \delta_i + 0.400$	$\gamma_N = 5$	$\gamma_A = 9.0$	$\gamma_E = 0.4$	$\gamma_I = 0.6$
$40000s_t$	$\delta = \delta_i + 0.710$	$\gamma_N = 9$	$\gamma_A = 10.6$	$\gamma_E = 0.4$	$\gamma_I = 0.6$

Table A.7: **Parameter change schedule for the combined V1 and FSA network.** The network with both V1 and the FSA from chapter 7 was also trained longer than is listed in table A.3. In iterations 20,000–28,000, the FSA was trained using the V1 schedule from 0–8,000 listed in table A.3. The remainder of the schedule, from 28,000–40,000, is shown above. Because each cortical area has an independent set of parameters, the values for V1 and the FSA are listed separately.

A.6 Conclusion

The parameters given above can be used to reproduce the simulation results in this thesis. Although all of the parameters were listed here for completeness, most can be left unchanged or calculated from known values. Most of the rest can be set systematically and without an extensive search for the correct values. Thus in practice, each simulation has relatively few free parameters, and it is usually straightforward to use the model to simulate new phenomena.

Bibliography

AAAI (2000). *Proceedings of the 17th National Conference on Artificial Intelligence*. Cambridge, MA: MIT Press.

Abramov, I., Gordon, J., Hendrickson, A., Hainline, L., Dobson, V., and LaBossiere, E. (1982). The retina of the newborn human infant. *Science*, 217(4556):265–267.

Acerra, F., Burnod, Y., and de Schonen, S. (2002). Modelling aspects of face processing in early infancy. *Developmental Science*, 5(1):98–117.

Achermann, B. (1995). Full-faces database. Copyright 1995, University of Bern, all rights reserved. http://iamwww.unibe.ch/ fkiwww/Personen/achermann.html.

Adorján, P., Levitt, J. B., Lund, J. S., and Obermayer, K. (1999). A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. *Visual Neuroscience*, 16:303–318.

Albrecht, D. G., Farrar, S. B., and Hamilton, D. B. (1984). Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *Journal of Physiology*, 347:713–739.

Albrecht, D. G., and Geisler, W. S. (1994). Visual cortex neurons in monkey and cat: Contrast response nonlinearities and stimulus selectivity. In Lawton, T., editor, *Computational Vision Based on Neurobiology*. Bellingham, Washington: SPIE.

Alexander, D. M., Bourke, P. D., Sheridan, P., Konstandatos, O., and Wright, J. J. (1999). Emergence under Hebbian learning of local maps in the primary visual cortex: Orientation preference in the tree shrew. Submitted.

Alpert, D., and Avnon, D. (1993). Architecture of the Pentium microprocessor. *IEEE Micro*, 13(3):11–21.

Amari, S.-I. (1980). Topographic organization of nerve fields. *Bulletin of Mathematical Biology*, 42:339–364.

Anderson, J. A., and Rosenfeld, E., editors (1988). *Neurocomputing: Foundations of Research*. Cambridge, MA: MIT Press.

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.

Barrow, H., and Bray, A. (1993). An adaptive neural model of early visual processing. In Eeckman, F., and Bower, J. M., editors, *The Neurobiology of Computation: Proceedings of the Annual Computation and Neural Systems Meeting*. Dordrecht; Boston: Kluwer.

Barrow, H. G. (1987). Learning receptive fields. In *Proceedings of the IEEE First International Conference on Neural Networks* (San Diego, CA), vol. IV, 115–121. Piscataway, NJ: IEEE.

Barrow, H. G., and Bray, A. (1992). An adaptive neural model of early visual processing. In Aleksander, I., and Taylor, J., editors, *Proceedings of the International Conference on Artificial Neural Networks 1992*. New York: Elsevier.

Bartlett, M. S., and Sejnowski, T. J. (1997). Viewpoint invariant face recognition using independent component analysis and attractor networks. In Mozer, M. C., Jordan, M. I., and Petsche, T., editors, *Advances in Neural Information Processing Systems 9*, 817. Cambridge, MA: MIT Press.

Bartlett, M. S., and Sejnowski, T. J. (1998). Learning viewpoint-invariant face representations from visual experience in an attractor network. *Network – Computation in Neural Systems*, 9(3):399–417.

Bartrip, J., Morton, J., and de Schonen, S. (2001). Responses to mother's face in 3-week to 5-month-old infants. *British Journal of Developmental Psychology*, 19:219–232.

