Modelling the Development of the Cat Lateral Geniculate Nucleus with Hebbian Learning

Stephen Eglen

School of Cognitive and Computing Sciences
University of Sussex, Brighton BN1 9QH, England

E-mail: stephene@cogs.susx.ac.uk

July 1995

CSRP 383

Abstract

The development of the cat Lateral Geniculate Nucleus depends on a combination of activity dependent and activity independent processes. A Hebbian model of the development of topography and ocular segregation in the LGN was first presented by Keesing, Stork, and Shatz (1992). This paper investigates two key features of the model: the normalisation schemes used and the nature of the retinal inputs. It is shown that the normalisation of weights into an LGN cell is not necessary, and that small positive between eye correlations do not prevent normal development.

1 Introduction

Mammalian visual systems are highly complex, occupying a large portion of the cortex (for example, Felleman and Van Essen (1991) estimate that around 55 % of the macaque cortex is devoted to some form of visual processing.) An important question to ask is how these visual systems develop both prenatally and postnatally. Given the complexity of visual systems, it is tempting to think that the visual system develops according to some predefined genetic program. However, it is thought that normal development requires a combination of both activity independent (genetic) and activity dependent processes. (See (Goodman & Shatz, 1993) for a review of these processes.)

For example, experiments by (Wiesel & Hubel, 1963) showed that cell responses in the kitten visual cortex are affected by early postnatal experience. If one of the kitten's eyes was closed shut for the first few weeks of life and then later reopened, it was found that cortical cells could no longer be driven by the eye that was deprived of vision. This is an example of an activity dependent processes, since the eye must be sending signals to the cortex in order for the cortical cells to develop responses to that eye.

Further evidence for activity dependent processing comes from (Sur, Garraghty, & Roe, 1988). They showed that when ferret retinal inputs are re-routed into parts of the thalamus and cortex normally devoted to auditory processing, these cells can then be visually driven (See Figure 1). Furthermore, some cells in the auditory cortex were directionally selective or orientation selective. Sur et al. conclude by suggesting that "the modality of a sensory thalamic nucleus or cortical area may be specified by its inputs during development." In other words, it could be that there is nothing intrinsically special about the visual cortex: it becomes the main area for visual processing just because it receives visual input.

In order to look at the role of activity dependent processes in more detail, one particular structure in the visual pathway will be considered, namely the Lateral Geniculate Nucleus (LGN). In order to avoid cross species comparisons, only the cat LGN will be considered here.

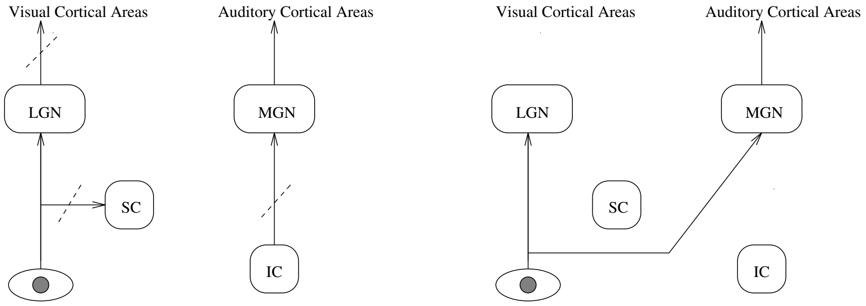


Figure 1: Experimental Procedure used to re-route retinal inputs in the ferret (Sur et al., 1988). The left hand figure shows the surgery performed: dotted lines indicates a pathway that was cut. Retinal inputs were re-routed from the Superior Colliculus (SC) to the Medial Geniculate Nucleus (MGN). Cells in the MGN and auditory cortex were found to be visually driven (IC = Inferior Colliculus).

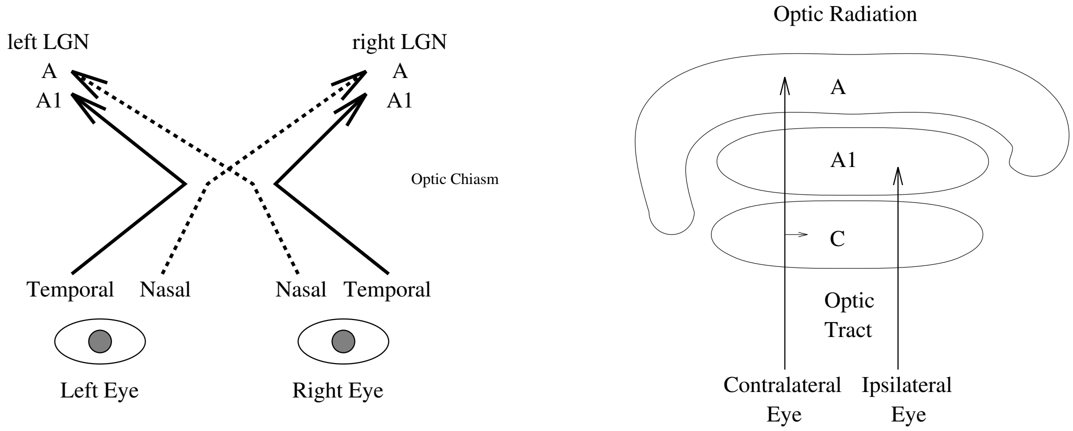


Figure 2: Organization of the cat retinogeniculate pathway. Each LGN receives input from only half of the visual field. The figure on the right shows a coronal section of the LGN, revealing its laminar structure.

1.1 The Lateral Geniculate Nucleus (LGN)

The LGN is the primary destination for retinal axons in the cat. There are two LGNs in each cat brain, as shown in Figure 2. Each LGN receives input from both eyes in such a way that visual information from the left half of the visual field goes to one LGN, and information from the right half goes to the other LGN. This partial crossover of retinal axons is performed in the optic chiasm.

The LGN has a layered structure, as shown in Figure 2. Layers A and A1 receive X and Y retinal ganglion cell inputs, whereas layer C receives mainly W retinal ganglion cell inputs (Sherman, 1985). Much more is known about X and Y ganglion cells than the W cells, and so the C layer will not be further discussed here. The LGN can be regarded as a relay station, with retinal inputs arriving via the optic tract and then passing onto the cortex via the optic radiation. However, this view is too simplistic, as only about 10 to 20% of the synapses onto geniculate relay neurons come from the retina, whereas about 50% come from the visual cortex, indicating that the LGN is doing more than just passing the signals from the retina to the cortex. (Sherman & Koch, 1986).

At birth, the cat LGN has two structural features of interest. Firstly, there is ocular segregation: contralateral eye inputs (retinal inputs from the eye on the opposite side of the body to the LGN) project to layer A, whereas ipsilateral inputs (retinal inputs from the eye on the same side of the body as the LGN) project to layer A1. Secondly, there is a topographic organization of the visual field onto layers A and A1. Neighbouring regions of the retina are mapped onto neighbouring regions in the LGN. Sanderson (1971) gives a good review of the precise topographic mapping for those interested.

Day	Event
E23	Retinal axons are yet to arrive at the optic chiasm. (Sherman, 1985, p352)
E30–E40	Correlated bursting activity begins (Wong 1994, personal communication.)
E32	First retinal axons reach the LGN. During next two weeks, there is considerable overlap: most cells in A and A1 are binocular (Sherman, 1985, p351)
E38	Functional synaptic transmission between ganglion cell axon and LGN neuron (Shatz, 1994, p535)
E47	Segregation of retinal inputs into laminae begins.
E48	Loss of ganglion cells begins (Sherman, 1985, p353)
E63	Birth.
P10	Eye opening (Wong et al., 1993, p935)

Table 1: Timescale of cat retinogeniculate development. E = embryonic, P = postnatal.

1.2 Waves of Activity in the Prenatal Retina.

Since the topography and ocular segregation are present at birth, postnatal experience is not necessary to develop these features. Prenatal experience, however, is required. Even before the photoreceptors in the retina have developed, the retinal ganglion cells spontaneously fire. This firing is spatially and temporally correlated, producing 'waves of activity' in the ganglion cell layer of the retina (Wong, Meister, & Shatz, 1993; Meister, Wong, Baylor, & Shatz, 1991). Typically, during early development, a patch of retina is active for just a few seconds and then is quiet for about a minute.

This activity is thought to be necessary for the ocular segregation to develop, because blocking this spontaneous activity with Tetrodotoxin (a drug that acts to block the Na^+ channel in the ganglion cell axons) prevents the segregation of axons into layer A and A1 (Shatz & Stryker, 1988).

Further evidence linking these waves of activity to the development of ocular segregation is given by examining the timescale of events during development. Table 1 shows the main events in the development of the retinogeniculate pathway. At around E32 (Embryonic day 32), the first retinal axons reach the LGN. Sometime between E30 and E40, it is estimated that the correlated bursting activity begins in the cat. (R. Wong, personal communication) By E47, the segregation of inputs into layers A and A1 begins. This in itself does not prove that that the waves cause the segregation, but they do occur at the right time during development for them to be a plausible mechanism for the segregation.

So far, no mention of the development of topography has been made. Since the waves of activity produce correlated activity in neighbouring ganglion cells, these waves can also provide information to refine the topographic mapping of the retinogeniculate pathway. (Early simulations by Willshaw and von der Malsburg (1976) showed that neighbouring presynaptic cell activity is sufficient to provide local topographic order.)

Clearly the development of the LGN requires activity driven processes, but are they sufficient on their own to produce the features of the LGN that are present at birth? One way to test this hypothesis is to build a computer model in which we can vary the influence of both activity dependent and independent processes.

2 The Computer Model

An initial model of the retinogeniculate pathway to investigate these questions was developed by (Keesing et al., 1992). The architecture of the neural network model is displayed in Figure 3. For simplicity, the two hemi retinae are one dimensional with wrap around. The two retinae are fully connected to a two dimensional LGN, arranged in a grid 10 cells wide by 8 cells "deep". Hence, the model is of a coronal slice through the LGN. The equations describing the activity of LGN cells are given in Table 2.

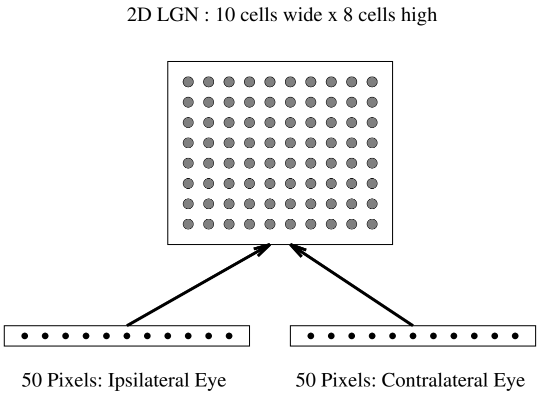


Figure 3: Architecture of the Network. Each retina is one dimensional and fully connects to the two dimensional LGN. This represents a coronal slice of the LGN.

r_i	Activity of the ith retinal cell.
nR	Number of retinal cells (100)
nL	Number of LGN cells (80)
w_{ij}	Weight from the ith retinal cell to the jth LGN cell.
o_j	Activity of the jth LGN cell.
	$o_j = \sum_{k=1}^{nr} w_{kj} r_k$

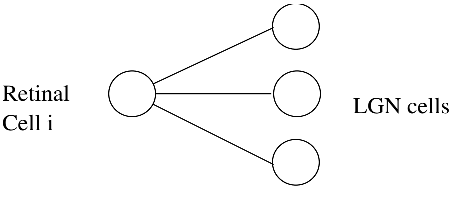
Table 2: Equations describing the activity of the network.

Learning

The weights are modified after every iteration using a correlational Hebbian rule, subject to the constraint that all weights should be non negative:

$$\Delta w_{ij} = \varepsilon(r_i - \alpha)(o_j - \beta)$$

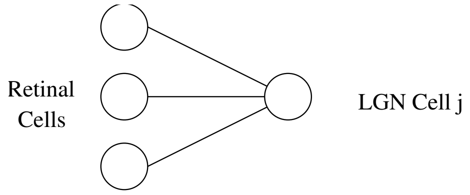
Here, α, β and ε are all parameters (see Appendix A). In the original model, two forms of normalisation were included to prevent unbounded weight growth and to introduce competition. Firstly, the weights from each retinal cell were constrained using divisive normalisation, such that $\sum_{j=1}^{nL} w_{ij} = tR$. Here, tR is the target sum of the retinal weights. Since this form of normalisation constrains the weights from a retinal cell, it shall be referred to as retinal normalisation. The new weight values w'_{ij} are given by:



Retinal Normalisation

$$w'_{ij} = \frac{w_{ij} \cdot tR}{\sum_{k=1}^{nL} w_{ik}}$$

Secondly, the weights to each LGN cell are constrained using subtractive normalisation such that $\sum_{i=1}^{nR} w_{ij} = tL$. Here, tL is the target sum of the LGN weights. This scheme will be called geniculate normalisation. The new weight values w'_{ij} are given by:



Geniculate Normalisation

$$w'_{ij} = w_{ij} - \frac{1}{nR}\left(tL - \sum_{k=1}^{nL} w_{kj}\right)$$

Subtractive normalisation is actually an iterative procedure, since there is an additional constraint that each weight must not be negative. If any weights to go below zero, then they are set to zero and the normalisation is repeated on the remaining non zero weights. In this model, if weights do go to zero, they are allowed to increase again. (Some models, for example Miller, Keller, and Stryker (1989) freeze weights when they go to zero.)

The final aspect of the original model by (Keesing et al., 1992) was to use a growth rule to encourage a retinal cell to increase the synaptic strength to neighbouring LGN cells:

$$\Delta w_{ij} = \gamma \sum_{k=\text{neigh}(j)} w_{ik}$$

So, if a retinal cell i synapses onto several neighbouring cells of a LGN cell j, the weight w_{ij} should be increased by some fraction. In this model, a LGN cell at location (x_0, y_0) in the grid is considered as a neighbour of a cell (x, y) if $|x_0 - x| \leq radius$ and $|y_0 - y| \leq radius$.

The value of $radius$ determines how large the neighbourhood is. The value of $radius$ can decrease over time to produce smaller neighbourhoods, in a similar fashion to the Kohonen network. Additionally, there is a toroidal wrap around in the X dimension of the LGN grid, but not in the Y dimension, so that cells at the end of each row are considered neighbours. This is demonstrated in Figure 4.

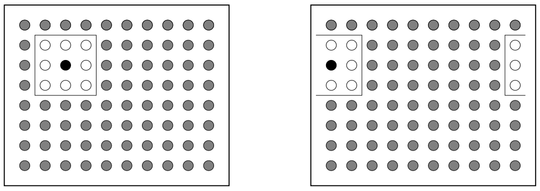


Figure 4: Neighbourhood function. In both pictures, the central node is coloured black, and its neighbours (using $radius$ = 1) are coloured white. The picture on the right shows the wrap around present in the X dimension of the network.

Waves of Retinal Activity

The waves of activity are modelled here as gaussian distributions around the centre of the wave. A wave moves across a retina at a rate of one pixel per time step. For a wavefront centred at element d in the input vector, the activity of retinal cell r_i is given by:

$$r_i = exp^{\frac{i-d}{2\sigma^2}}$$

The parameter σ effectively describes the width of the retinal wave. In the simplest experiments, a wave is present in one eye whilst the other eye is silent. An example set of input vectors are shown in Figure 5.

Training Paradigm

The weights w_{ij} are initially assigned random values in the range [0,1]. The network is trained by presenting each input vector from the set of inputs and updating the weights with the Hebbian rule. The weights are

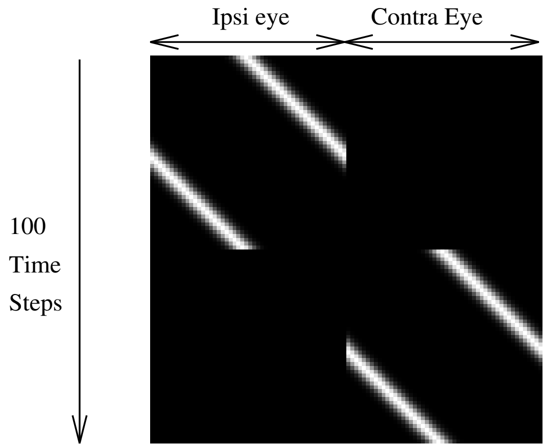


Figure 5: Retinal inputs shown over one epoch (100 iterations). In this example, only one eye is active at a time.

normalised either after every iteration (presentation of one input vector) or after every epoch (presentation of all input vectors). Growth learning takes place at random, typically so that it happens once or twice an epoch.

3 Results

This section describes the experiments performed with the network. Figure 6 explains how to interpret the weight diagrams shown in this paper. Hinton diagrams are also shown, but these have only been used on a subset of the weight matrix for reasons of space. Ocular dominance plots are also used – these plots colour each LGN cell according to the nature of its inputs. LGN cells receiving at least 80% of its weights from the ipsilateral (contralateral) eye are coloured black (white). All other cells are coloured grey to indicate they receive both ipsilateral and contralateral input.

Five sets of experiments are described here, listed below. Tests 1 to 3 replicate the results described in (Keesing et al., 1992). The remaining tests examine certain aspects of this model.

1. Simultaneous arrival of all retinal afferents.
2. Biased arrival times of retinal afferents.
3. Chemical Gradients used to Bias Topology.
4. Importance of growth term.
5. Importance of normalisation.

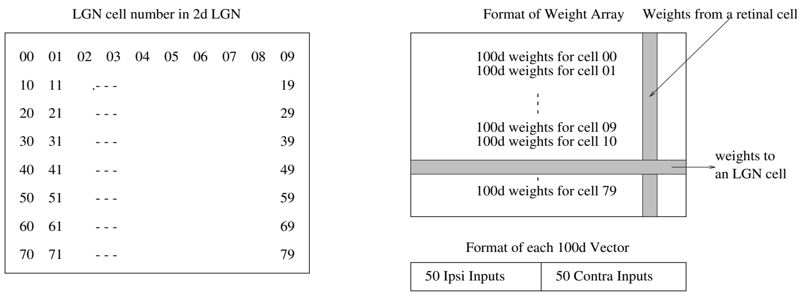


Figure 6: Interpreting the weight diagrams. Each LGN cell is numbered as shown, from 0 to 79. The right hand figure shows how the weights are displayed: row n of the weights array corresponds to the 100 dimensional vector for LGN cell n. Each weight vector can be subdivided into the 50 weights for the ipsilateral eye and 50 weights for the contralateral eye. Each element of the weight array is represented by a pixel: the larger the weight, the brighter the pixel.

3.1 Test 1: Simultaneous Arrival of All Retinal Afferents

Figure 7 shows the results of training the model with completely random initial weights. As can be seen, the weights gradually refine to produce a local topographic ordering. This can be seen more clearly in Figure 8 which shows a Hinton plot of weights for LGN cells 60-69 (the 7th row of weights in the LGN). Clearly, a global topographic mapping has not been achieved. (A global topographic mapping is achieved when the weights are concentrated along a diagonal in the Hinton Diagram, or appear as repeated diagonal lines in the weight diagram.)

By the end of training, all but 7 of the LGN cells became monocular (38 ipsilateral, 35 contralateral), as shown in picture 3 of Figure 7. This shows that most of the cells have become monocular, but they have organized into patches, rather than just the 2 layers as seen in the cat LGN.

It is interesting to compare these results with the innervation pattern seen when ferret retinal inputs are rerouted into the MGN (see Figure 10b of Roe, Garraghty, Esguerra, & Sur, 1993). The rerouted ferret retinal inputs cluster into patches, rather into the normal layers. This evidence can be interpreted as suggesting that for the formation of layers, extra information must be provided, and that it is likely that this information is specified by activity independent mechanisms (Sur, 1995).

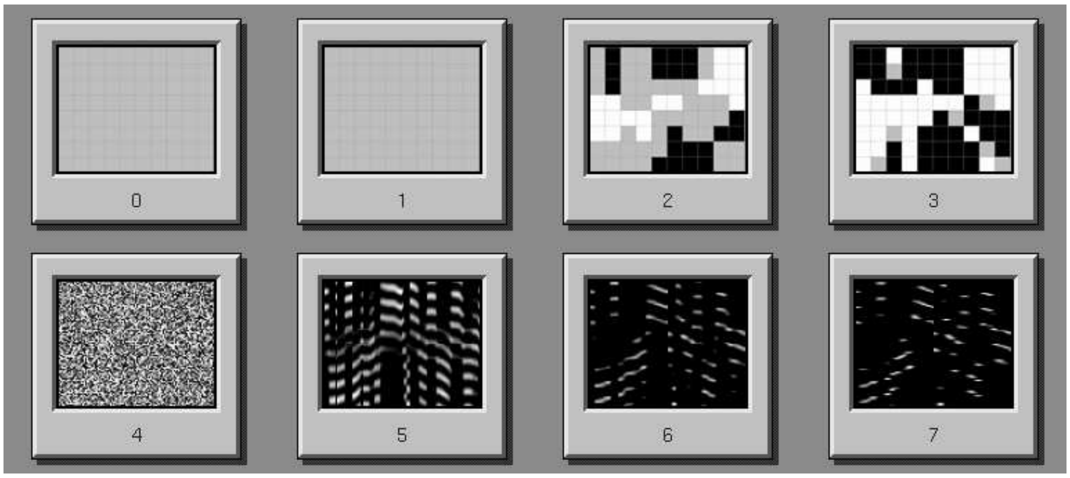


Figure 7: Test 1: Development of the weights when the initial weights are completely random. The upper figure shows the ocular dominance plots with the corresponding weight diagrams underneath. (0) Initial Weights, (1) 200 epochs, (2) 500 epochs, (3) 800 epochs.

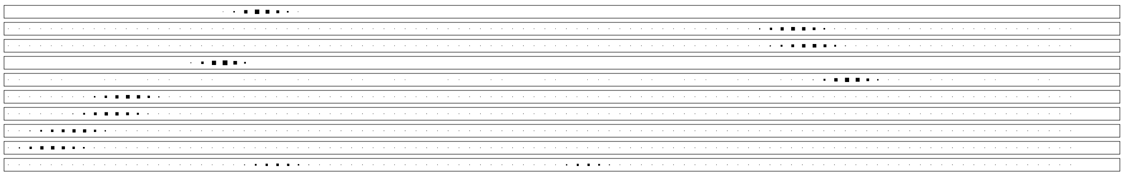


Figure 8: Hinton diagram for one row of LGN cells (row 7: cells 60-69), from picture 7 in Figure 7. The larger the square, the larger the corresponding weight.

3.2 Test 2: Biased Arrival Times of Retinal Afferents

The previous test showed that given completely random initial weights, global topography and layered ocular segregation do not develop. It is known however that contralateral eye inputs innervate the LGN several days earlier than ipsilateral eye inputs (Shatz, 1994, p535). This means that the contralateral inputs have an advantage in innervating the upper half of the LGN, namely layer A. (This earlier arrival time for contralateral inputs explains why the same pattern of ocular segregation always arises in the cat LGN.)

To model this difference in arrival times, we can bias the initial weights so that only the contralateral inputs can initially innervate the upper half of the LGN model. To do this, we initialise the weights to random values as before, and then set the weights from the ipsilateral eye to the upper half of the LGN to zero. This is demonstrated in Figure 9. (Note that just because some of the initial weights are set to zero, they can be modified during training, just like all other weights.)

As can be seen from Figure 9, the LGN starts out with the upper half being dominated by contralateral inputs, but the bottom half of the LGN receives both contralateral and ipsilateral input. Very quickly, the upper half of the LGN is innervated by ipsilateral inputs as well (picture 1), but then gradually produces a contralateral and ipsilateral layer with a binocular border in the middle (picture 2). Finally, the binocular region diminishes, leaving just 5 binocular cells (picture 3).

If we examine the topography produced by this experiment, we do get a better topography, but it is still not a continuous mapping: clear breaks can be seen. These mappings are very similar to those produced by (Willshaw & Von der Malsburg, 1979, p213), which are called piece-wise continuous projections.

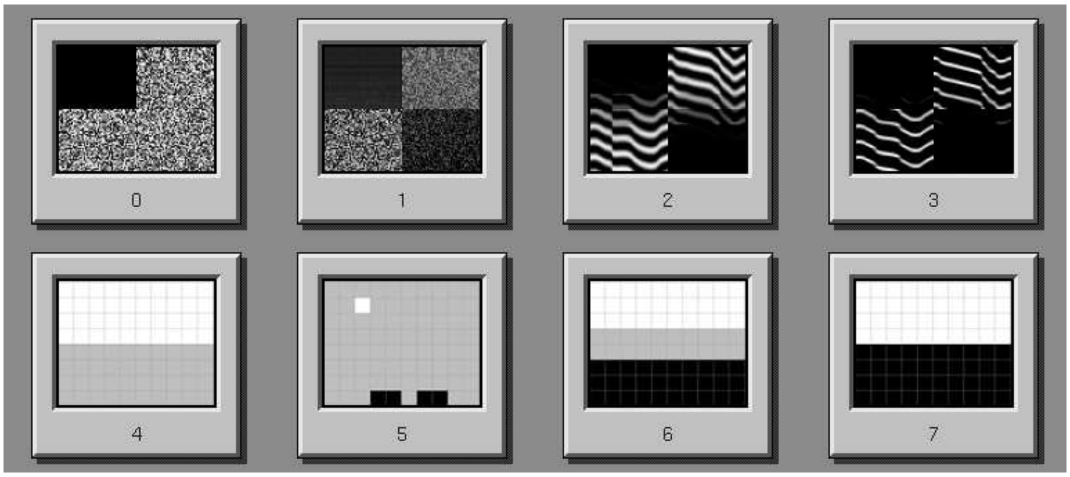


Figure 9: Test 2: Biased arrival times of retinal axons. Picture 0 shows the initial weights: contralateral axons innervate all 8 layers of the LGN, whereas the ipsilateral axons only innervate the bottom 4 layers. The top row of pictures show the weight diagrams, with corresponding ocular dominance plots underneath. The final weights reveal clear ocular segregation, but only a piecewise continuous topographic projection. (0) Initial weights, (1) 1 Epoch, (2) 400 Epochs , (3) 750 epochs

3.3 Test 3: Chemical Gradients used to Bias Topology

So far, the model can generate local topographic mappings, but there is no global mapping from the retina to each row of the LGN. As argued by Willshaw and Von der Malsburg (1979), some initial specification of the orientation information must be required (for example, which end of the retina connects to which end of the LGN), as there is nothing in the model to predispose it to one particular orientation.