Bartsch, A. P., and van Hemmen, J. L. (2001). Combined Hebbian development of Geniculocortical and lateral connectivity in a model of primary visual cortex. *Biological Cybernetics*, 84:41–55.

Bednar, J. A. (1997). *Tilt Aftereffects in a Self-Organizing Model of the Primary Visual Cortex*. Master's thesis, Department of Computer Sciences, The University of Texas at Austin. Technical Report AI-97-259.

Bednar, J. A., Kelkar, A., and Miikkulainen, R. (2002). Scaling self-organizing maps to model large cortical networks. Submitted. Available as Technical Report AI-00-285, Department of Computer Sciences, University of Texas at Austin.

Bednar, J. A., and Miikkulainen, R. (2000a). Self-organization of innate face preferences: Could genetics be expressed through learning?. In (AAAI 2000), 117–122.

Bednar, J. A., and Miikkulainen, R. (2000b). Tilt aftereffects in a self-organizing model of the primary visual cortex. *Neural Computation*, 12(7):1721–1740.

Bednar, J. A., and Miikkulainen, R. (2001). Learning innate face preferences. Technical report, Department of Computer Sciences, The University of Texas at Austin, Austin, TX. Technical Report AI-01-291.

Bednar, J. A., and Miikkulainen, R. (2003). Self-organization of spatiotemporal receptive fields and laterally connected direction and orientation maps. In *Computational Neuroscience: Trends in Research, 2003*. To appear.

Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, 2:32–48.

Blais, B. S., Shouval, H., and Cooper, L. N. (1999). The role of presynaptic activity in monocular deprivation: Comparison of homosynaptic and heterosynaptic mechanisms. *Proceedings of the National Academy of Sciences, USA*, 96.

Blakemore, C., and Cooper, G. F. (1970). Development of the brain depends on the visual environment. *Nature*, 228:477–478.

Blakemore, C., and van Sluyters, R. C. (1975). Innate and environmental factors in the development of the kitten's visual cortex. *Journal of Physiology (London)*, 248:663–716.

Blasdel, G. G. (1992a). Differential imaging of ocular dominance columns and orientation selectivity in monkey striate cortex. *Journal of Neuroscience*, 12:3115–3138.

Blasdel, G. G. (1992b). Orientation selectivity, preference, and continuity in monkey striate cortex. *Journal of Neuroscience*, 12:3139–3161.

Blasdel, G. G., and Salama, G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature*, 321:579–585.

Bolhuis, J. J. (1999). Early learning and the development of filial preferences in the chick. *Behavioural Brain Research*, 98(2):245–252.

Bosking, W. H., Zhang, Y., Schofield, B., and Fitzpatrick, D. (1997). Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *Journal of Neuroscience*, 17(6):2112–2127.

Bray, A. J., and Barrow, H. G. (1996). Simple cell adaptation in visual cortex: A computational model of processing in the early visual pathway. Technical Report CSRP 331, Sussex University, UK.

Bronson, G. W. (1974). The postnatal growth of visual capacity. *Child Development*, 45:873–890.

Burger, T., and Lang, E. W. (1999). An incremental Hebbian learning model of the primary visual cortex with lateral plasticity and real input patterns. *Zeitschrift für Naturforschung C—A Journal of Biosciences*, 54:128–140.

Bushnell, I. W. R. (1998). The origins of face perception. In (Simion and Butterworth 1998), 69–86.

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development*, 10(1/2):67–74.

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.

Cai, D., DeAngelis, G. C., and Freeman, R. D. (1997). Spatiotemporal receptive field organization in the lateral geniculate nucleus of cats and kittens. *Journal of Neurophysiology*, 78(2):1045–1061.

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.

Casagrande, V. A., and Norton, T. T. (1989). Lateral geniculate nucleus: A review of its physiology and function. In Leventhal, A. G., editor, *The Neural Basis of Visual Function*, vol. 4 of *Vision and Visual Dysfunction*, 41–84. Boca Raton, Florida: CRC Press.

Catania, K. C., Lyon, D. C., Mock, O. B., and Kaas, J. H. (1999). Cortical organization in shrews: Evidence from five species. *Journal of Comparative Neurology*, 410(1):55–72.

Chapman, B. (2000). Necessity for afferent activity to maintain eye-specific segregation in ferret lateral geniculate nucleus. *Science*, 287(5462):2479–2482.