An extra mechanism to provide this global mapping is likely to exist in the retinogeniculate pathway: it is widely thought that chemical gradients are responsible for guiding the retinal ganglion cell axons to the LGN and then projecting in a coarse grained topographic manner (Goodman & Shatz, 1993, p92).

To incorporate this mechanism into the model, the initial weights of the network are biased so that the weights at an intermediate distance away from the main diagonal of a row of weights are set to zero, or to small random values, as shown in Figure 10.

It is important to note when biasing the weights for topography, there is no need to bias all of the weights: it is sufficient to bias just one row in each of the contralateral / ipsilateral parts of the weight array. Once there is a bias in one row for each eye, the growth rule ensures that this bias is transferred into the other rows of the LGN.

Figure 11, picture 0 shows an initial set of weights which are biased for both topography and ocular segregation: the contralateral eye innervates all eight rows of the LGN, whereas the ipsilateral eye innervates only the bottom four rows. As can be seen from this sequence of pictures, the network transfers the bias to the remaining layers for each array. A complete topographic mapping can be found in each row of the LGN. The ocular segregation develops in the same way as for Test 2.

Similar experiments were carried out, varying the number of layers that were innervated by each eye. For example, Figure 12 shows a similar test, but in this case, the pattern of innervation is changed: the contralateral eye innervates the bottom four layers of the LGN, whereas the ipsilateral eye only innervates the bottom two layers of the LGN. Again, one row of topographic bias is given to each eye. Both topography and ocular dominance develop as normal. In general, as long as the contralateral eye innervates higher into the LGN than the ipsilateral eye, then ocular segregation and topography develop as normal. This demonstrates that the network will settle into the same 4 rows for each eye, regardless of the exact number of rows that are initially innervated.

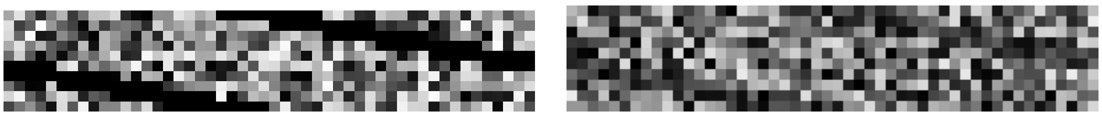


Figure 10: Example detail of initial biased weights for one row of LGN cells.

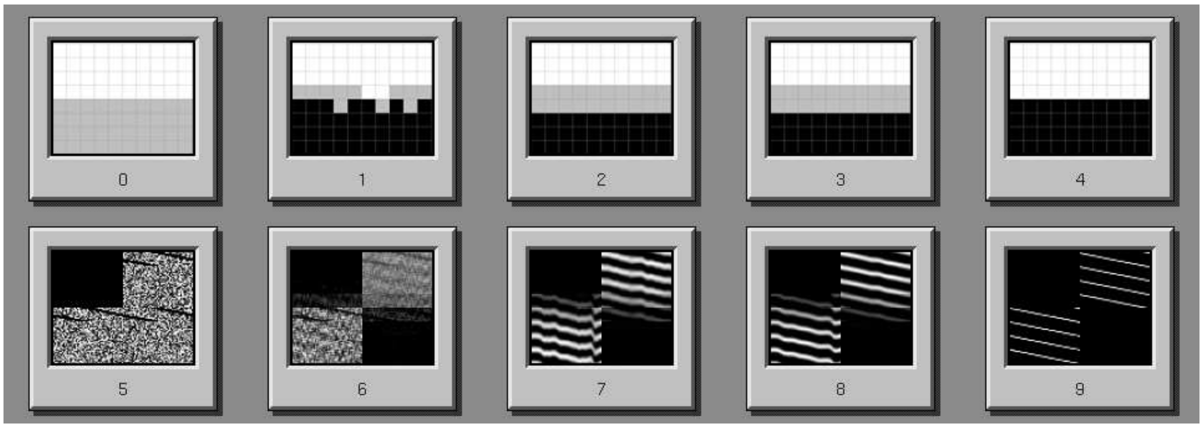


Figure 11: Test 3: Initial weights (picture 5) are biased for both ocular segregation and topography. Over time, the topographic bias transfers to all rows of the LGN and refines to produce a good topographic mapping everywhere. The ocular segregation develops in the same fashion as in Figure 9, but this time all of the cells are monocular.

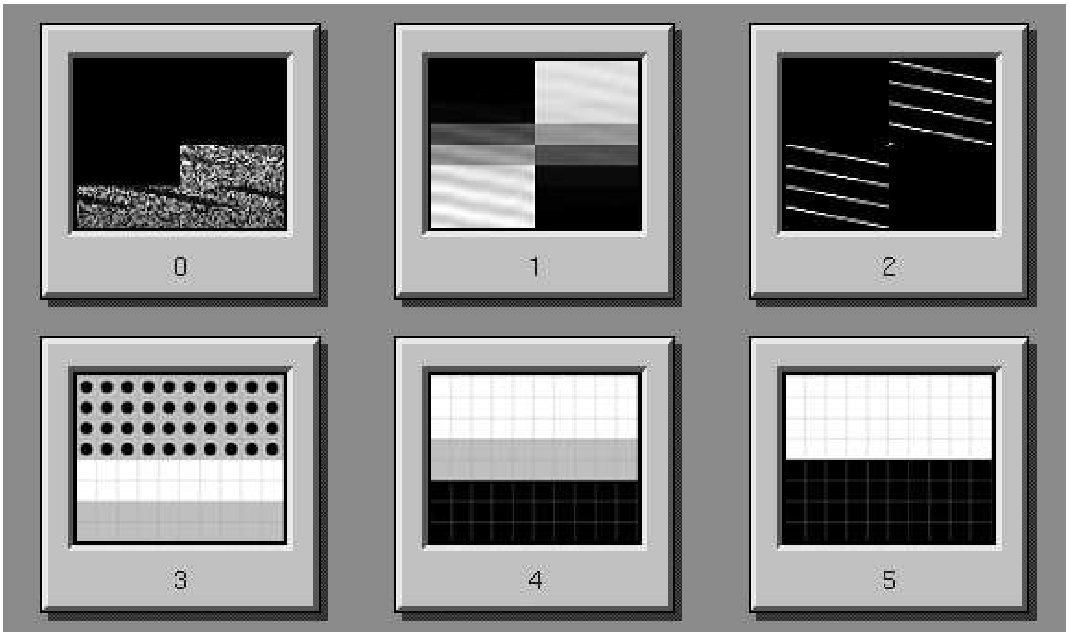


Figure 12: Test 3: Initial weights (picture 0) are biased for topography, but this time the contralateral axons initially innervate the bottom four layers of the LGN, and the ipsilateral axons only innervate the bottom two layers. Topography and ocular segregation develop as normal. (0) Initial weights, (1) 200 epochs, (2) 1200 epochs (This experiment required longer training to allow the global topography to develop.) Circles in cells of picture 3 indicate that the cell does not initially receive any input, as all of its weights are zero.

3.4 Test 4: Importance of Growth Term

This test investigates the importance of the growth term in the learning rule: what happens if there is no growth learning? Figure 13 shows the results of a typical experiment when there is no growth learning in the model. As can be seen, although the ocular segregation is formed, the topography completely breaks down on both a global and a local scale. This is not surprising, given the nature of the growth term. It seems to have two responsibilities:

- Make sure that neighbouring LGN cells become responsive to *similar* retinal cells within a row.
- Make sure neighbouring LGN cells in a column respond to the *same* retinal cells within a column. It is this aspect which ensures that the topographic bias spreads from one row of the LGN to the other unbiased rows.

Associated with the growth term is the $radius$ parameter, which controls the size of the neighbourhood for growth learning. In the previous test, $radius$ was set to 0 throughout training. However, in normal experiments, it is usually set to an initial value and then decreased over time to zero. The smaller the final value of $radius$, the more refined the final topography is, as shown in Figure 14.

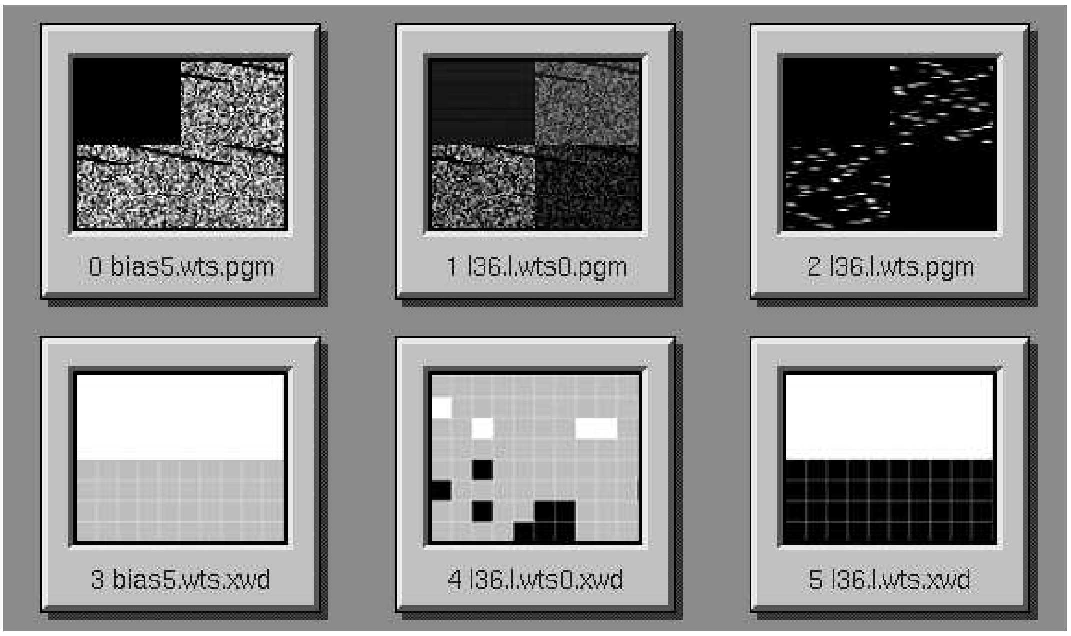


Figure 13: Test 4: Development without the growth term. Each LGN develops to a small patch of retina in just one eye, but there is no global or local topography: neighbouring cells respond to different patches of retina. (0) Initial weights, (1) 1 Epoch, (2) 500 Epochs.

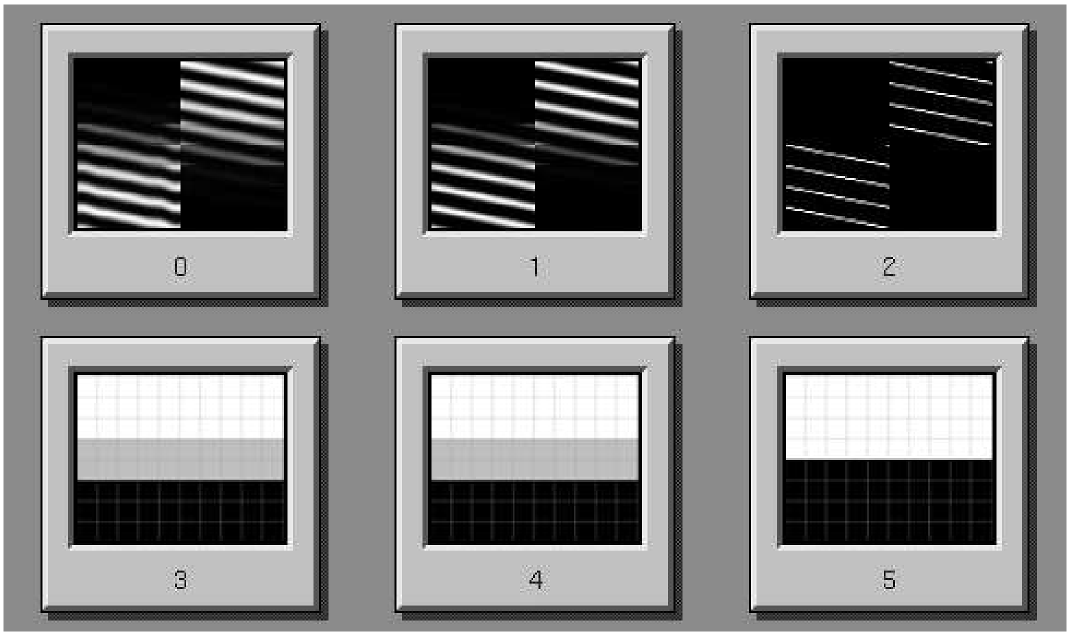


Figure 14: Effect of the radius parameter of the neighbourhood. (0) Weights after 500 epochs of training with $radius$ equal to 2, (1), $radius$ is then decreased to 1 for another 500 epochs, and (2) finally $radius$ is decreased to 0 and the network trained for another 500 epochs. The smaller the final value of $radius$, the more refined the topography is. As shown by the corresponding ocular dominance plots, the binocular region in the middle of the LGN disappears only when rad = 0.