Chapman, B., and Bonhoeffer, T. (1998). Overrepresentation of horizontal and vertical orientation preferences in developing ferret area 17. *Proceedings of the National Academy of Sciences, USA*, 95:2609–2614.

Chapman, B., Gödecke, I., and Bonhoeffer, T. (1999). Development of orientation preference in the mammalian visual cortex. *Journal of Neurobiology*, 41(1):18–24.

Chapman, B., and Stryker, M. P. (1993). Development of orientation selectivity in ferret primary visual cortex and effects of deprivation. *Journal of Neuroscience*, 13(12):5251–5262.

Chapman, B., Stryker, M. P., and Bonhoeffer, T. (1996). Development of orientation preference maps in ferret primary visual cortex. *Journal of Neuroscience*, 16(20):6443–6453.

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. (1996). Self-organization and segmentation with laterally connected spiking neurons. Technical Report AI96-251, Department of Computer Sciences, The University of Texas at Austin.

Choe, Y., and Miikkulainen, R. (1997). Self-organization and segmentation with laterally connected spiking neurons. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, 1120–1125. San Francisco, CA: Morgan Kaufmann.

Choe, Y., and Miikkulainen, R. (1998). Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing*, 21:139–157.

Choe, Y., and Miikkulainen, R. (2000a). Contour integration and segmentation with self-organized lateral connections. Technical Report AI2000-286, Department of Computer Sciences, The University of Texas at Austin.

Choe, Y., and Miikkulainen, R. (2000b). A self-organizing neural network for contour integration through synchronized firing. In (AAAI 2000), 123–128.

Choe, Y., and Miikkulainen, R. (2002). Contour integration and segmentation in a self-organizing map of spiking neurons. Technical Report 2002-1-2, Department of Computer Science, Texas A&M University. Submitted.

Chouvet, G., Blois, R., Debilly, G., and Jouvet, M. (1983). [The structure of the occurrence of rapid eye movements in paradoxical sleep is similar in homozygotic twins] (French). *Comptes Rendus des Seances de l'Academie des Sciences – Serie III, Sciences de la Vie*, 296(22):1063–1068.

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.

Coppola, D. M., White, L. E., Fitzpatrick, D., and Purves, D. (1998). Unequal representation of cardinal and oblique contours in ferret visual cortex. *Proceedings of the National Academy of Sciences, USA*, 95(5):2621–2623.

Cover, T. M., and Thomas, J. (1991). *Elements of Information Theory*. Wiley.

Crair, M. C. (1999). Neuronal activity during development: Permissive or instructive?. *Current Opinion in Neurobiology*, 9:88–93.

Crair, M. C., Gillespie, D. C., and Stryker, M. P. (1998). The role of visual experience in the development of columns in cat visual cortex. *Science*, 279:566–570.

Crowley, J. C., and Katz, L. C. (2000). Early development of ocular dominance columns. *Science*, 290:1321–1324.

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.

Daw, N. (1995). *Visual Development*. New York: Plenum Press.

de Boysson-Bardies, B., editor (1993). *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*. Dordrecht; Boston: Kluwer.

de Gelder, B., and Rouw, R. (2000). Configural face processes in acquired and developmental prosopagnosia: Evidence for two separate face systems. *Neuroreport*, 11(14):3145–50.

de Gelder, B., and Rouw, R. (2001). Beyond localisation: A dynamical dual route account of face recognition.. *Acta Psychologica (Amsterdam)*, 107(1-3):183–207.

de Haan, M. (2001). The neuropsychology of face processing during infancy and childhood. In Nelson, C. A., and Luciana, M., editors, *Handbook of Developmental Cognitive Neuroscience*. MIT Press.

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.

Diamond, S. (1974). Four hundred years of instinct controversy. *Behavior Genetics*, 4:237–252.

Durbin, R., and Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, 343:644–647.

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.

Egan, J. P. (1975). *Signal Detection Theory and ROC Analysis*. Series in Cognition and Perception. Academic Press.

Eglen, S. (1997). *Modeling the Development of the Retinogeniculate Pathway*. PhD thesis, University of Sussex at Brighton, Brighton, UK. Technical Report CSRP 467.