3.5 Test 5: Importance of Normalisation

In the original model, two methods of normalisation were imposed: divisive retinal normalisation and subtractive geniculate normalisation. An important question is whether both these normalisation schemes are necessary for proper development. The normalisation methods have very different properties, as shown in (Miller & Mackay, 1994; Goodhill & Barrow, 1994). Does the network rely on particular features of each normalisation technique in order to develop, or can it use any normalisation technique? Furthermore, are the two forms of normalisation (geniculate and retinal) normalisation required? If normalisation is being used here just to keep the weights bounded, then only one normalisation scheme is sufficient.

These questions have been investigated by systematically varying the form of normalisation used at each site. Normalisation can either be subtractive or divisive at each site, or alternatively there can be no normalisation. This gives 3×3 alternatives, which are documented in Table 3. (In each case, the initial weights and the model parameters used were the same.)

		Retinal Normalisation		
		Divisive	Subtractive	None
Geniculate Normalisation	Divisive	Picture 0 Top: Global OS: Good	Picture 1 Top: Multiple OS: Good	Picture 2 Top: Small patch OS: All Contra
	Subtractive	Picture 3 Top: Global OS: Good	Picture 4 Top: Multiple OS: OK	Picture 5 Top: One cell OS: All Contra
	None	Picture 6 Top: Global OS: Good	Picture 7 Top: Single OS: OK	Picture 8 Top: Small patch OS: All Contra

Ocular Segregation Term	Meaning
Good	Top 4 layers contralateral, bottom 4 layers ipsilateral
OK	Most cells are monocular, but a few are monocular (or dead).
All Contra	All cells are contralaterally innervated.

Topography Term	Meaning
Global	Good global topography on each row of the LGN
Multiple	No global or local topography. A patch of retina may project to more than one LGN cell.
Single	No global or local topography. A patch of retina projects only to one LGN cell.
Small patch	Only a few retinal cells are connected to all LGN cells.
One cell	Only one retinal cell is connected to all LGN cells.

Table 3: Normalisation experiments. Each combination of the three normalisation methods (divisive/subtractive/none) is applied to the outputs from a retinal cell and inputs to an LGN cell. Each table entry in the top table describes the resulting topography (Top) and ocular segregation (OS). The bottom two tables provide keys for interpreting the table entries. (These two tables are ranked in order from best to worst.) Each final weight diagram and ocular dominance plot can be found under the relevant picture number in Figures 15 and 16.

From the results shown in Table 3 and Figures 15 and 16, two main points can be made. Firstly, the retinal normalisation must be present for normal development. Furthermore, this normalisation must be divisive. If subtractive retinal normalisation is used, then the topography is lost, although every retinal cell connects to the LGN, and layered ocular segregation is often achieved. The worst case is when there is no retinal normalisation, since it is possible that all the weights from a retinal cell go to zero. This means that the retinal cell is effectively dead, as its activity does not affect any LGN cells. (Most of the retinal cells in pictures 2 and 5 of Figure 15 are dead.)

This result is not surprising. Divisive and subtractive normalisation have different properties, and in this model, those differences are important. Subtractive normalisation tends to force weight values to extremes, making a local representation. Divisive normalisation, on the other hand, forms a distributed representation. This is shown in Figure 17, which compares the effects of subtractive and divisive normalisation. In this model, divisive retinal normalisation is required, as each retinal cell must synapse with LGN cells in more than one row. As shown in pictures 1,4 and 7 of Figure 15, subtractive retinal normalisation will tend to make each retinal cell synapse onto LGN cells in only one or two rows.

Secondly, as long as the retinal normalisation is divisive, the form of geniculate normalisation is irrelevant, as shown by pictures 0,3 and 6 in Figure 15. Importantly, as shown by picture 6, normal development occurs even in the absence of any geniculate normalisation. If there is no geniculate normalisation, there is no constraint that the sum of the weights to an LGN cell should be constant: it can vary from cell to cell. For example, Figure 18 shows the sum of weights to each LGN cell when there is only divisive retinal normalisation. (Figure 19 shows the corresponding histogram of weights.) It can be seen that the sum does vary for LGN cells, according to the row of the LGN to which the cell belongs: LGN cells at the top and bottom of the LGN (rows 1 and 8) receive more synaptic input than cells in the middle of the LGN (rows 4 and 5). This symmetry around the center rows of the LGN is likely to be caused by the boundary conditions of the network, although this has not been tested.

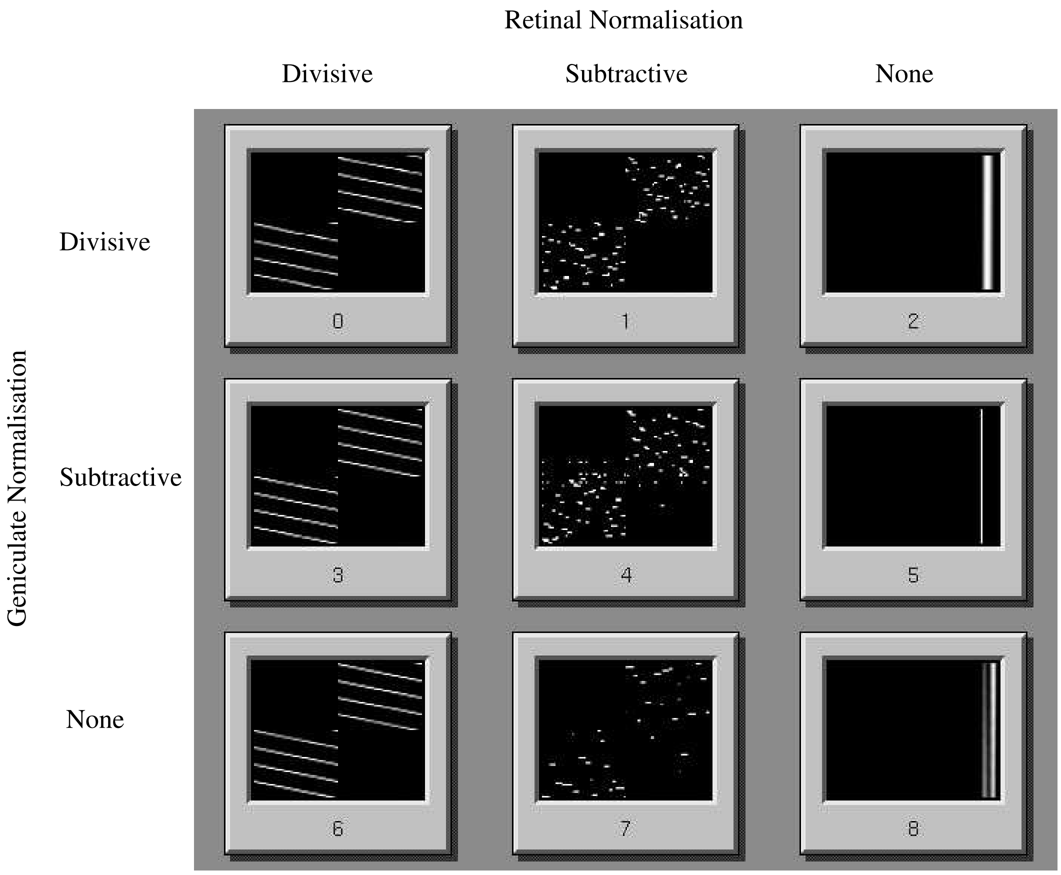


Figure 15: Final weight diagrams for the normalisation tests listed in Table 3. Note: In picture 8, the weights are free to grow without bounds and eventually cause very large weights, since there is no upper bound on individual weight strengths. (Weights are still constrained to be 0 or greater.) The largest weight value by epoch 200 is approximately 2×10^7, and so has not been plotted on the same scale as the other figures.

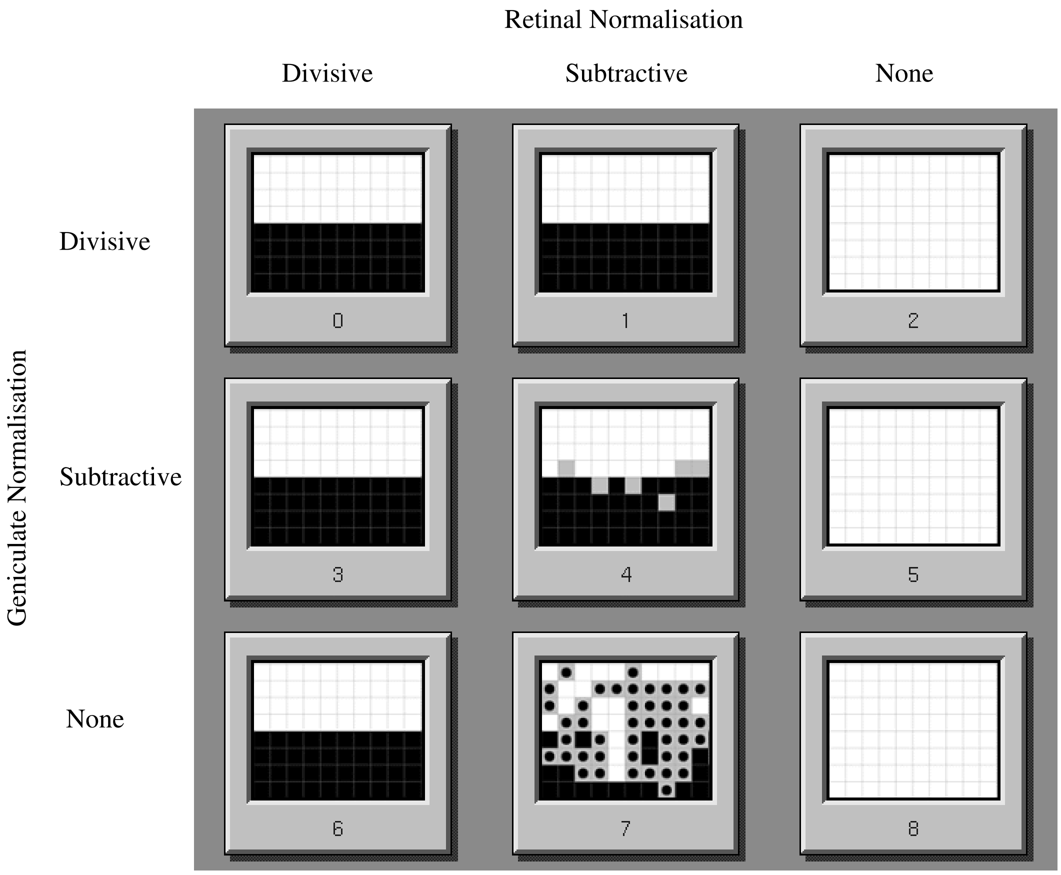
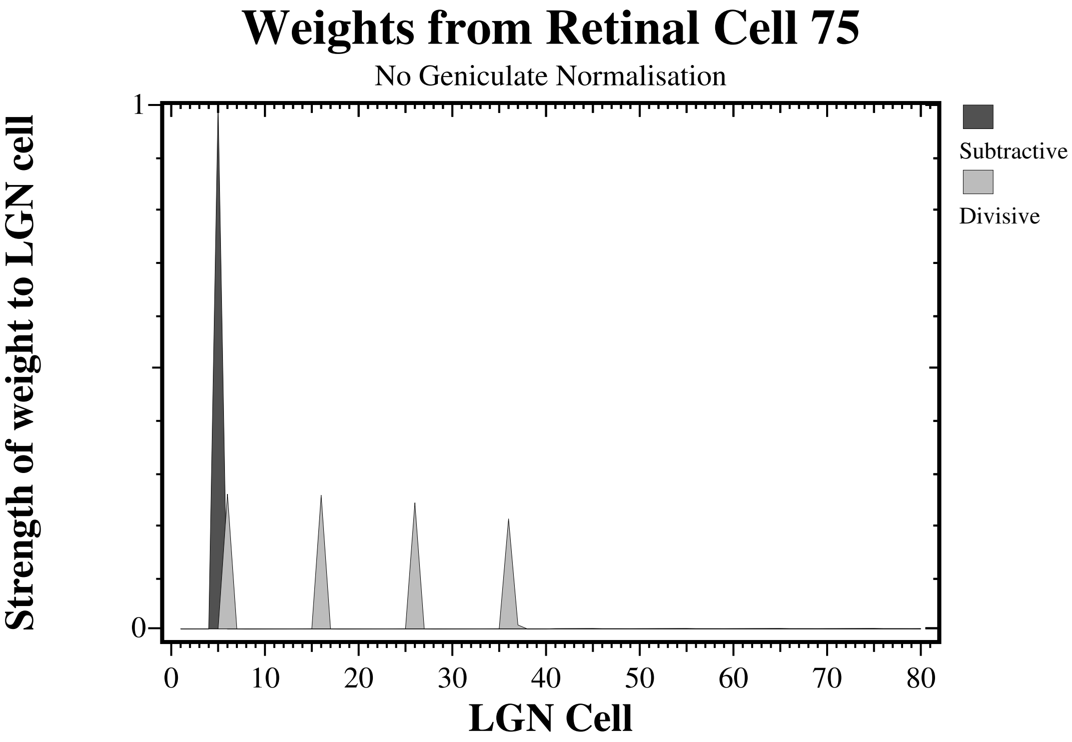


Figure 16: Final ocular dominance plots for the normalisation tests listed in Table 3. In picture 7, the cells marked with a circle indicate that all of the weights to the cell are zero – the cell is effectively "dead", rather than binocular. All other cells in this test are monocular.