Einhauser, W., Kayser, C., Konig, P., and Kording, K. P. (2002). Learning the invariance properties of complex cells from their responses to natural stimuli. *European Journal of Neuroscience*, 15(3):475–486.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Ernst, U. A., Pawelzik, K. R., Sahar-Pikielny, C., and Tsodyks, M. V. (2001). Intracortical origin of visual maps. *Nature Neuroscience*, 4(4):431–436.

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.

Farah, M. J., Wilson, K. D., Drain, M., and Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review*, 105(3):482–498.

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.

Ferrari, F., Manzotti, R., Nalin, A., Benatti, A., Cavallo, R., Torricelli, A., and Cavazzutti, G. (1986). Visual orientation to the human face in the premature and fullterm newborn. *Italian Journal of Neurological Sciences*, 5(Suppl):53–60.

Ferster, D. (1994). Linearity of synaptic interactions in the assembly of receptive fields in cat visual cortex. *Current Opinion in Neurobiology*, 4(4):563–568.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, 6:559–601.

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.

Findlay, J. (1998). Active vision: Visual activity in everyday life. *Current Biology*, 8:R640–642.

Gauthier, I., and Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology*, 11(2):219–224.

Geisler, W. S., and Albrecht, D. G. (1997). Visual cortex neurons in monkeys and cats: Detection, discrimination, and identification. *Visual Neuroscience*, 14(5):897–919.

Ghose, G. M., and Ts'o, D. Y. (1997). Form processing modules in primate area V4. *Journal of Neurophysiology*, 77(4):2191–2196.

Gilbert, C. D. (1998). Adult cortical dynamics. *Physiological Reviews*, 78(2):467–485.

Gilbert, C. D., Hirsch, J. A., and Wiesel, T. N. (1990). Lateral interactions in visual cortex. In *Cold Spring Harbor Symposia on Quantitative Biology, Volume LV*, 663–677. Cold Spring Harbor Laboratory Press.

Gilbert, C. D., and Wiesel, T. N. (1983). Clustered intrinsic connections in cat visual cortex. *Journal of Neuroscience*, 3:1116–1133.

Gilbert, C. D., and Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and cortico-cortical connections in cat visual cortex. *Journal of Neuroscience*, 9:2432–2442.

Gödecke, I., Kim, D. S., Bonhoeffer, T., and Singer, W. (1997). Development of orientation preference maps in area 18 of kitten visual cortex. *European Journal of Neuroscience*, 9(8):1754–1762.

Gomez, F., and Miikkulainen, R. (1997). Incremental evolution of complex general behavior. *Adaptive Behavior*, 5:317–342.

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.

Gray, M. S., Lawrence, D. T., Golomb, B. A., and Sejnowski, T. J. (1995). A perceptron reveals the face of sex. *Neural Computation*, 7(6):1160–1164.

Grinvald, A., Lieke, E. E., Frostig, R. D., and Hildesheim, R. (1994). Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual cortex. *Journal of Neuroscience*, 14:2545–2568.

Gross, C. G., Rocha-Miranda, C. E., and Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, 35(1):96–111.

Haith, G. L. (1998). *Modeling Activity-Dependent Development in the Retinogeniculate Projection*. PhD thesis, Department of Psychology, Stanford University, Stanford, CA.

Halgren, E., Dale, A. M., Sereno, M. I., Tootell, R. B., Marinkovic, K., and Rosen, B. R. (1999). Location of human face-selective cortex with respect to retinotopic areas. *Human Brain Mapping*, 7(1):29–37.

Hata, Y., Tsumoto, T., Sato, H., Hagihara, K., and Tamura, H. (1993). Development of local horizontal interactions in cat visual cortex studied by cross-correlation analysis. *Journal of Neurophysiology*, 69:40–56.

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.

Hayles, T., and Bednar, J. A. (2000). Attribute-based system and method for configuring and controlling a data acquisition task. U.S. patent 6067584, National Instruments Corporation.

Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.

Henry, G. H. (1989). Afferent inputs, receptive field properties and morphological cell types in different laminae of the striate cortex. In Leventhal, A. G., editor, *The Neural Basis of Visual Function*, vol. 4 of *Vision and Visual Dysfunction*, 223–245. Boca Raton, Florida: CRC Press.

Hershenson, M., Kessen, W., and Munsinger, H. (1967). Pattern perception in the human newborn: A close look at some positive and negative results. In Wathen-Dunn, W., editor, *Models for the Perception of Speech and Visual Form*. Cambridge, MA: MIT Press.