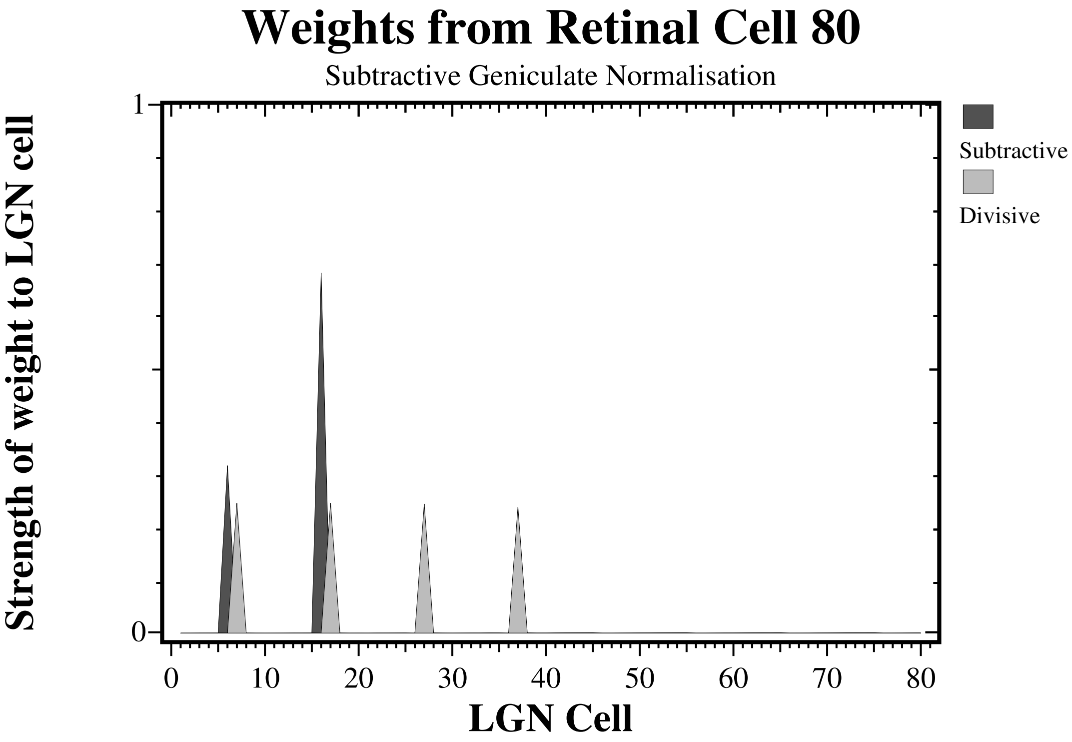


Figure 17: Detailed view of weights from one retinal cell to LGN cells under different retinal normalisation schemes. In both graphs, the total weight from a retinal cell is constrained to be 1.0. In the upper graph, there is no geniculate normalisation. In the lower graph, subtractive geniculate normalisation is also applied. In the upper graph, subtractive retinal normalisation makes a local representation, whereas the divisive retinal normalisation produces a distributed representation. In the lower graph, the situation is similar, although in the subtractive retinal normalisation case, the weights have not been pushed right to their extremes. This is because of the geniculate normalisation that is also in action. Data taken from Figure 15: Upper graph – Subtractive data taken from column 75 of picture 7, Divisive data taken from column 75 of picture 6. Lower graph – Subtractive data taken from column 80 of picture 4, Divisive data taken from column 80 of picture 3.

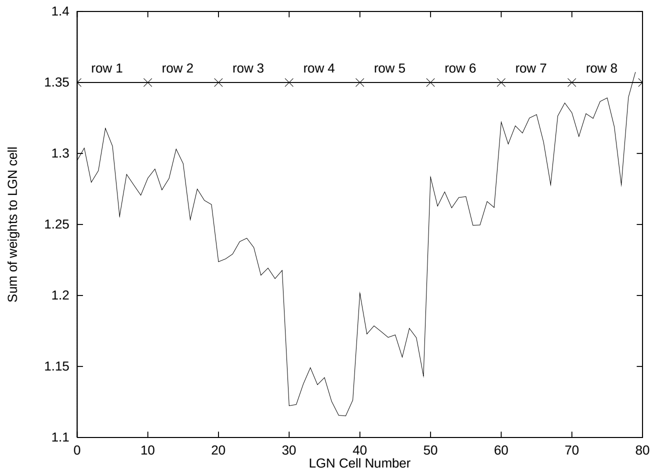


Figure 18: Sum of weights to LGN cells after learning with divisive retinal normalisation and no geniculate normalisation. Row numbers in the graph refer to the row number in the LGN of the cell.

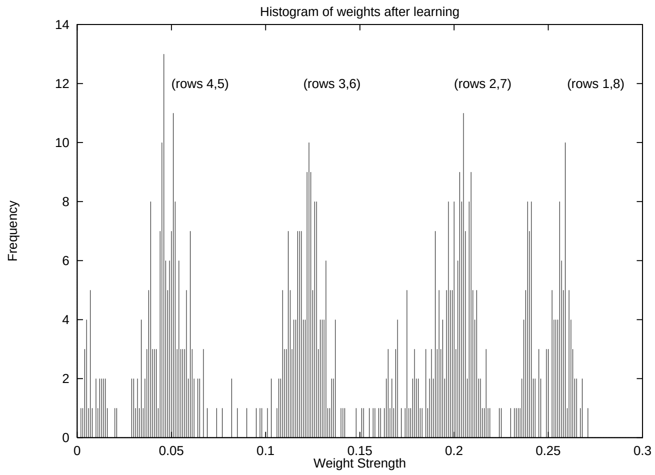


Figure 19: Histogram of weight strengths after learning with divisive retinal normalisation. Weights with zero strength are not included for clarity (7326 out of 8000 weights go to zero.) In this example, there is no geniculate normalisation, and so there is no constraint to keep the weights to a LGN at a constant value. The four distributions of weights corresponds to the different rows of the LGN as marked on the graph. (Weights used are the ones shown in Picture 6 of Figure 15). Row numbers are the same as in Figure 18.

3.5.1 Is geniculate normalisation necessary?

To investigate whether geniculate normalisation is redundant in this model, two aspects of the model were looked at in more detail: the initial conditions and the nature of the inputs.

Initial conditions

In the previous experiments, the weights were normally biased so that some of the weights were only innervated by the contralateral eye. If some LGN weights are initially monocular, could this cause the other weights to become monocular in the absence of any geniculate normalisation? This hypothesis was tested by initially allowing both eyes to innervate only the bottom two layers of the LGN. The weights were then adapted with the Hebbian rule, using only divisive retinal normalisation. As can be seen from Figure 20, after training, only 6 out of the 80 LGN cells are binocular – the rest are monocular. Similar results were attained when both eyes initially innervated all of the LGN. LGN cells therefore become monocular without geniculate normalisation regardless of whether any LGN cells are initially monocular.

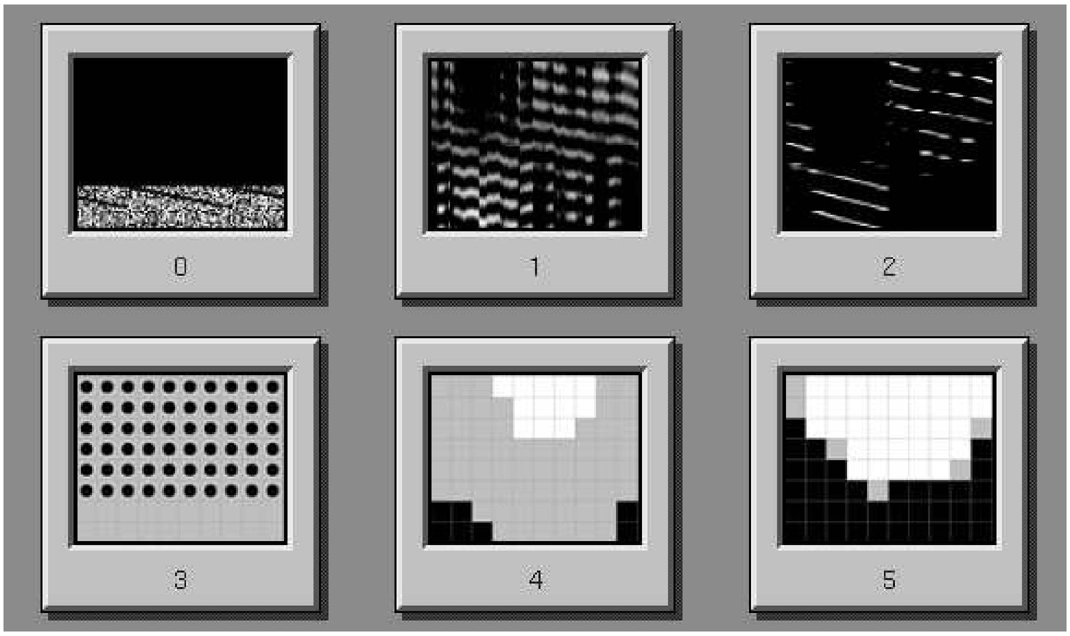


Figure 20: Testing the initial conditions of the network. All LGN cells are initially binocular or "dead" (indicated by a filled circle). (0) Initial weights, (1) 100 epochs, (2) 700 epochs. Pictures 3,4 and 5 show the corresponding ocular dominance plots.

Nature of the Inputs: Fixed Overlapping Waves

In the previous experiments, the inputs to the network were those shown in Figure 5: while one eye is active, the other is silent. A more challenging set of inputs for the network is to have a period when the two eyes are active at the same time, as shown in Figure 21, which will be called the fixed overlapping waves. The percentage of time for which both eyes are simultaneously active is given by:

$$\text{overlap \%} = \frac{n - offset}{n + offset} \times 100$$

The retinal waves are again modelled as gaussian distributions. Divisive retinal normalisation was used, both with and without subtractive geniculate normalisation with these retinal inputs. The results are shown in Figure 22. In both cases, normal topography and ocular dominance did not develop, even when the offset was as high as 45 (an overlap of 5.3%). Instead, some LGN cells responded to a few cells in both

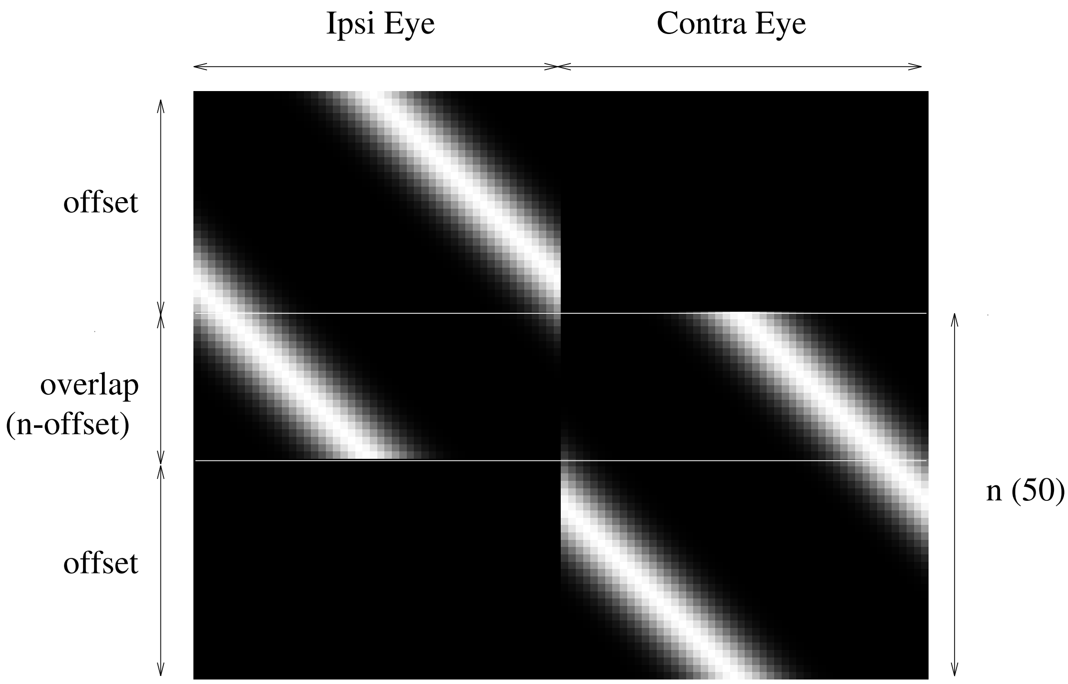


Figure 21: Fixed Overlapping Waves. The offset in this case is 30, and the period of overlap is 20 (25% overlap). The region of overlap is outlined by the two horizontal white lines drawn on top of the image. n corresponds to the number of pixels in each eye (50 in this case).

retinae. Not surprisingly, these retinal cells were the ones that were always active at the same time as cells in the other eye, as shown in Figure 23. This is to be expected, as the Hebbian learning rule will increase the weights of retinal cells that fire at the same time onto an LGN cell, regardless of whether the retinal cells are in the same eye or not.

As shown in Figure 25, these fixed overlapping waves have large positive between eye correlations. Previous correlational models, eg (Miller et al., 1989) require zero between eye correlations for normal ocular dominance development – the addition of positive between eye correlations causes cells to become binocular (Dayan & Goodhill, 1992; Goodhill, 1992), although very small positive between eye correlations do not prevent the development of monocular cells (Miller & Mackay, 1994, Unpublished observations, p114). This model also seems to require the same constraint on the correlations between input cells to ensure that cells become monocular.

Random Overlapping Waves

To investigate whether this model could develop normally using inputs that do have a period of overlap when both eyes are active, a new form of retinal inputs were used. For each epoch, a new set of retinal inputs were created by allowing the starting point and direction of each wave to vary at random:

Time step	Action
0	Select a random starting position [0-49] and direction [left/right] for the ipsilateral eye. Start the wave.
offset	Pick a random starting position and direction for the contralateral eye. Start the wave.
50	Wave in ipsilateral eye finishes: eye goes quiet.
50 + *offset*	Contralateral eye wave finishes: eye goes quiet.

Using these retinal inputs, there is still a period of overlap when both eyes are active. However, in this case, the retinal cells which are firing at the same time as cells in the other eye should always be different every epoch. This effectively reduces the between eye correlations, whilst still allowing cells in both eyes to fire at the same time.

The results of using these new retinal inputs are shown in Figure 24, varying the value of the offset parameter. Divisive retinal normalisation was used, either with or without subtractive geniculate normalisation. When the offset is 30 (25%), the topography and ocular segregation develops more or less as normal. However, as the degree of overlap increases, the global topography begins to break up: retinal cells tend to synapse onto two different rows of LGN cells rather than the normal four. The ocular segregation also tends to break down around the border region between rows four and five of the LGN. It is interesting to note though that even with complete overlap (offset 0), cells still become monocular.

Comparing the effects of the geniculate normalisation, it seems, on a quantitative basis, that for each value of *offset* chosen, the results with no geniculate normalisation were slightly better than with the subtractive geniculate normalisation.

It therefore seems that the model can develop normally when there is a limited period of overlap between the two eyes firing, as long as the retinal cells firing during the period of overlap changes from epoch to epoch. The effect of this on the correlation matrix of the inputs can be seen in Figure 25: the between eye correlations are spread out, so that many retinal cells between eyes are slightly correlated rather than a few retinal cells in different eyes being highly correlated. Furthermore, the network can develop as normal without geniculate normalisation.

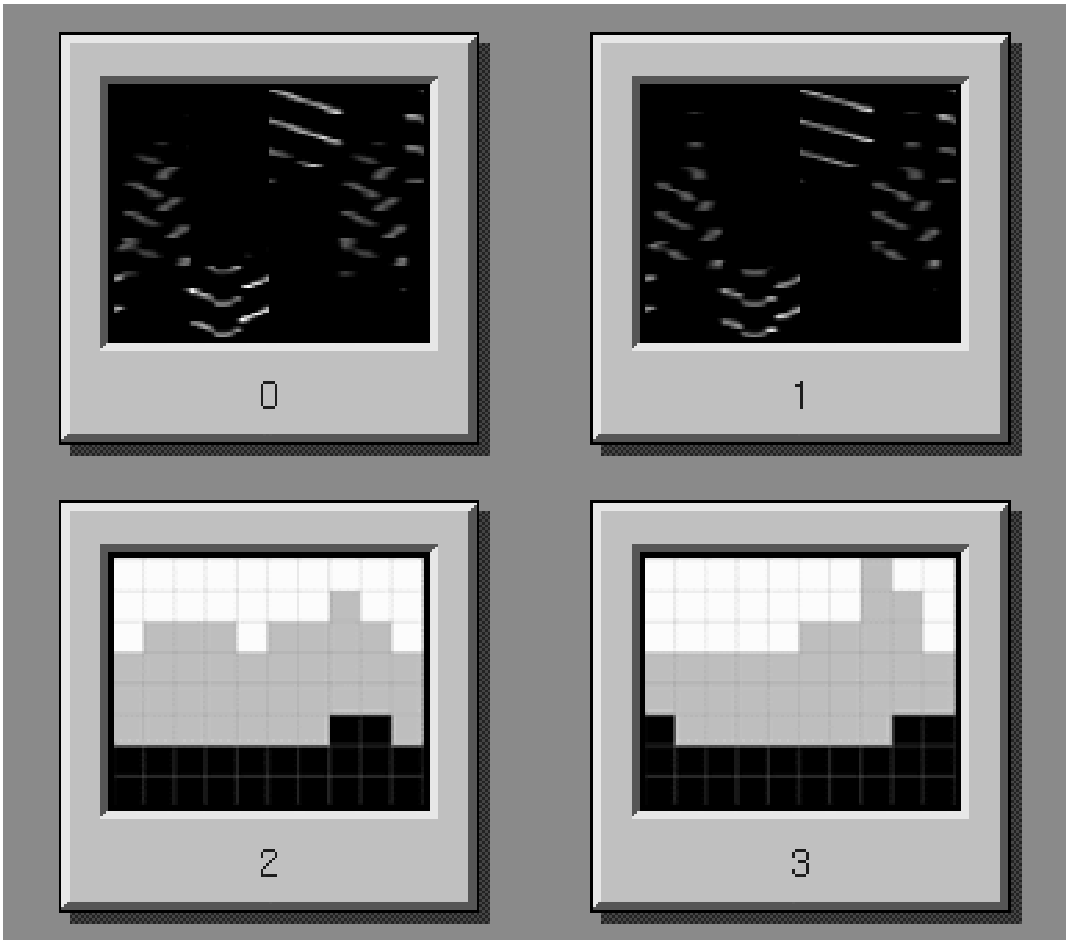


Figure 22: Results of training the network with the fixed overlapping waves (shown in Figure 21) with subtractive geniculate normalisation (picture 0) and without any geniculate normalisation (picture 1). Corresponding ocular dominance plots are shown underneath the weight diagrams.

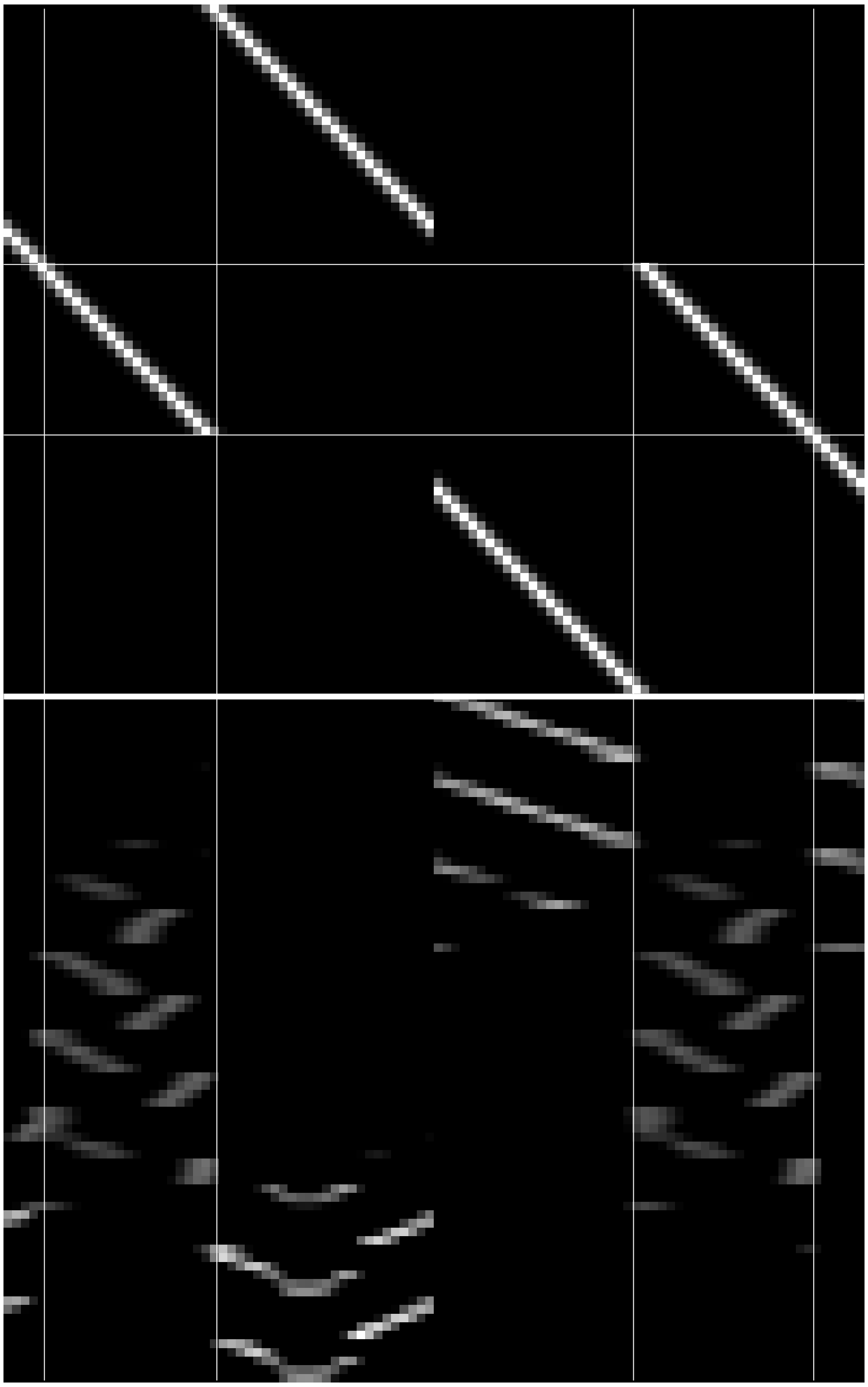


Figure 23: Composite image showing the fixed overlapping waves and the final weights from picture 0 of Figure 22. The topography and ocular segregation is lost in the LGN cells which respond to retinal cells that are always active at the same time as retinal cells in the other eye. (Extra horizontal and vertical lines have been added to the image to show the connection between the area of overlap in the input vectors and the loss in topography and ocular segregation in the weights diagram.)