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., Alonso, J. M., Reid, R. C., and Martinez, L. M. (1998a). Synaptic integration in striate cortical simple cells. *Journal of Neuroscience*, 18(22):9517–9528.

Hirsch, J. A., Gallagher, C. A., Alonso, J. M., and Martinez, L. M. (1998b). Ascending projections of simple and complex cells in layer 6 of the cat striate cortex. *Journal of Neuroscience*, 18(19):8086–8094.

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.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Ann Arbor, MI: University of Michigan Press.

Horn, G. (1985). *Memory, Imprinting, and the Brain: An Inquiry Into Mechanisms*. Oxford: Clarendon Press.

Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160:106–154.

Hubel, D. H., and Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 28:229–289.

Hubel, D. H., and Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology (London)*, 195:215–243.

Hyvärinen, A., and Hoyer, P. O. (2001). A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 41(18):2413–2423.

Issa, N. P., Trepel, C., and Stryker, M. P. (2001). Spatial frequency maps in cat visual cortex. *Journal of Neuroscience*, 20:8504–8514.

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.

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

Jouvet, M. (1980). Paradoxical sleep and the nature-nurture controversy. In McConnell, P. S., Boer, G. J., Romijn, H. J., van de Poll, N. E., and Corner, M. A., editors, *Adaptive Capabilities of the Nervous System*, vol. 53 of *Progress in Brain Research*, 331–346. New York: Elsevier.

Jouvet, M. (1998). Paradoxical sleep as a programming system. *Journal of Sleep Research*, 7(Suppl 1):1–5.

Kaas, J. H. (2000). Why is brain size so important: Design problems and solutions as neocortex gets bigger or smaller. *Brain and Mind*, 1:7–23.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (1991). *Principles of Neural Science*. New York: Elsevier. Third edition.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience*, 3(8):759–763.

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.

Kasamatsu, T., Kitano, M., Sutter, E. E., and Norcia, A. M. (1998). Lack of lateral inhibitory interactions in visual cortex of monocularly deprived cats. *Vision Research*, 38(1):1–12.

Katz, L. C., and Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science*, 274:1133–1138.

Keesing, R., Stork, D. G., and Shatz, C. J. (1992). Retinogeniculate development: The role of competition and correlated retinal activity. In Moody, J. E., Hanson, S. J., and Lippmann, R. P., editors, *Advances in Neural Information Processing Systems 4*, 91–97. San Francisco, CA: Morgan Kaufmann.

Kiorpes, L., and Kiper, D. C. (1996). Development of contrast sensitivity across the visual field in macaque monkeys (*Macaca nemestrina*). *Vision Research*, 36(2):239–247.

Kleiner, K. (1993). Specific vs. non-specific face-recognition device. In (de Boysson-Bardies 1993), 103–108.

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. *Infant Behavior and Development*, 10:49–59.

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

Kolen, J. F., and Pollack, J. B. (1990). Scenes from exclusive-OR: Back propagation is sensitive to initial conditions. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, 868–875. Hillsdale, NJ: Erlbaum.

Lander, E. S., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Lee, A. B., Blais, B. S., Shouval, H. Z., and Cooper, L. N. (2000). Statistics of lateral geniculate nucleus (LGN) activity determine the segregation of ON/OFF subfields for simple cells in visual cortex. *Proceedings of the National Academy of Sciences, USA*, 97(23):12875–12879.

Linsker, R. (1986a). From basic network principles to neural architecture: Emergence of orientation columns. *Proceedings of the National Academy of Sciences, USA*, 83:8779–8783.

Linsker, R. (1986b). From basic network principles to neural architecture: Emergence of orientation-selective cells. *Proceedings of the National Academy of Sciences, USA*, 83:8390–8394.

Linsker, R. (1986c). From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proceedings of the National Academy of Sciences, USA*, 83:7508–7512.

Lippe, W. R. (1994). Rhythmic spontaneous activity in the developing avian auditory system. *Journal of Neuroscience*, 14(3):1486–1495.

Löwel, S., and Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science*, 255:209–212.

Luhmann, H. J., Martínez Millán, L., and Singer, W. (1986). Development of horizontal intrinsic connections in cat striate cortex. *Experimental Brain Research*, 63:443–448.

Markman, A. B., and Dietrich, E. (2000). Extending the classical view of representation. *Trends in Cognitive Science*, 4(12):470–475.

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.

Maurer, D., and Barrera, M. (1981). Infants' perception of natural and distorted arrangements of a schematic face. *Child Development*, 52(1):196–202.

Mayer, N., Herrmann, J. M., and Geisel, T. (2001). Signatures of natural image statistics in cortical simple cell receptive fields. *Neurocomputing*, 38:279–284.

Meltzoff, A. N., and Moore, A. K. (1993). Why faces are special to infants — On connecting the attraction of faces and infants' ability for imitation and cross-modal processing. In (de Boysson-Bardies 1993), 211–226.

Merigan, W. H., and Maunsell, J. H. R. (1993). How parallel are the primate visual pathways? *Annual Review of Neuroscience*, 16:369–402.

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.

Miller, K. D., Erwin, E., and Kayser, A. (1999). Is the development of orientation selectivity instructed by activity? *Journal of Neurobiology*, 41:44–57.

Miyashita-Lin, E. M., Hevner, R., Wassarman, K. M., Martinez, S., and Rubenstein, J. L. (1999). Early neocortical regionalization in the absence of thalamic innervation. *Science*, 285(5429):906–909.

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.

Movshon, J. A., and van Sluyters, R. C. (1981). Visual neural development. *Annual Review of Psychology*, 32:477–522.

Müller, T., Stetter, M., Hubener, M., Sengpiel, F., Bonhoeffer, T., Gödecke, I., Chapman, B., Löwel, S., and Obermayer, K. (2000). An analysis of orientation and ocular dominance patterns in the visual cortex of cats and ferrets. *Neural Computation*, 12(11):2573–2595.

Nachson, I. (1995). On the modularity of face recognition: The riddle of domain specificity. *Journal of Clinical & Experimental Neuropsychology*, 17(2):256–275.

National Park Service (1995). Image database. http://www.freestockphotos.com/NPS.

Nolfi, S., and Parisi, D. (1994). Desired answers do not correspond to good teaching inputs in ecological neural networks. *Neural Processing Letters*, 1(2):1–4.

Obermayer, K., and Blasdel, G. G. (1993). Geometry of orientation and ocular dominance columns in the monkey striate cortex. *Journal of Neuroscience*, 13:4114–4129.

Obermayer, K., Ritter, H. J., and Schulten, K. J. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proceedings of the National Academy of Sciences, USA*, 87:8345–8349.

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.

Oksenberg, A., Shaffery, J. P., Marks, G. A., Speciale, S. G., Mihailoff, G., and Roffwarg, H. P. (1996). Rapid eye movement sleep deprivation in kittens amplifies LGN cell-size disparity induced by monocular deprivation. *Brain Research. Developmental Brain Research*, 97(1):51–61.

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., and Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development*, 18:79–85.

Penn, A. A., and Shatz, C. J. (1999). Brain waves and brain wiring: The role of endogenous and sensory-driven neural activity in development. *Pediatric Research*, 45(4):447–458.

Perrett, D. I. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London Series B*, 335(1273):23–30.

Pfeifer, R., and Scheier, C. (1998). Representation in natural and artificial agents: An embodied cognitive science perspective. *Z Naturforsch [C]*, 53(7-8):480–503.

Phillips, P. J., Wechsler, H., Huang, J., and Rauss, P. (1998). The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5):295–306.

131

Pompeiano, O., Pompeiano, M., and Corvaja, N. (1995). Effects of sleep deprivation on the postnatal development of visual-deprived cells in the cat's lateral geniculate nucleus. *Archives Italiennes de Biologie*, 134(1):121–140.

Puce, A., Allison, T., Gore, J. C., and McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology*, 74(3):1192–1199.

Purves, D. (1988). *Body and Brain: A Trophic Theory of Neural Connections*. Cambridge, MA: Harvard University Press.

Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends Cogn Sci*, 4(5):197–207.

Rakic, P. (1988). Specification of cerebral cortical areas. *Science*, 241:170–176.

Rao, R. P. N., and Ballard, D. H. (1995). Natural basis functions and topographic memory for face recognition. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 10–17. San Francisco, CA: Morgan Kaufmann.

Rao, S. C., Toth, L. J., and Sur, M. (1997). Optically imaged maps of orientation preference in primary visual cortex of cats and ferrets. *Journal of Comparative Neurology*, 387(3):358–370.

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.