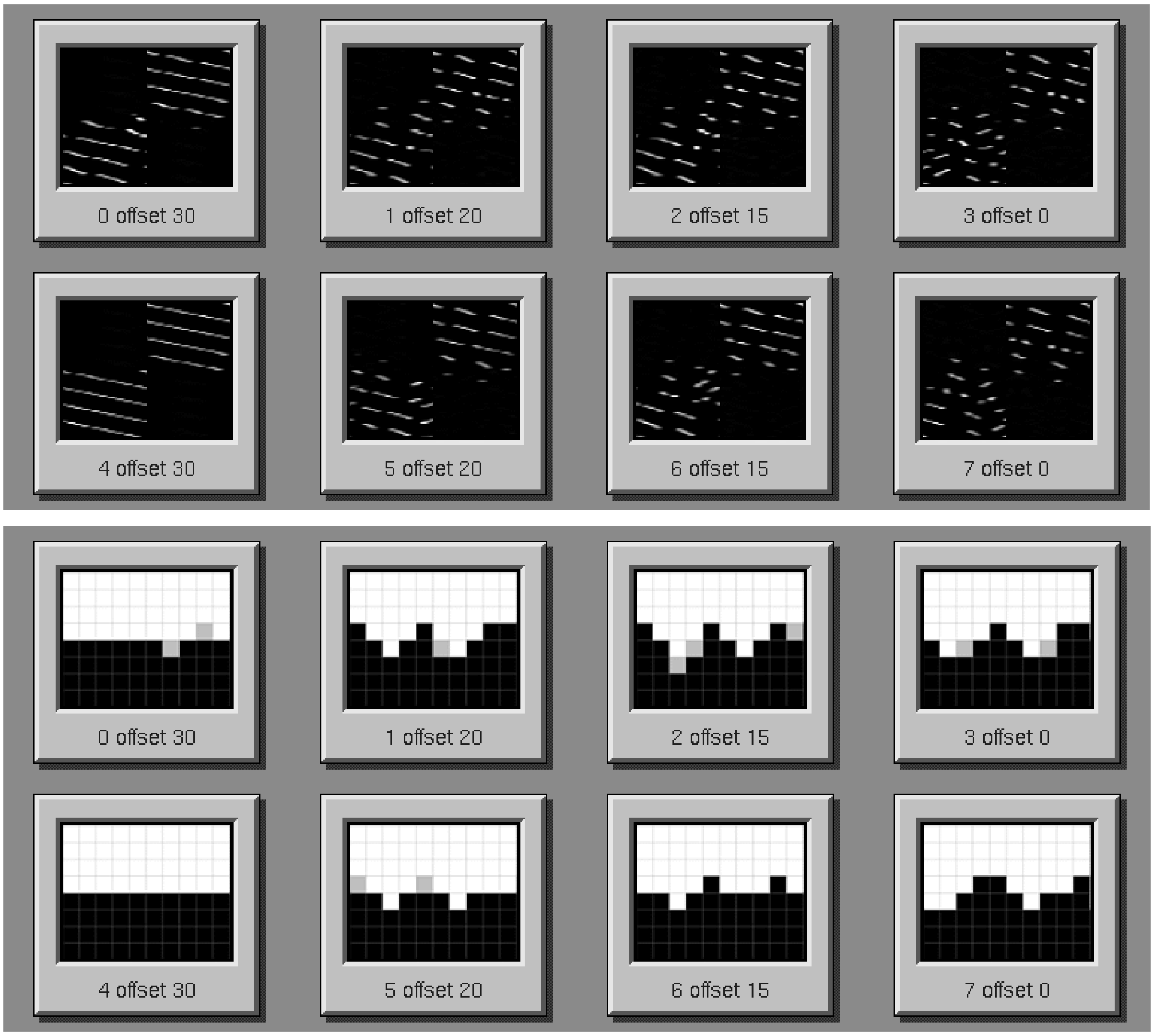
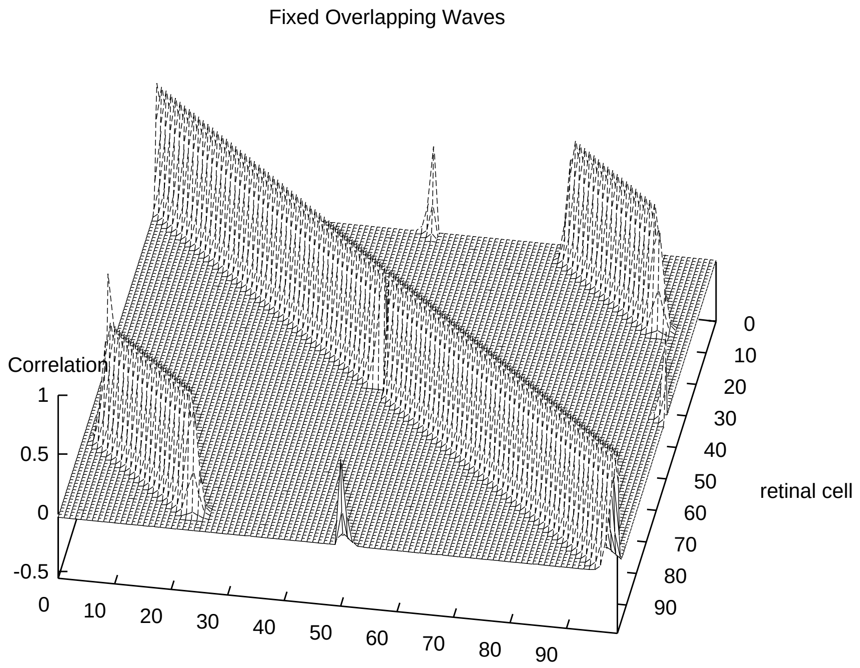
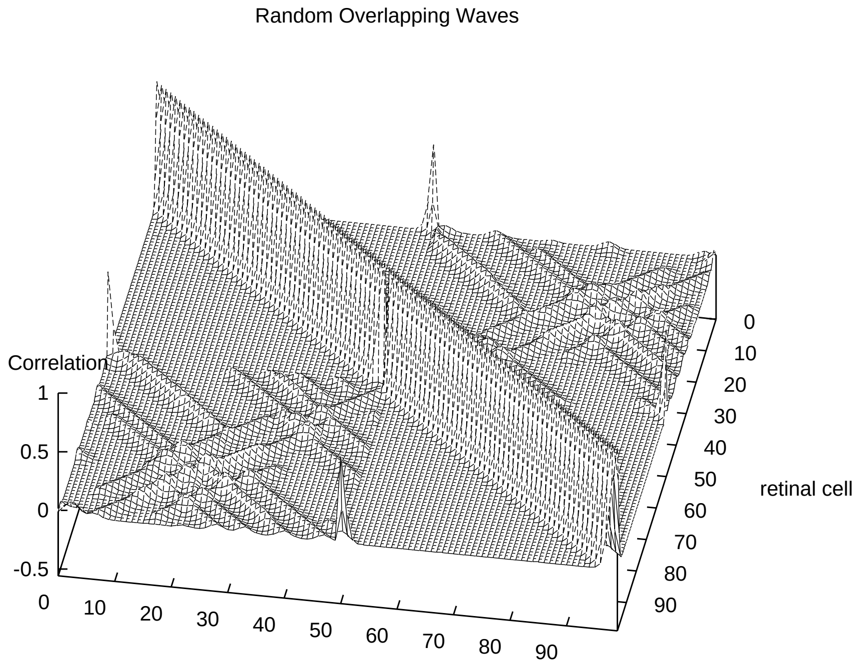


Figure 24: Final weight patterns using various values of the *offset* parameter to create the random overlapping waves. Pictures 0,1.2 and 3 are with subtractive geniculate normalisation. Pictures 4,5,6 and 7 are without any geniculate normalisation.



(a)



(b)

Figure 25: Correlation matrices for the 2 forms of retinal waves: a) Fixed overlapping waves, b) Random overlapping waves. Correlation matrix for random overlapping waves created by generating 20 epochs of random waves. In both graphs, the peaks at (0,50), (50,0), (100,50) and (50,100) are caused by the toroidal nature of each retina. The correlation matrix is defined in Appendix B.

Imposing maximum weight values

It has already been shown that the retinal normalisation must be divisive in order to produce the normal topographic mapping (as shown in Figure 15). Since there is no individual maximum weight value imposed in the model, subtractive normalisation normally forces all of the synaptic weight strength to one cell, as shown by (Miller & Mackay, 1994, p111) and by Figure 17. If however there is a maximum individual weight strength ($maxWt$), then more than one weight is forced to be non zero. (This assumes $maxWt < tR$. The case when $maxWt \geq tR$ is uninteresting.)

Experiments were therefore conducted to see what happens when subtractive retinal normalisation is subject to the constraint that each weight must be no larger than $maxWt$. Hence weights are now bounded between $[0, maxWt]$. (This constraint must also be enforced during both hebbian learning and growth steps). This form of normalisation will be called capped normalisation.

The results of these experiments are shown in Figure 26. As can be seen from picture 1, capped subtractive retinal normalisation can produce the desired topography, but only under certain conditions:

1. Normalisation must take place at the end of an epoch rather than after every iteration.
2. The subtractive normalisation must be applied "slowly", using the parameter rs, which describes the rate at which subtractive normalisation is applied [1]. (Compare pictures 0 and 1 in Figure 26.)
3. The value of $maxWt$ is critical for normal development. (Compare picture 1 ($maxWt$ = 0.20) with picture 2 ($maxWt$ = 0.50) in Figure 26.)

Also in these experiments, there was no geniculate normalisation, which seemed to produce slightly better results.

However, there is one main difference between the weights learnt by divisive retinal normalisation and subtractive capped retinal normalisation: the weights with the divisive normalisation are distributed quite evenly in the range [0,0.27] (as shown in Figure 19), whereas 98 % of the weights learnt with subtractive capped retinal normalisation are either 0 or $maxWt$. Therefore, capped subtractive normalisation also pushes weights to extreme values.

[1] The value of rs determines how far a weight vector is moved to its fully normalised value, and should be in the range [0,1]. The modified equation for implementing subtractive normalisation is then $w'_{ij} = w_{ij} - \frac{rs}{nR}\left(tL - \sum_{k=1}^{nL} w_{kj}\right)$ Compare this with the usual form of subtractive normalisation on page 5 .

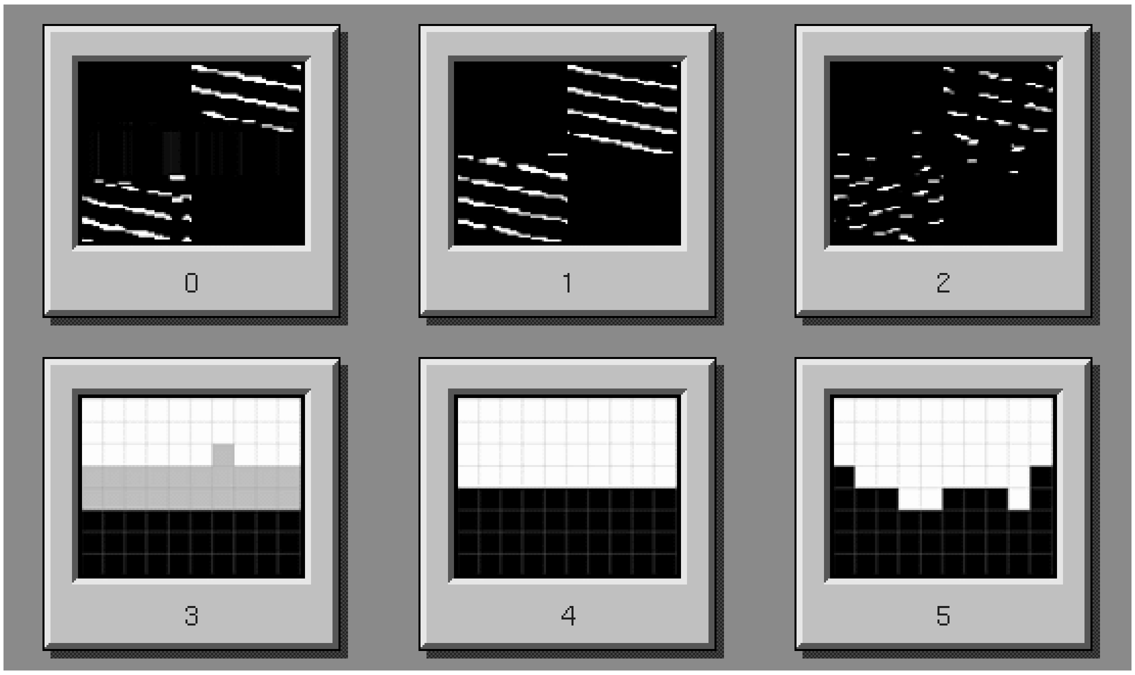


Figure 26: Effect of learning with subtractive retinal normalisation constrained also such that each weight is bounded in the range $[0, maxWt]$. No geniculate normalisation, (0) rs = 1.0, $maxWt$ = 0.20 (1) rs = 0.2, $maxWt$ = 0.2, (2) rs = 0.2, $maxWt$ = 0.50. Pictures 3,4 and 5 show the corresponding ocular dominance plots. Non overlapping retinal inputs were used for these experiments, but experiments with the random overlapping inputs produced similar results.

Satisfying both forms of normalisation.

When both retinal and geniculate normalisation is being used in the model, there is the problem of how to satisfy both forms of normalisation at the same time. If the normalisations are applied sequentially, there is no guarantee that imposing the second normalisation will not destroy the constraints imposed by the first normalisation. Some tests were therefore performed to see how model performance is affected by the interactions between the two forms of normalisation.

In the previous experiments, at the end of every epoch, retinal normalisation was applied first followed by geniculate normalisation. In this set of experiments, the probability of geniculate normalisation occurring first (pN) was allowed to vary between 0 (divisive retinal normalisation is always applied first for this epoch) and 1 (subtractive geniculate normalisation is always applied first for this epoch). The results are shown in Figure 27: it is clear that the normal topography develops only if the divisive retinal normalisation is applied before the subtractive geniculate normalisation. (In contrast, both Miller et al. (1989) and Goodhill (1992) apply the output cell normalisation before the input cell normalisation.)

Why does applying subtractive geniculate normalisation first produce strange results?

One possibility why applying subtractive geniculate normalisation first fails to produce the normal topography could be that both forms of normalisation cannot be satisfied if the geniculate normalisation is applied first. To quantify how well the two forms of normalisation can be simultaneously satisfied, two error measures were produced:

$$\text{retinal error} \qquad re = \left[\frac{1}{nR} \sum_{i=1}^{nR} \left(tR - \sum_{j=1}^{nL} w_{ij} \right)^2 \right]^{\frac{1}{2}}$$

$$\text{geniculate error} \qquad ge = \left[\frac{1}{nL} \sum_{j=1}^{nL} \left(tL - \sum_{i=1}^{nR} w_{ij} \right)^2 \right]^{\frac{1}{2}}$$

These measures give the root mean squared difference of the expected normalised weights and the actual weights. Table 4 shows the error values for weights at different times during learning, and for different values of pN. From this table, it can be seen that the difference between the actual normalised values and the desired normalised values are quite small, even though the normalisation that is applied first (for example, retinal normalisation when $pN = 0.00$) tends to have a larger error. From these error measures, it is reasonable to conclude that both forms of normalisation can be approximately satisfied at the same time, regardless of whether retinal or geniculate normalisation is applied first. The possibility that both forms of normalisation cannot be simultaneously satisfied when the subtractive geniculate normalisation is applied first must therefore be rejected. (Unpublished observations reported in (Miller & MacKay, 1992, p31) also confirm that in practice both forms of normalisation can be satisfied.)