Rochester, N., Holland, J. H., Haibt, L. H., and Duda, W. L. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Transactions on Information Theory*, 2:80–93. Reprinted in Anderson and Rosenfeld 1988.

Rodieck, R. W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(11):583–601.

Rodman, H. R. (1994). Development of inferior temporal cortex in the monkey. *Cerebral Cortex*, 4(5):484–98.

Rodman, H. R., Skelly, J. P., and Gross, C. G. (1991). Stimulus selectivity and state dependence of activity in inferior temporal cortex of infant monkeys. *Proceedings of the National Academy of Sciences, USA*, 88(17):7572–7575.

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.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London Series B*, 335(1273):11–21.

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

Roque Da Silva Filho, A. C. (1992). *Investigation of a Generalized Version of Amari's Continuous Model for Neural Networks*. PhD thesis, University of Sussex at Brighton, Brighton, UK.

Rowley, H. A., Baluja, S., and Kanade, T. (1998). Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38.

Ruthazer, E. S., and Stryker, M. P. (1996). The role of activity in the development of long-range horizontal connections in area 17 of the ferret. *Journal of Neuroscience*, 16:7253–7269.

Sackett, G. P. (1966). Monkeys reared in isolation with pictures as visual input: Evidence for an innate releasing mechanism. *Science*, 154(755):1468–1473.

Schmidt, K. E., Kim, D. S., Singer, W., Bonhoeffer, T., and Lowel, S. (1997). Functional specificity of long-range intrinsic and interhemispheric connections in the visual cortex of strabismic cats. *Journal of Neuroscience*, 17(14):5480–5492.

Sengpiel, F., Stawinski, P., and Bonhoeffer, T. (1999). Influence of experience on orientation maps in cat visual cortex. *Nature Neuroscience*, 2(8):727–732.

Sergent, J. (1989). Structural processing of faces. In Young, A. W., and Ellis, H. D., editors, *Handbook of Research on Face Processing*, 57–91. New York: Elsevier.

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.

Shatz, C. J., and Stryker, M. P. (1978). Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *Journal of Physiology (London)*, 281:267–283.

Shmuel, A., and Grinvald, A. (1996). Functional organization for direction of motion and its relationship to orientation maps in cat area 18. *The Journal of Neuroscience*, 16:6945–6964.

Shouval, H., Intrator, N., and Cooper, L. (1997). BCM network develops orientation selectivity and ocular dominance in natural scene environment. *Vision Research*, 37:3339–3342.

Shouval, H., Intrator, N., Law, C. C., and Cooper, L. N. (1996). Effect of binocular cortical misalignment on ocular dominance and orientation selectivity. *Neural Computation*, 8(5):1021–1040.

Shouval, H. Z., Goldberg, D. H., Jones, J. P., Beckerman, M., and Cooper, L. N. (2000). Structured long-range connections can provide a scaffold for orientation maps. *Journal of Neuroscience*, 20(3):1119–1128.

Siegel, J. M. (1999). The evolution of REM sleep. In Lydic, R., and Baghdoyan, H. A., editors, *Handbook of Behavioral State Control*, 87–100. Boca Raton: CRC Press.

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.

Sincich, L. C., and Blasdel, G. G. (2001). Oriented axon projections in primary visual cortex of the monkey. *Journal of Neuroscience*, 21:4416–4426.

Sirosh, J. (1995). *A Self-Organizing Neural Network Model of the Primary Visual Cortex*. PhD thesis, Department of Computer Sciences, The University of Texas at Austin, Austin, TX. Technical Report AI95-237.

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., and Miikkulainen, R. (1997). Topographic receptive fields and patterned lateral interaction in a self-organizing model of the primary visual cortex. *Neural Computation*, 9:577–594.

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 Johnson, S. P. (1998). Visual sensory and perceptual abilities of the newborn: Beyond the blooming, buzzing confusion. In (Simion and Butterworth 1998), 121–142.

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., Morison, V., and Somers, M. (1988). Orientation discrimination and cortical function in the human newborn. *Perception*, 17:597–602.

Stellwagen, D., and Shatz, C. J. (2002). An instructive role for retinal waves in the development of retinogeniculate connectivity. *Neuron*, 33(3):357–367.

Sur, M., Angelucci, A., and Sharma, J. (1999). Rewiring cortex: The role of patterned activity in development and plasticity of neocortical circuits. *Journal of Neurobiology*, 41:33–43.

Sur, M., and Leamey, C. A. (2001). Development and plasticity of cortical areas and networks. *Nature Reviews Neuroscience*, 2(4):251–262.

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

Switkes, E., Mayer, M. J., and Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, 18(10):1393–1399.

Tarr, M. J., and Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3(8):764–769.

Tavazoie, S. F., and Reid, R. C. (2000). Diverse receptive fields in the lateral geniculate nucleus during thalamocortical development. *Nature Neuroscience*, 3(6):608–616.

Thomas, H. (1965). Visual-fixation responses of infants to stimuli of varying complexity. *Child Development*, 36:629–638.

Thompson, E., and Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Science*, 5(10):418–425.

Thompson, I. (1997). Cortical development: A role for spontaneous activity?. *Current Biology*, 7:R324–R326.

Tovée, M. J. (1998). Face processing: Getting by with a little help from its friends. *Current Biology*, 8:R317–R320.

Troyer, T. W., Krukowski, A. E., Priebe, N. J., and Miller, K. D. (1998). Contrast-invariant orientation tuning in cat visual cortex: Thalamocortical input tuning and correlation-based intracortical connectivity. *Journal of Neuroscience*, 18(15):5908–5927.

Ts'o, D. Y., Frostig, R. D., Lieke, E. E., and Grinvald, A. (1990). Functional organization of primate visual cortex revealed by high resolution optical imaging. *Science*, 249:417–420.

Turrigiano, G. G. (1999). Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same. *Trends in Neurosciences*, 22(5):221–227.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., and Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:845–846.

Valentin, D., Abdi, H., O'Toole, A. J., and Cottrell, G. W. (1994). Connectionist models of face processing: A survey. *Pattern Recognition*, 27:1209–1230.

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.

Van Essen, D. C., Anderson, C. H., and Felleman, D. J. (1992). Information processing in the primate visual system: An integrated systems perspective. *Science*, 255:419–423.

Venter, J. C., et al. (2001). The sequence of the human genome. *Science*, 291(5507):1304–1351.

Viola, P., and Jones, M. (2001). Robust real-time object detection. In *Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing, and Sampling*.

von der Malsburg, C. (1973). Self-organization of orientation-sensitive cells in the striate cortex. *Kybernetik*, 15:85–100. Reprinted in Anderson and Rosenfeld 1988.

von Melchner, L., Pallas, S. L., and Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, 404(6780):871–876.

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.

Wallis, G. M. (1994). *Neural Mechanisms Underlying Processing in the Visual Areas of the Occipital and Temporal Lobes*. PhD thesis, Corpus Christi College, Oxford University, Oxford, UK.

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

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.

Wandell, B. A. (1995). *Foundations of Vision*. Sunderland, Massachusetts: Sinauer Associates, Inc.

Weber, C. (2001). Self-organization of orientation maps, lateral connections, and dynamic receptive fields in the primary visual cortex. In *Proc. Intl. Conf. on Artificial Neural Networks (ICANN-2001)*, Lecture Notes In Computer Science 2130, 1147–1152. Springer-Verlag.

Weliky, M., Bosking, W. H., and Fitzpatrick, D. (1996). A systematic map of direction preference in primary visual cortex. *Nature*, 379:725–728.

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.

Weliky, M., and Katz, L. C. (1997). Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature*, 386(6626):680–685.

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.

Yang, M.-H., Kriegman, D., and Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(1):34–58. In press.

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.

Vita

Jim Bednar was born in Houston, Texas on February 28, 1971, where he resided until 1989. He enrolled at the University of Texas at Austin in fall 1989, where he received a B.S. in Electrical Engineering (December 1993), a B.A. in Philosophy (May 1994), and an M.A. in Computer Science (May 1997). During several internships at National Instruments Corporation he developed the DAQCard-700 PCMCIA card, twice voted one of the top 100 research and development products of the year, and developed the patented NLI-API data acquisition software (Hayles and Bednar 2000). He is currently the lead author of the Topographica brain modeling software package, under development through a Human Brain Project grant from the National Institutes of Health.

Permanent Address: 3806 Laurel Ledge Ln.
Austin, Texas 78731 USA
`jbednar@alumni.utexas.net`
`http://www.cs.utexas.edu/users/jbednar/`

This dissertation was typeset with LaTeX 2_ε[1] by the author.

---

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin, and extended by Bert Kay and James A. Bednar.