Weights tested	Normalisation errors.							
	$pN = 0.00$		$pN = 0.05$		$pN = 0.20$		$pN = 1.00$	
	re	ge	re	ge	re	ge	re	ge
Epoch 1	1.09e–02	7.39e–06	1.09e–02	7.39e–06	1.09e–02	7.39e–06	1.02e–06	5.70e–01
Epoch 350	6.03e–04	8.87e–06	1.61e–06	3.31e–03	1.45e–06	5.45e–03	6.00e–07	9.36e–03
Epoch 700	6.76e–05	7.85e–06	9.91e–04	6.48e–06	1.74e–03	7.34e–06	4.80e–07	2.75e–02

Table 4: Normalisation error measures re and ge for different values of pN at different times during learning. See Figure 27 for the corresponding weight diagrams and ocular dominance plots.

Gradual Application of Subtractive geniculate normalisation.

Returning to the original paper by Keesing et al. (1992), the geniculate normalisation is introduced by saying:

> The weights supported by each LGN cell are also renormalised, *gradually* driving them toward some target value T: $w_{ij}(t+1) = w_{ij}(t) + [T - \sum_k w_{kj}(t)]$ [emphasis added].

Unfortunately, details are not given as to how the normalisation is gradually applied. One way of implementing this form of normalisation is to include an extra term, rs, to the normalisation, as shown on page 27. If this term is included into the geniculate normalisation, the order in which the two normalisations are performed no longer affects the resulting topography. For example, if the probability of geniculate normalisation occurring first (pN) is 0.2 and $rs = 0.1$, the normal patterns of topography and ocular segregation develop. (Compare this with picture 2 of Figure 27, when $pN = 0.2$, and rs is implicitly set to 1.0). One consequence of gradually applying the subtractive geniculate normalisation is that the average geniculate error (the average value of ge during learning) is much higher: it is typically 1.17e–1 when $rs = 0.1$, compared to 1.44e–3 when $rs = 1.0$.

These results show that the order of normalisation is important for normal development of the LGN, unless the subtractive geniculate normalisation is applied slowly.

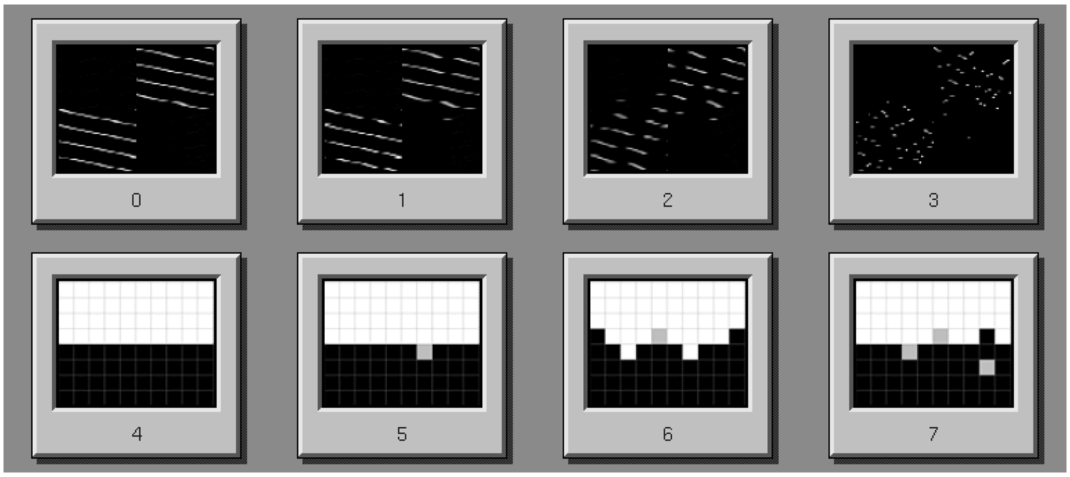


Figure 27: Effect of varying the order of normalisation. Both divisive retinal normalisation and subtractive geniculate normalisation are used. The probabilities of geniculate normalisation occurring first are: (0) $pN = 0.00$ (1) $pN = 0.05$ (2) $pN = 0.20$ (3) $pN = 1.00$. Corresponding ocular dominance plots are shown beneath each weight diagram. For these experiments, the inputs were random overlapping waves with *offset* equal to 35.

Conclusions and Further Work

Simple Hebbian learning rules, together with a growth rule and normalisation, can reproduce the patterns of ocular segregation and topography as found in coronal sections of the cat LGN. However, these activity dependent processes are not sufficient on their own, and must be used alongside activity independent processes. In the cat LGN, these activity independent processes are the biased arrival times of the contralateral axons and the topographic bias formed by chemical gradients. These have been modelled by introducing biases into the initial weights of the network.

As with other correlational models of ocular dominance, for example (Miller et al., 1989), the nature of the correlations between the eyes is important: the between eye correlations are usually zero. It has been shown here that this does not imply the eyes cannot fire together: using the random overlap waves, both eyes can be active for at least 25% of the time (producing small between eye correlations) without affecting the development of ocular segregation and topography. This seems to support the hypothesis suggested by Goodman and Shatz (1993):

> These waves could act to provide the local correlations necessary for topographic fine tuning of retinal projections to the LGN, and because they are random in direction in each eye, would rarely if ever be correlated between the two eyes. (Goodman & Shatz, 1993, p91)

More research is required however to investigate the precise nature of the between eye correlations in the developing cat, and to see how this compares with the maximum between eye correlations allowed in this model. Figures from a related species, the ferret,[2] look promising. For example, at postnatal day 5, the average duration of a wave is between 2–4 seconds with a mean inter-burst interval of around 100 seconds (Wong et al., 1993, p926). These waves are likely to produce very low correlations (if any) between the eyes.

Normalisation

The model by Keesing et al. (1992) used both retinal and geniculate normalisation. The results in this paper show that these normalisation schemes were not used just to keep the weights bounded, but to take advantage of their other properties for producing certain types of weight structures.

Firstly, it has been shown that the retinal normalisation in this model must be divisive. Divisive normalisation has the property of producing a graded or distributed representation (Miller & Mackay, 1994, p121), and this property is essential to allow the topography to develop in each row of the LGN. If there is no retinal normalisation in this model, it fails to develop the normal topography. This also holds for subtractive normalisation, unless individual weights are capped. (However, the capped subtractive retinal normalisation is very fragile: it is highly sensitive to parameter values. The divisive retinal normalisation on the other hand is much more robust, and does not need an externally imposed maximum weight value.) This result agree with the conclusions of Miller and Mackay that:

> different choices of competitive [normalisation] mechanism can yield different outcomes, so it is important for the modellers to know whether and how their results depend on these choices (Miller & Mackay, 1994, p121)

It has also been shown that geniculate normalisation is not required in this model. This was slightly surprising, given the earlier results by Miller and Mackay (1994) and Goodhill (1992). These results however are not directly applicable to the work presented here for the following reasons.

Firstly, the model of ocular dominance and topography developed by Goodhill (1992) required subtractive normalisation of weights to an output cell, along with divisive normalisation of weights from an input cell. This result is not contradictory to the results presented here, since Goodhill's model differs in two aspects: (1) the model uses fully distributed patterns of activity in each retina in comparison to the local retinal waves used here, and (2) the model uses a Kohonen like competitive learning rule, rather than the correlational rule used here.

[2] The ferret is normally studied in preference to the cat for several practical reasons. One main reason is that ferrets are born much earlier in their development than cats, and so much of the development is postnatal, which is easier to study.

Secondly, Miller and Mackay (1994) reported that the form of normalisation for output cells depends on the form of correlations in the inputs. If there are anticorrelations in the inputs, divisive normalisation of output cells is sufficient. If there are only positive correlations in the inputs, subtractive normalisation of weights to an output cell must be used. However, this analysis only considered normalisation of weights to an output cell (geniculate normalisation): normalisation of weights from an input cell (retinal normalisation) is not considered in detail.

Work is currently underway analysing the model in terms of the eigenvectors of the correlation matrix of the inputs. This will permit a more detailed comparison to be made with the results of Miller and Mackay (1994). The aim of this comparison will be to see if a similar argument can be made for normalisation of weights from an output cell to the argument presented in the previous paragraph for normalisation of weights to an output cell.

More simulation work is also to be performed. Firstly, the LGN model will be three dimensional rather than two dimensional, with corresponding two dimensional retinal inputs. More realistic retinal inputs will also be considered, including the introduction of different retinal ganglion cell classes.

Acknowledgements

I would like to thank Harry Barrow, Julian Budd and Geoff Goodhill for reading draft versions of this report.

A Parameters Used.

Table 5 shows the parameters of the model, along with typical values.

Parameter	Use/Meaning	Value
α	Average Input Value, used in learning rule	0.1
β	Average Output Value, used in learning rule	0.0125
γ	Constant for growth term	0.1
$radius$	radius of neighbourhood	[0,1,2]
nL	Number of LGN cells	80
nR	Number of Retinal cells	100
tR	Target Sum for weights from each Retinal Cell	1.0
tL	Target sum for weights to each LGN cell	1.25

Table 5: Parameters and typical values used in the simulation.

There is a relationship between the target sums tR and tL:

$$\text{average weight value} = \frac{tR}{nl} = \frac{tL}{nR}$$
$$\frac{tL}{tR} = \frac{nR}{nL}$$

B The Correlation Matrix

The definition of the correlation matrix for Figure 25 is defined as follows. If x_i is the activity of retinal cell i, then the correlation matrix is defined as:

$$Corr(i,j) = \frac{Cov(i,j)}{\sqrt{Cov(i,i) \times Cov(j,j)}}$$

where Cov is the covariance matrix:

$$Cov(i,j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

and $\langle\ \rangle$ denotes the averaging operator over all input patterns.

References

Dayan, P. S., & Goodhill, G. J. (1992). Perturbing hebbian rules. In Moody, J. E., Hanson, S. J., & Lippman, R. P. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 4. Morgan Kauffman, CA.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*, 1–47.

Goodhill, G. J. (1992). Correlations, Competition and Optimality: Modelling the Development of Topography and Ocular Dominance. DPhil Thesis, School of Cognitive and Computing Sciences, Sussex University, UK.

Goodhill, G. J., & Barrow, H. G. (1994). The role of weight normalization in competitive learning. *Neural Computation*, *6*, 255–269.

Goodman, C. S., & Shatz, C. J. (1993). Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell*, *72*, 77–98. Joint Supplement with Neuron, Vol 10.

Keesing, R., Stork, D. G., & Shatz, C. J. (1992). Retinogeniculate development: the role of competition and correlated retinal activity. Tech. rep. CRC-TR-92-02, Ricoh California Research Centre (Also in Neural Information Processing Systems 4, ed. J. E. Moody et al. 1992).

Meister, M., Wong, R. O. L., Baylor, D. A., & Shatz, C. J. (1991). Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science*, *252*, 939–943.

Miller, K. D., Keller, J. B., & Stryker, M. P. (1989). Ocular dominance column development — analysis and simulation. *Science*, *245*(4918), 605–615.

Miller, K. D., & MacKay, D. J. C. (1992). The role of constraints in hebbian learning. Tech. rep., CNS Memo 19, California Institute of Technology.

Miller, K. D., & Mackay, D. J. C. (1994). The role of constraints in hebbian learning. *Neural Computation*, *6*, 100–126.

Roe, A. W., Garraghty, P. E., Esguerra, M., & Sur, M. (1993). Experimentally-induced visual projections to the auditory thalamus inferrets - evidence for a w cell pathway. *Journal of Comparative Neurology* , *334*, 263–280.

Sanderson, K. J. (1971). The projection of the visual field to the lateral geniculate and medial interlaminar nuclei in the cat. *Journal of Comparative Neurology*, *143*, 101–118.

Shatz, C. J. (1994). Role for spontaneous neural activity in the patterning of connections between retina and LGN during visual system development. *International Journal of Developmental Neuroscience*, *12*(6), 531–546.

Shatz, C. J., & Stryker, M. P. (1988). Prenatal tetrodotoxin infusion blocks segregation of retinogeniculate afferents. *Science*, *242*, 87–89.

Sherman, S. M. (1985). Development of retinal projections to the cat's lateral geniculate nucleus. *Trends in Neuroscience*, *8*(8), 350–355.

Sherman, S. M., & Koch, C. (1986). The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus. *Experimental Brain Research*, *63*, 1–20.

Sur, M. (1995). Visual activity and cortical development. In *Cerebral Cortex: Function and Development, (Lyon, France)*.

Sur, M., Garraghty, P. E., & Roe, A. W. (1988). Experimentally induced visual projections into auditory thalamus and cortex. *Science*, *242*, 1437–1441.

Wiesel, T. N., & Hubel, D. H. (1963). Single cell responses in striate cortex of kittens deprived of vision in one eye. *Journal of Neurophysiology*, *26*, 1003–1017.

Willshaw, D., & von der Malsburg, C. (1976). How patterned neural connections can be set up by self organization. *Procedings of the Royal Society of London B*, *194*, 431–445.

Willshaw, D. J., & Von der Malsburg, C. (1979). A marker induction mechanism for the establishment of ordered neural mappings: Its application to the retinotectal problem. *Philosophical Transansctions of the Royal Society, London B*, *287*(1021), 203–243.

Wong, R. O. L., Meister, M., & Shatz, C. J. (1993). Transient period of correlated bursting activity during development of the mammalian retina. *Neuron*, *11*(5), 923–938.